Advances in Intelligent Systems and Computing 687

Plamen Angelov
Jose Antonio Iglesias
Juan Carlos Corrales *Editors*

Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change

Proceedings of the International Conference of ICT for Adapting Agriculture to Climate Change (AACC'17), November 22–24, 2017, Popayán, Colombia

Springer

Advances in Intelligent Systems and Computing

Volume 687

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

More information about this series at http://www.springer.com/series/11156

Plamen Angelov · Jose Antonio Iglesias
Juan Carlos Corrales
Editors

Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change

Proceedings of the International Conference of ICT for Adapting Agriculture to Climate Change (AACC'17), November 22–24, 2017, Popayán, Colombia

Springer

Editors

Plamen Angelov
School of Computing and Communications
Lancaster University
Lancaster
UK

Juan Carlos Corrales
Campus de Tulcán
University of Cauca
Popayán
Colombia

Jose Antonio Iglesias
Computer Science Department
Carlos III University of Madrid
Leganés, Madrid
Spain

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-70186-8 ISBN 978-3-319-70187-5 (eBook)
https://doi.org/10.1007/978-3-319-70187-5

Library of Congress Control Number: 2017957851

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Agriculture is a significant economic activity for almost every country in the world, but this one is highly exposed to climate change since it assumes that the farming production depends on weather conditions. Therefore, different research challenges have been proposed and established to improve the sector's profitability. In this sense, over the last decade, the Informatics and Communication Technology (ICT) has emerged as a solution to improve the agricultural production. This trend can be reflected in the multiple software applications that provide services to lessen the effects caused by diseases and pests in crops, or in the real-time monitoring of weather conditions and the water quality parameters used for production. Management activities, decision-making, and management of complex events are cornerstones assumed by the ICT solutions applied to the agriculture.

This international conference looked at emerging problems and new ICT solutions to address the effects of climate change and climate variability on agricultural sector, proposing common strategies and guidelines for incorporating risk management and adaptation to climate change. Topics such as smart farming, systems for prediction of disease or pests, water efficiency, climate and its effects on crop production, land cover modeling for forest and crop systems, remote sensing for crop production and management, meteorological data integration for agriculture were covered by the authors.

The conference was organized by the Inter-Institutional Network for Climate Change and Security Food of Colombia—RICCLISA, COLCIENCIAS, University of Cauca, Colombia (Telematics Engineering Group) and the Carlos III University of Madrid, Spain (CAOS research group), and technically sponsored by Springer.

The conference received 29 submissions from 68 authors from 9 countries. This volume collects 19 papers accepted and presented, confirm thus ascertaining its status of the international event. The papers were grouped into the following sessions: decision support and early warning systems for agriculture, sustainable water management, predictive models of growth and crop production, land cover dynamics for agricultural systems, and networking for supporting the adaptation of climate change and food security.

This conference provided a friendly atmosphere and will be a leading international forum focusing on discussing problems, research, results, and future directions in the application of information and communications technology to solutions that allow improving the agricultural production in climate change scenarios.

Finally, we would like to thank the hard work and dedication of the Program Committee members and Organizing Committee members. Thank you.

September 2017

Juan Carlos Corrales
José Antonio Iglesias
Plamen Angelov
Conference General Chairs

Organization

AACC 2017 was organized by Red Insterinstitucional de Cambio Climático y Seguridad Alimentaria de Colombia—RICCLISA, Departamento Administrativo de Ciencia, Tecnología e Innovación COLCIENCIAS—Colombia, Universidad del Cauca—Colombia, Universidad Carlos III de Madrid—España, and Centro Regional de Productividad e Innovación del Cauca CREPIC—Colombia.

General Chairs

Juan Carlos Corrales	Universidad del Cauca
José Antonio Iglesias	Universidad Carlos III de Madrid
Plamen Angelov	Lancaster University

Program Co-chairs

Apolinar Figueroa	Universidad del Cauca
Agapito Ledezma	Universidad Carlos III de Madrid
Carlos Alberto Alves Meira	Empresa Brasileira de Pesquisa Agropecuária, EMBRAPA

Organizing Committee

Carlos Arturo León	RICCLISA
Rafael Bermudez	CREPIC
Sandra Patricia Rebolledo Acosta	CREPIC
Carolina Quiñonez	CREPIC

Program Committee

Wuletawu Abera	Centro Internacional de Agricultura Tropical (CIAT)
Plamen Angelov	Lancaster University
Jayme Barbedo	Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)
Sandro Bimonte	Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture (IRSTEA)
Oscar Caicedo	Universidad del Cauca
David Camilo Corrales	Universidad Carlos III de Madrid
Juan Carlos Corrales	Universidad del Cauca
Alexandre Coutinho	Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)
Gil De Sousa	Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture (IRSTEA)
Julio Esquerdo	Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)
Apolinar Figueroa	Universidad del Cauca
Christian Figueroa	Universidad Santiago de Cali
Aryeverton Fortes de Oliveira	Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)
German Gutierrez	Universidad Carlos III de Madrid
Jorge Gómez	Universidad de Córdoba
Jose Antonio Iglesias	Universidad Carlos III de Madrid
Emmanuel Lasso	Universidad del Cauca
Agapito Ledezma	Universidad Carlos III de Madrid
Diego López	Universidad del Cauca
Iván López	Universidad del Cauca
Carlos Meira	Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)
Juan Mendoza	Universidad Santo Tomás Seccional Tunja
Jordi Morato	Universidad Politécnica de Catalunya
Hugo Armando Ordóñez	Universidad de San Buenaventura - Cali
José Armando Ordóñez	Fundación Universitaria de Popayán
Mario Milver Patiño Velasco	Universidad del Cauca
Leonairo Pencue	Universidad del Cauca
Edier Humberto Perez	Universidad del Cauca
Francois Pinet	Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture (IRSTEA)
Gustavo Ramirez Gonzalez	Universidad del Cauca

Alvaro Rendon	Universidad del Cauca
Araceli Sanchis	Universidad Carlos III de Madrid
Luz Santamaría	Universidad Santo Tomás Seccional Tunja
M. Paz Sesmero Lorente	Universidad Carlos III de Madrid
Fernando Aparicio Urbano	Universidad del Cauca

Additional Reviewers

Eraso, Luis
Estrada-Solano, Felipe
Perez, Edier Humberto
Rodriguez-Vivas, Angela
Simmonds, Jose
Solano Correa, Yady Tatiana
Vivas, Fulvio

Sponsors

Red Insterinstitucional de Cambio Climático y Seguridad Alimentaria (RICCLISA)

Universidad del Cauca

Grupo de Ingeniería Telemática, Universidad del Cauca

Colciencias

REPÚBLICA DE COLOMBIA
COLCIENCIAS
Ciencia, Tecnología e Innovación

Crepic

CREPIC
CENTRO REGIONAL DE PRODUCTIVIDAD
E INNOVACIÓN DEL CAUCA

Universidad Carlos III de Madrid

UNIVERSIDAD CARLOS III DE MADRID
homo homini sacra res
Universidad
Carlos III de Madrid

Contents

Knowledge Inference from a Small Water Quality Dataset with Multivariate Statistics and Data-Mining

Jose Simmonds[1(✉)], Juan A. Gómez[2], and Agapito Ledezma[1]

[1] Departamento de Informática, Universidad Carlos III de Madrid,
Avenida de la Universidad 30, 28911 Leganés, Spain
jose.simmonds@alumnos.uc3m.es, ledezma@inf.uc3m.es
[2] Departamento de Biología Marina, Universidad de Panamá,
Estafeta Universitaria, Apartado 3366, Panamá 4, Panamá
juanay05@hotmail.com

Abstract. Multivariate analysis (MV) and data mining (DM) techniques were applied to a small water quality dataset obtained from the surface waters at three water quality monitoring stations in the Petaquilla River Basin, Panama, during the hydrological period of 2008 through 2011 for the assessment and understanding of the ongoing environmental stress within the river basin. From Factor Analysis (PCA/FA), results indicated that the factors which changed the quality of the water for the two seasons differed. During the dry (low flows) season, water quality showed to be strongly influenced by turbidity (NTU) and total suspended solids (TSS) concentrations. In contrast, during the wet (high flows) season the main changes on water quality sources were characterized by an inverse relation of NTU and TSS with the electrical conductivity (EC) and chlorides (CL), followed by significant sources of agricultural pollution. To complement the MV analysis, DM techniques like cluster analysis (CA) and classification (CLA) was applied to the data. Cluster analysis was used to separate the stations based on their levels of pollution and the classification of stations was implemented by C5.0 algorithm to classify stations of unknown origin into one of the several known groups of water quality constituents. The study demonstrated that the major water pollution threats to the Petaquilla River Basin are industrial and urban development in character and uses of agricultural and grazing land which are defined as non-point sources. The use of DM techniques was to complement the MV analysis. Taking into account the limited data, the usage of these methodologies is regarded useful in aiding water managers for implementing water monitoring campaigns and in setting priorities for improving and protecting water quality sources that are impaired due to land disturbances from anthropogenic activities.

Keywords: Multivariate · Cluster analysis · Classification · Decision tree · Water quality

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_1

1 Introduction

Minera Panamá S.A. (MPSA), wholly owned by Minera Panama S.A-First Quantum Minerals Ltd (MPSA-FQML), is investigating the feasibility of developing the MPSA Project Mina de Cobre Panamá (the Project). The proposed Project would mine and process copper sulfide ore in the Petaquilla Concession, Panamá. This concession covers an area of 130 square kilometers (km^2) and is located in the District of Donoso, Colón Province, in north-central Panamá. The concession contains at least three spatially distinct copper ore bodies (Colina, Botija and Valle Grande) and three conventional open pit mines are currently planned to exploit these ore bodies [1].

The copper sulfide ore will be mined using conventional open pit mining and will be processed using crushing, milling, flotation recovery and concentrate dewatering. The proposed design ore feed to the processing plant is 150,000 tons per day (t/d). It is expected that this will be expanded to 225,000 t/d at year ten by the addition of a third processing line. The Project will export materials through a port site to be constructed on the Caribbean coast at Punta Rincón and linked to the main Project site by a road, a power line corridor, and buried pipelines for transfer of products and other materials. As the nation develops, increasing industrialization and urbanization has led to a wide-scale contamination of many surface water resources from industrial effluents, domestic sewage discharges, and excessive use of fertilizers, pesticides and the emerging mining activities. Then, it may be inferred that the increased anthropogenic pressures and natural processes are accounting for degradation in surface water and groundwater quality [2]. Hence, given these pressures experienced on the water resources in the area, the main objectives of conservation must be in the control and minimization of pollution occurrences and problems facing these pollutants and to provide water of an adequate quality that can serve different purposes, such as drinking water, irrigation water [3, 4]. Then, the monitoring of water quality for any water body must be one of the highest priorities for their protection policy [5].

Giving that the physical and chemical characteristics of water bodies affect species composition, abundance, productivity and physiological conditions of aquatic organisms [6], water quality monitoring must be of highest priorities in any environmental protection policy [7].

Multivariate statistical methods such as factor analysis and principal components have been used successfully in hydrochemistry for many years. Nowadays, with the emerging technique offered by data mining techniques, the water quality of a given river state can reveal features otherwise not seen by conventional methods. The evaluation of the quality of surface water, groundwater and environmental assessment techniques use multiple components that are well defined in the literature [8, 9]. Multivariate techniques allow us to discover the information hidden in the data set about the possible environmental influences on water quality [10]. Today, data mining is popular among researchers of water quality investigations, for example in regard to chlorophyll levels researchers in [11] proposed Decision-making tree to forecast levels for the next day. Also in [12], it was proposed the use of fuzzy c-means clustering method to classify and assess rural surface water quality built on monitoring data from 33 water quality stations in 23 rural rivers and 4 reservoirs in Lianyungang city.

Multivariate methods have several shortcomings such as the presence of mathematical calculations, equal treatment and process to the old and new data, problems with prediction and classification task due to multivariate overlapping of the parameters. Notwithstanding, data mining and machine learning techniques have shown to achieve great success in many disciplines [13]. Nevertheless, it is a well-known fact that data mining algorithms work best on large data sets, there are several studies which encourages its application on small databases [14–16].

In this study, we evaluated the possibility that a smaller group of water quality parameters could provide sufficient information for assessing water quality. For this reason, Factor analysis and data mining methods were applied to water quality data obtained from the surface waters of three (3) water quality monitoring stations at the Petaquilla River Basin during two hydrological seasons (high and low flows). The first part of the evaluation of the data was done using the MATLAB 2011a Package, version 7.5 (March 18, 2011), (http://www.matworks.com/) for MAC OSX and the second part was evaluated by data mining approaches using the R statistical computing language [16]. The water quality record with 12 observations of 41 water quality parameters selected for estimation of the physical-chemical characteristics were: Total Alkalinity (T_Alk), Electrical Conductivity (EC), Oxygen Redox Potential (ORP), pH, Total Suspended Solids (TSS), Turbidity (NTU), Chlorides (Cl), Fluorides (F), Nitrogen-Nitrite (N_NO2), Nitrogen-Nitrate (N_NO3), Sulfate (SO4), Total Cyanide (CN_T), Weak Acid Dissociable Cyanide (CN_W), Ammonia (NH3), Aluminum (Al), Boron (B), Barium (Ba), Beryllium (Be), Bismuth (Bi), Calcium (Ca), Cobalt (Co), Chromium (Cr), Copper (Cu), Iron (Fe), Potassium (K), Lithium (Li), Magnesium (Mg), Manganese (Mn), Sodium (Na), Nickel (Ni), Phosphorus (P), Selenium (Se), Silicon (Si), Zirconium (Sr), Titanium (Ti), Vanadium (V), Zinc (Zn), Phenol (PHEN), Total Kjeldahl Nitrogen (TKN), Total Nitrogen (TN) and Dissolved Organic Carbon (DOC). Water samples were collected on site and physical-chemical analysis was performed in a local accredited laboratory according to the methodology contained in the Standard Methods for the Examination of Water and Wastewater 21st Ed 2005.

The paper is organized as follows. Section 2, describes the background of the overall study location. Section 3 illustrates the methodology that has been used to tackle the problem. The experimental setup and analysis results of the study are presented in Sect. 4. Finally, conclusions and future work are discussed in Sect. 5.

2 Background

2.1 Data Source and Study Area

The Petaquilla basin is the westernmost drainage basin at the mine site. Surface runoff in this basin reports to the Petaquilla River, where it subsequently flows northwest and discharges directly to the Caribbean Sea. Two open pits (Colina and Valle Grande), as well as the southwest waste rock storage facility and associated sedimentation ponds, will be developed in the southeast region of the Petaquilla basin. The community of Nueva Lucha and the Faldalito sector are also located in this basin. The three-surface water baseline sampling stations established in the Petaquilla River Basin, specifically in the Petaquilla River, are described in Table 1 and shown in Fig. 1.

Table 1. Petaquilla River Basin water quality stations and community location.

Sampling stations	River	Description
W1	Petaquilla River	Farthest upstream sampling location in the Petaquilla River Basin, upstream of the Colina open pit and associated southwest waste rock storage facility
W2	Petaquilla River	Downstream of W1 and the Colina open pit and associated southwest waste rock storage facility, near the sector of Faldalito
W10	Petaquilla River	Located about 6 km upstream of the river mouth and 8 km downstream of W2; this location was established to characterize baseline conditions downstream of proposed mine facilities and the community of Nueva Lucha

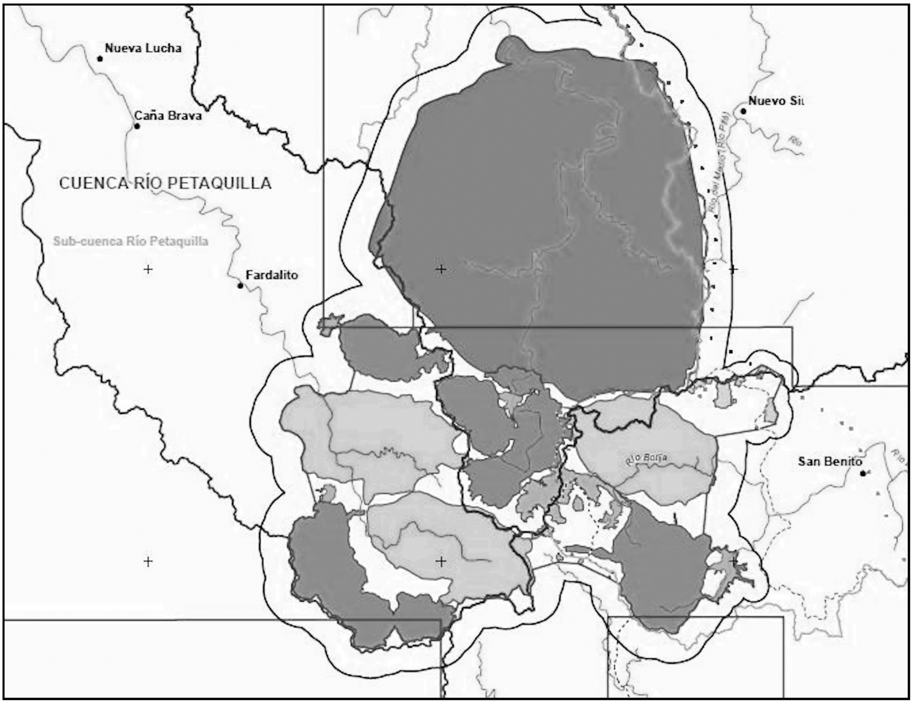


Fig. 1. Map showing the Petaquilla River Basin location. (Source: MPSA, ESIA, 2010)

3 Methods

The Isthmus of Panama has basically two seasons: the dry low flow season (January to April) and the high flow season (May to December). The climate in the region of the Petaquilla River Basin is typically governed by these two seasons. Therefore, the hydrological conditions during the low and high flow seasons are somewhat different. From there, the water quality was assessed separately for both seasons (periods of low flow conditions) and the wet season (period for high flow conditions) would assist in the understanding of the major pollutants of water quality sources and in establishing priorities for monitoring this water course during these periods.

Before we ran the experiments, it was necessary to impute missing data for the available small data record by means of PCA. No data is registered for the year 2010. The experiment was then divided into two parts. First, we applied principal components analysis (PCA) and factor analysis (FA) techniques to the 41 water quality parameters from the 12 surface water samples collected for both seasons, dry (January-April 2008) and rainy (May to December 2011, respectively) to orderly simplify the number of interrelated measures to a few representative constructs or factors [17]. From the PCA, FA, correlation and variance matrices, we chose 9 variables for the purposes of the experiment. On the one hand, and of particular notice, the FA left out the T_Alk, and by the other hand the PCA left out the N_NO2 variable. Then, completed this phase and given the scarcity of the data, and keeping in mind to potentiate the information within, the data was also treated with cluster analysis (CA) and classification analysis (CLA) with the purpose of discovering features not otherwise revealed by the PCA and FA.

The descriptive statistics of the data is presented in Tables 2 and 3, respectively.

Table 2. Descriptive statistics for water quality variables during low flow season

Variables	Units	N	Median	Mean	Std	Variance	CV	Min	Max
T_Alk	CaCO3 mg/L	12	10.78	10.78	2.40	5.78	0.22	7.00	16.00
EC	µS/cm	12	40.22	40.44	3.25	10.57	0.08	37.00	50.00
ORP	mV	12	154.67	146.33	54.41	2960.18	0.37	40.00	212.00
pH	–	12	6.92	6.92	0.17	0.03	0.02	6.73	7.29
TSS	mg/L	12	1.50	2.78	2.68	7.19	0.97	1.00	10.50
NTU	NTU	12	2.37	2.37	1.53	2.33	0.64	0.50	5.00
Cl	mg/L	12	3.46	3.51	0.50	0.25	0.14	2.90	4.59
F	mg/L	12	0.03	0.07	0.09	0.01	1.30	0.01	0.25
N_NO2	mg/L	12	0.00	0.00	0.00	0.00	0.26	0.00	0.00

Table 3. Descriptive statistics for water quality variables during high flow season.

Variables	Units	N	Median	Mean	Std	Variance	CV	Min	Max
T_Alk	CaCO3 mg/L	12	10.10	10.10	3.07	9.41	0.30	5.00	16.80
EC	µS/cm	12	37.44	37.44	7.89	62.20	0.21	19.00	47.00
ORP	mV	12	236.50	238.00	41.48	1720.91	0.17	186.00	327.00
pH	–	12	6.89	6.89	0.28	0.08	0.04	6.41	7.30
TSS	mg/L	12	5.50	32.78	69.29	4801.41	2.11	0.50	248.00
NTU	NTU	12	9.50	11.48	14.13	199.66	1.23	1.00	52.50
Cl	mg/L	12	2.71	2.71	0.54	0.29	0.20	1.70	3.40
F	mg/L	12	0.06	0.09	0.10	0.01	1.08	0.01	0.25
N_NO2	mg/L	12	0.00	0.00	0.00	0.00	0.29	0.00	0.00

4 Experimental Setup and Results

4.1 Assessment of the Water Quality Profile with Principal Components and Factor Analysis

Factor Analysis: Low flow season

The correlation matrix of the selected variables was generated and the factors extracted and rotated using the Promax criterion, a common oblique method which has the advantage of being fast and conceptually simple [18]. Eigenvalues, the percentage of total variance, the factor loadings and cumulative variance are in Tables 4 and 5 respectively.

Table 4. Matrix of factor loads and total variance explained for low flow season.

Variables	Factor		
	1	2	3
EC	0.2068	**0.6691**	0.2135
ORP	**0.9130**	−0.4014	0.0819
pH	0.0344	−0.0218	**0.9923**
TSS	**1.0114**	−0.1805	−0.0678
Turbidity (NTU)	−0.1357	**0.6195**	0.0100
Cl	**0.7206**	−0.0407	0.4287
F	0.2681	**1.0046**	−0.1179
N_NO2	0.1225	0.1020	0.0207
Eigenvalue	2.53	2.05	1.24
% TOTAL VAR	0.32	0.26	0.16
Cumulative %	0.34	0.60	0.76

Table 5. Matrix of factor loads and total variance explained for high flow season.

Variables	Factor		
	1	2	3
EC	**0.9580**	−0.0516	−0.1202
ORP	0.1723	0.5573	0.0500
pH	0.3238	**0.6885**	−0.4567
TSS	−0.5698	0.1198	**0.6456**
Turbidity (NTU)	**−0.7847**	0.2208	0.4273
Cl	**1.0529**	−0.0839	0.1268
F	0.0794	0.4725	0.4963
N_NO2	−0.2851	**1.0552**	−0.0156
Eigenvalue	3.19	2.19	1.09
% TOTAL VAR	0.40	0.27	0.14
Cumulative %	0.34	0.61	0.75

The FA for the dry season generated three (3) important factors (with eigenvalues of 2.53, 2.05 and 1.24 respectively) and this solution explaining 76% of the total variance in the data set for surface waters is kept for rotation. From the analysis, the following factors were indicators of the physical-chemical profiles of the water quality:

- Factor 1: TSS, ORP, Cl^-,
- Factor 2: F^-, EC, NTU,
- Factor 3: pH.

The first factor accounted for 32% of the explained variance. TSS and CL^- recorded high positive loadings (1.0114 and 0.7206 respectively). ORP showed a strong negative loading (−0.9130), meaning that they were inversely correlated with this factor.

Water bodies may vary in their physical-chemical composition and concentration on a seasonal, diurnal or even hourly basis. These variations may be related to patterns of land and water use and precipitation [19, 20].

Land uses such as urbanization and industrialization influences the water cycle through changes in flow and water quality. Urban land use (Na+, F^-, Cl^-) may be differentiated from other land disturbances such as agricultural runoff (F^-, EC, NTU). Subsurface drainage water would commonly include Salts such as sulfates, chlorides, carbonates, and bicarbonates of calcium, and magnesium. Excess surface water draining also may contain these salts, but generally in much lower concentrations than in drainage water [21]. Based on the results of the factor analysis and typical sources of water pollutants, it is concluded that factor 1 can be denoted as the erosion processes factor with the presence of TSS, ORP and Cl. Negative factor loading of ORP explained the disproportion between this parameter and Factor 1. Based on the factor analysis, these parameters are usually associated with masses of large land movements and disturbances on the levels of sediments in the study area. This is because TSS as the highest loadings (1.0114).

Factor 2 is strongly correlated with F^{-}, EC and NTU, assigned as the agricultural processes factor. Factor loadings were 1.0046, 0.6691 and 0.6195. pH is included in Factor 3 and the indicator of biochemical processes in water.

In summary, three factors representing three different processes are:

- Erosion processes
- Agricultural processes
- Biochemical processes

Hence, the water quality of the Petaquilla River Basin during the dry season period was mainly controlled by activities which trigger erosion processes.

Factor 1: High factor scores observed for TSS (erosion processes factor) Giving the presence of an already established gold mine, within the study area, illegal Artisanal Small-Scale Miners use domestic gravel pumping and flushing devices in their activities. The use of these devices in gold mining by the Artisanal Small-Scale Miners constitutes a point source of contamination by Suspended Solids.

Factor 2: High factor scores (agricultural processes factor) were observed in the west of the basin. The middle and eastern parts where high values were monitored were faced with pollution risks originating from agricultural uses.

Factor 3: (Biochemical processes factor) pH values were distributed in the basin almost uniformly and are indicative of decay. The mean value was 6.89.

Factor Analysis: High flow season

The wet season period may influence positive effects with a dilution of surface water by rain and stormwater. In contrast, runoff water increases pollutant concentrations, thereby decreases water quality. To assess the water quality of the Petaquilla River Basin under rainy conditions, factor analysis was applied to the datasets obtained from 3 monitoring stations between May-December. Statistical analyses of the data are shown in Table 3.

Results of the factor analysis including factor-loading matrix, eigenvalues, total and cumulative variance values are given in Table 5.

From the Factor Analysis, three factors that are indicated below explained 75% of total variance.

- Factor 1: Cl^{-}, EC, NTU,
- Factor 2: N_NO2, pH-,
- Factor 3: TSS.

It is proposed that factor 1 represents the agricultural processes shown by the presence of Cl^{-} and EC, with the NTUs indicating an inverse relation for that factor. This factor explained 40% of the variance. Factor 2 is strongly correlated with N_NO2 and pH which is mainly influenced by biochemical processes. Factor 3 was marked by TSS. Thus, agricultural processes were the major pollution source for this hydrological period.

Considering the distribution of factor scores and locations of the monitoring stations, it is concluded that:

Factor 1: High factor scores (agricultural processes factor) were observed at the southwest part, upstream of the basin.

Factor 2: Relatively high values of biochemical processes obtained throughout the entire basin where some agriculture and grazing of cattle are most important economic activities.

Factor 3: Significant scores of total suspended solids pollution were distributed in the basin, because Factor 3 depends on point pollution sources and is affected by eroded soils in the study area

4.2 Assessment of the Water Quality Profile with Cluster Analysis and Decision Tree

Spatial Similarity and Stations Grouping: Low flow season

The experimental approach involved the use of hierarchical cluster analysis (CA) with Euclidean distance and linkage method using ward.D2, from the R Statistical Package on the scaled data formed with data sorted by two seasons (dry and wet). For a clearer understanding of the stations' representation with respect to the date of sampling campaigns, we systematically arranged the stations with four levels with the purpose to identify specifically to which date corresponded the collection of the samples. So, station W-1 as the following labels (W-1.1 to W-1.4), station W-2 (W-2.1 to W-2.4) and finally W-10 (W-10.1 to W-10.4), such that each level corresponds to a sampling date (01-Jan-08, 01-Feb-09, 11-Jan-11, 05-Apr-11, 01-Jan-08, 01-Feb-09, 11-Jan-11, 05-Apr-11, 01-Jan-08, 01-Feb-09, 11-Jan-11 and 05-Apr-11 respectively). Temporal CA for the dry period generated a dendrogram as shown in Fig. 2, grouping 3 combined water quality stations into three clusters. Cluster I comprised Stations W-1 and W-2 and the cluster II included Stations W-1, W-2 and W-10, while the cluster III grouped all three stations in one for the various consecutive periods. The cluster III, shows similarities in the water quality for the respective sampling period between February 2009, January 2011 and April 2011. Figure 2 shows that the spatial patterns of the water quality are not purely consistent with the dry season. From the monitoring campaigns, at Stations W-1 and W-2, the highest pollution levels for (TSS = 5.0 and 10.5 mg/l, respectively) were recorded in February 2009. In cluster II, for stations W-1, W-2 and W-10, W-1 had the highest value of (T_Alk = 13 CaCO3 mg/l) in January 2008. Aside from the eroded land scape, these high levels of TSS recorded in the dry season can be probably attributed to the location of Station W-1 been directly in the project foot print and W-2 been downstream of it. Notwithstanding, it is also reported the activities of artisanal mining by peasants in the area of station W-2, as was previously mentioned. With an increase of the TSS, it is more likely that the concentration of other water quality constituents like chlorides (Cl^{-}) to be present as was depicted from the factor loadings in Table 4.

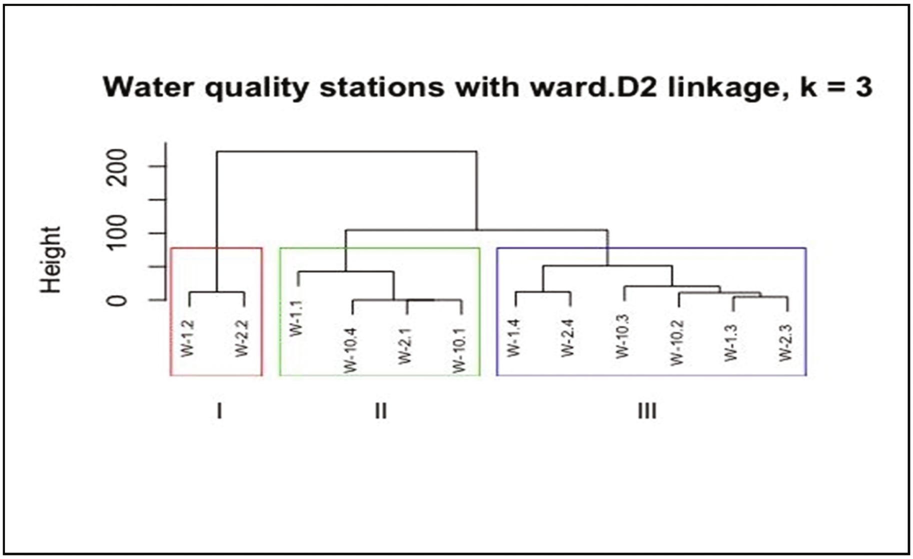


Fig. 2. Cluster analysis on sampling stations along Petaquilla River during low flow season.

Spatial Similarity and Stations Grouping: High flow season

For the wet season period, the information depicted in Fig. 3, shows the grouping of 3 clusters, been the third one the largest group. In cluster I, station W-1 is isolated from the other stations, cluster II groups Stations W-1 and W-2 and cluster III all three stations. As with the dry season, the systematic arrangement of the stations with respect

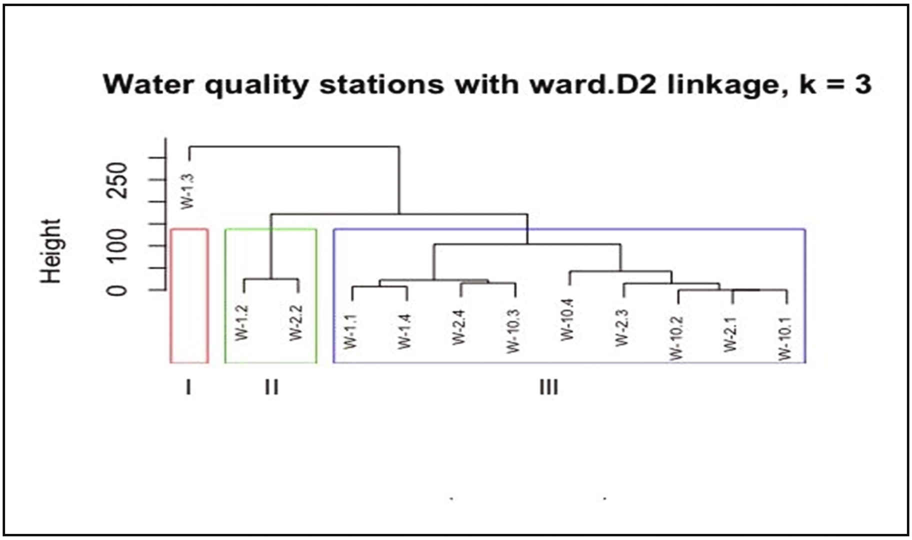


Fig. 3. Cluster analysis on sampling stations along Petaquilla River during high flow season.

to the sampling dates or as follows: 1-Sep-08, 1-Oct-08, 1-Nov-08, 13-Oct-10, 1-Jun-08, 1-Oct-08, 1-Nov-08 27-Sep-10, 1-Jun-08, 1-Sep-08, 13-Oct-10 and 15-Jun-11. Again, Station W-1 isolated by cluster I for November 2008, shows to be the most polluted in terms of the (TSS = 248 mg/l). Cluster II shows high pollution of (Cl^{-} = 3.36 and 3.33 mg/l, respectively) at Stations W-1 and W2 in October 2008. Cluster III shows lower levels compared to the other two clusters. From the FA loadings in Table 4, it is observed the pollution of the water constituents that follows the high levels of eroded soils during high water flow seasons (agricultural and biochemical) and is reflected in all three stations for the sampling campaign.

The results showed that CA technique can be useful in classification of the river water [22] in the study region and was able to disclose the sampling sites and its associated monitoring periods. in spite of the short database, the small data record was used without missing much information.

Decision Tree: Low flow season

Decision trees are one of the most widely used inductive inference algorithms and it is a kind of approaching discrete value function method which is good and robust for dealing with noise and can analyze and extract expressions. Decision trees are usually generated from top to bottom and its routines procedures use greedy search traversal method to traverse. Every decision or event may extract two or more events and cause different results. This kind of decision branches can be depicted in figures which resemble the branching of trees, reason why it is called decision tree. The decision tree algorithms commonly used include CART, ID3, C4.5 AND C5.0 algorithms [23–25].

In this section of the experiment, the objective was to test the applicability of decision trees to the limited data set and find rules that could have just as well provide information that leads to unseen patterns in the data and also corroborates the other analysis carried out previously. With the R package for data mining, after serval trial and errors with the Party, RPart, Random Forest and the C5.0 algorithms, a classifier was finally built with the C5.0 algorithm. The C5.0 was feasible to deal with the twelve observations, we could only go as far as dividing the data in half for training and testing, as other combination approaches of data splitting did not suffice for the task. Below in Fig. 4 the output tree for the dry season period representation.

As shown in Fig. 4 the decision tree C5.0 model, which ranked the water quality of the three stations according to the content of two key indicators (T_Alk and NT) generated the following equations of water quality: when T_Alk <= 8: W-1, when T_Alk > 8: and the NTU > 3: W-1, when NTU <= 3: and NTU <= 1: W-2 and NTU > 1: W-10.

The design of this tree model was built by the sampled data of the 12 observations. The performance error rate was 16.7%. The other metrics of the model are shown in Fig. 4. The prediction with the test data output, suggested a model performance of 83% with only two miss classified instances for station W-2.

The information provided by the decision tree corroborates the information given previously by the FA and the CA analysis in regard to impairments of the water quality profile at stations W-1, W-2 and W-10 with respect to pollution by high levels of TSS and on the other hand, it reveals the direct effects of the TSS on the water column as a consequence of the increased NTU levels.

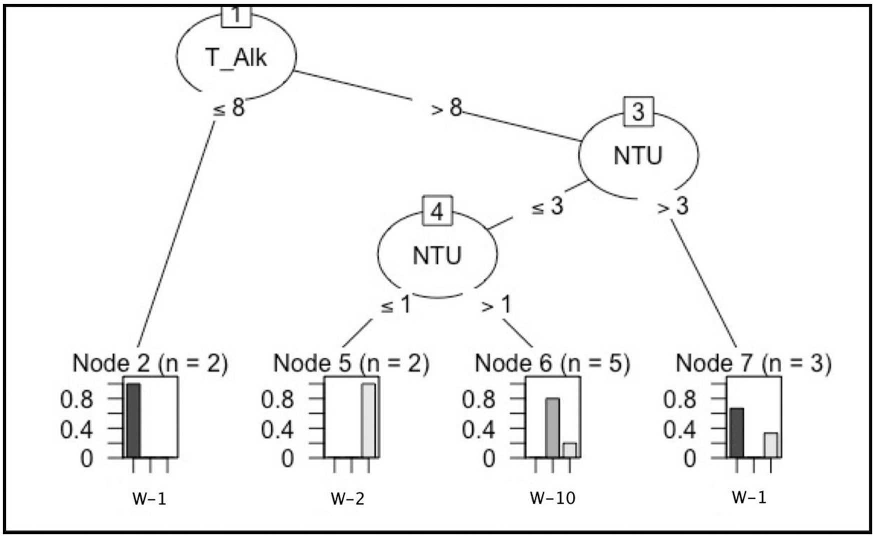


Fig. 4. Decision tree generated by C5.0 for the water quality stations during dry season period.

Decision tree: High flow season

Conditions of the water quality under the wet season as labeled by the decision tree is shown in Fig. 5. As can be seen, three water quality indicators (EC, Cl and TSS) in this case are responsible for the governing pattern of the water quality profile for Stations

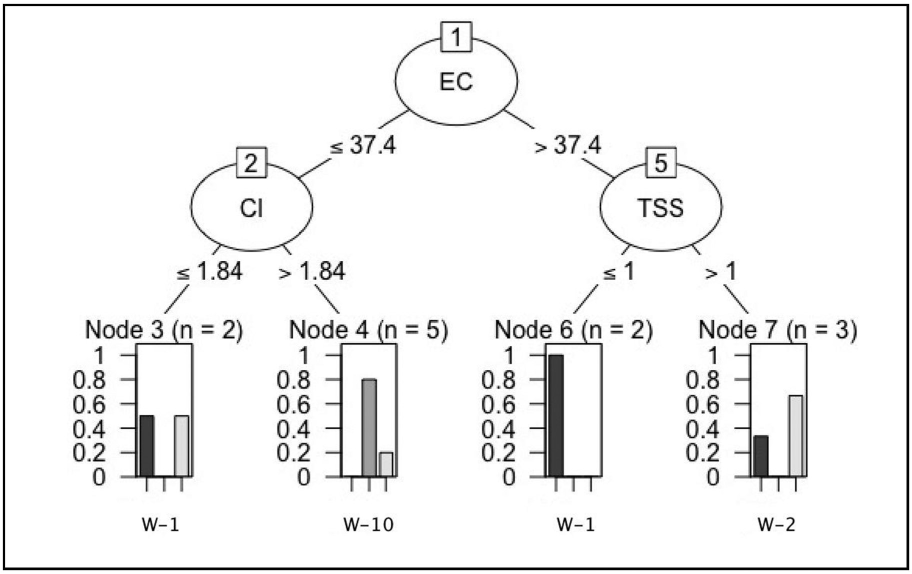


Fig. 5. Decision tree generated by C5.0 for the water quality stations during wet season period.

W-1, W-2 and W-10 respectively. The water quality equations generated for this scenario are: when EC <= 37.4 and Cl <= 1.84: W-1, and Cl > 1.84: W-10; when EC > 37.4 and TSS <= 1: W-1 and TSS > 1: W-2. The error rate for the performance of this model is 25%, 0.9% more than the model built under dry season condition. From the confusion matrix, this model classified correctly three instances out of four for Station W-1, all four instances for Station W-10 were classified correctly and only two for Station W-2 out of four were classified correctly. The prediction of the model of the classes a 75% accuracy.

5 Conclusions and Future Work

In this study, the application of multivariate techniques demonstrated that the water quality of the Petaquilla River Basin is under environmental stress.

The factors obtained from the FA indicated that the water quality for the two distinct hydrological periods and locations differed slightly. From the FA, Factor model 1 is interpreted to be of water affected by erosion processes, in spite of low flow periods. This may be inferred to illegal Artisanal Small-Scale miners who in their search for gold in the river bed, employs gravel pumping and flushing devices to dredge the river bed. Added to this, is the presence of an abandoned gold mining firm and recently to ongoing land movement activities in the study area. On the other hand, during the wet season, water quality was controlled by agricultural pollutant sources, this was explained by 40% of the total variance and for the dry season conditions, it was under only by 26%. So, the major pollutant source changed from erosion to agricultural processes during the wet season. This is attributed to the negative effects of runoff to surface water quality.

Another contribution to this study has been achieved by complementing the multivariate approach with data mining techniques. The results of this study unveiled the importance and usefulness of cluster analysis of a small and complex database to obtain better information concerning the surface water quality. Hierarchical CA grouped the 12 observed sampling campaigns into three clusters and classified 3 sampling sites based on the similarity sites of water quality parameters. The temporal pattern shows that February and November have high pollution levels in contrast with the rest of the months. The spatial pattern shows that the sampling sites (W-1 and W-2) have highest levels of pollution. We also used the water quality data with the C5.0 algorithm found in R data mining statistical tool. The experiment with the 12 observations and nine attributes of water quality data revealed interesting information with the built decision tree. See Sect. 4.2 where it is shown the If…then rules extracted and the generated trees. This information is enriching, which otherwise would not have been gained solely from a PCA/FA only.

It is obvious from this study, that major pollution threats during the dry and wet seasons were erosion and agricultural processes which are defined as nonpoint pollution sources. This situation suggests that it is imperative to establish an environmental monitoring and assessment plan to promote minimization of these sources as to protect and improve water quality in the basin.

The above analysis demonstrates the use of multivariate statistics and complementation with data mining techniques to study a small database of physical-chemical parameters in surface water systems. It's a useful methodology that could aid in decision making as to determine the extent and severity of pollution via practical pollution indicators. It could provide guidelines for selecting the priorities of keen preventative measures in the proper management of the surface water resources of the basin. Future work should involve fetching new and current data, and to explore or/and implement new algorithms to help foster the analysis of small databases of river basins.

Acknowledgement. The authors of this experiment will like to express their appreciation to Minera Panama S.A., Environmental Department for providing the necessary data. This work has been partially supported by the Spanish MICINN under projects: TRA2015–63708-R, and TRA2016-78886-C3-1-R.

References

1. Panama, M.: Environmental Impact Assessment Study (2010)
2. Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N., Smith, V.H.: Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol. Appl. **8**(3), 559–568 (1998)
3. Wetzel, R.G.: Gradient-dominated ecosystems: sources and regulatory functions of dissolved organic matter in freshwater ecosystems. Hydrobiologia **229**(1), 181–198 (1992)
4. Dinar, A., ed.: Restoring and Protecting the World's Lakes and Reservoirs, vol. 289. World Bank Publications (1995)
5. Lewis, W.M.: Basis for the protection and management of tropical lakes. Lakes Reserv. Res. Manage. **5**(1), 35–48 (2000)
6. Bagenal, T.B.: Fecundity in eggs and early life history (Bagenal, T.B., Braum, E Part 1). In: Bagenal, T.B. (ed.) Methods for Assessment of Fish Production in Freshwaters, 3rd edn. pp. 166–178 (1978)
7. Simeonov, V., Einax, J.W., Stanimirova, I., Kraft, J.: Environmetric modeling and interpretation of river water monitoring data. Anal. Bional. Chem. **374**(5), 898–905 (2002)
8. Praus, P.: Water quality assessment using SVD-based principal component analysis of hydrological data. Water SA **31**(4), 417–422 (2005)
9. Jayakumar, R., Siraz, L.: Factor analysis in hydrogeochemistry of coastal aquifers–a preliminary study. Environ. Geol. **31**(3-4), 174–177 (1997)
10. Spanos, T., Simeonov, V., Stratis, J., Xristina, X.: Assessment of water quality for human consumption. Microchim. Acta **141**(1), 35–40 (2003)
11. Lu, J., Huang, T.: Data mining on forecast raw water quality from online monitoring station based on decision-making tree. In: Fifth International Joint Conference on INC, IMS and IDC, NCM 2009, pp. 706–709. IEEE (2009)
12. Fu-Cheng, L., Xue-Zhao, H.: Application of fuzzy c-means clustering for assessing rural surface water quality in Lianyungang City. In: 2013 Fifth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), pp. 291–295. IEEE (2013)
13. Mjolsness, E., DeCoste, D.: Machine learning for science: state of the art and future prospects. Science **293**(5537), 2051–2055 (2001)

14. Jiang, Y., Li, M., Zhou, Z.-H.: Mining extremely small data sets with application to software reuse. Softw. Pract. Exper. **39**(4), 423–440 (2009). https://doi.org/10.1002/spe.905
15. Andonie, R.: Extreme data mining: inference from small datasets. Int. J. Comput. Commun. Control **5**(3), 280–291 (2010)
16. Natek, S., Zwilling, M.: Student data mining solution–knowledge management system related to higher education institutions. Expert Syst. Appl. **41**(14), 6400–6407 (2014)
17. R Core Team: A language and environment for statistical computing. R Foundation for Statistical Computing Department of Agronomy, Faculty of Agriculture of the University of the Free State. Vienna, Austria (2017). https://www.R-project.org/
18. Hendrickson, A.E., White, P.O.: Promax: a quick method for rotation to oblique simple structure. Br. J. Stat. Psychol. **17**, 65–70 (1964)
19. Ho, R.: Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS. CRC Press (2006)
20. Abel, P.D.: Water pollution biology. CRC Press (1996)
21. Ayoade, A.A., Fagade, S.O., Adebisi, A.A.: Dynamics of limnological features of two man-made lakes in relation to fish production. Afr. J. Biotechnol. **5**(10), 1013–1021 (2006)
22. Fataei, E., Shiralipoor, S.: Evaluation of surface water quality using cluster analysis: a case study. World J. Fish Mar. Sci. **3**, 366–370 (2011)
23. Areerachakul, S., Sanguansintukal, S.: Classification and regression trees and MLP neural network to classify water quality of canals in Bangkok, Thailand. Int. J. Intell. Comput. Res. (IJICR) **1**(1/2), 43–50 (2010)
24. Quinlan, J.R.: Induction of Decision Trees. Mach. Learn. **1**(1), 81–106 (1986)
25. Salzberg, S.L.: C4.5: programs for machine learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., Mach. Learn. **16**(3), 235–240 (1993)

A Multiscale Based Rainfall Amount Prediction Using Multiple Classifier System

Cristian Valencia-Payan(✉) and Juan Carlos Corrales

Universidad del Cauca, Popayán, Colombia
{chpayan,jcorral}@unicauca.edu.co

Abstract. Rain is vital for environmental and human processes such as temperature regulation, vegetation growth, agriculture, power generation, domestic use, and others. However, most of the previous research has focused mainly on the prediction of the probability of occurrence of rainfall or the rain rate estimation, leaving aside the amount of rain. Most of the previous studies use information on the same spatial scale, losing important information on other scales that affect rain-related climatological processes. In this paper, we present an approximation for estimating daily rainfall using data from Meteoblue, GOES, TRMM and MODIS Vegetation indices on different spatial scales based on Bagging with Random Forest.

Keywords: Rainfall · Machine Learning · Bagging · Stacking · Random Forest · Multiscale data

1 Introduction

Rain forecast is one of the most challenging tasks in meteorology due to all the factors that have the potential to affect the rain processes. These factors are related not only to the rain, but to each other. Rain forecast can be divided into three questions: where is it going to rain, when is it going to rain, and how much rain will fall? [1]. Each question is relevant, depending on the goal of the forecast itself and data availability. Forecasting the amount of rain has vital importance in areas such as the High Colombian Andes where the rain affects most of the country's population. The amount of rainfall over the Andes affects the levels on two of the most important rivers in Colombia; the Cauca river and the Magdalena river that together support more than 80% of the population as well as productive activities such as agriculture, fish farming, domestic use, and more.

Climatological databases have a great volume and variety of data stored in multiple formats from plain text files to Geo-referenced files. However, the forecast objective is limited by the data itself and the processing capacity. For this reason, the Bigdata Value Chain must be applied, especially Bigdata Analytics. This encompasses data mining processes, statistical analysis, and data modeling to extract value [2], to facilitate decision making and predictions of future events based on the database analyzed. In recent years, the authors of [3–6] and others have focused on the forecast of rain events, the probability of

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_2

rain, the expected rain rate or the interval of an expected amount of rain during storms, leaving aside the amount of rain under normal rain conditions.

Although some of these investigations have used data from various spatial scales, in many cases this data is restricted simply to corroborating the data generated by the model: they are not used directly in the forecast. This paper evaluates multiple machine learning approaches to estimate the daily amount of rain using multiple sources of data on different spatial scales based on machine learning (ML) algorithms.

This paper is organized as follows: the first part outlines studies related to the main topics addressed around rain forecasting using ML; subsequently, ML architectures are presented and explained and data sources for testing are described; ML architectures are validated by experimental evaluation results; and finally, the conclusions of this research are presented.

2 Related Work

The study carried out in [4] presents an approach for rainfall forecasting with ten years of weather station data, using Regression and Neural Network methods. Results show potential in estimating the daily amount of rainfall but fail to reproduce data where peaks of rain are present. In [6] a model for using satellite and rain gauge data was presented, but this model uses only the information about rain rate, neglecting to consider information about other climatological variables that affect rain. Meanwhile, in [3], predictive capacity of meteorological data is evaluated using multiple machine learning methods to predict the range of the amount of rain expected the next day using information not only on rain but on temperature and wind.

In [7] the use of Genetic Algorithms (GA) is proposed to select the attributes related to rain and reduce the amount of data used in the forecast, but the authors do not consider data from multiple meteorological sources. The study presented in [8] shows a Deep Learning approach to forecast the daily amount of rain based on the historical data collected from a weather station. Elsewhere, [9, 10] propose the use of Goddard Cumulus Ensemble to estimate rain rates using meteorological information on a multiple scale, but this model requires a lot of data not available at our study zone. Most of the related studies have machine learning methods involved in the forecasting process. In every case, there is no multiscale data fusion, or the use of multi-spatial scale data is limited to validation only.

According to the prior literature review, it is important to consider multi-spatial scale data, to obtain rain estimates with a precision comparable to those using rain gauge data, but with a wider range. Moreover, no studies using multiple climatic data at multiple spatial scales were found. The closest studies to this proposal use only rain information, and have not been applied in the High Colombian Andes. This proposal is important because multi-spatial scale data will allow us to detect local rain events that have a great impact on water resource generation, an important topic for watershed management authorities and government entities. These local rain events are not detectable on low spatial scale data without losing the ability to forecast rain in large areas at the same time.

3 Materials and Methods

Machine learning was developed to learn and perform specific tasks [11] for example to predict the amount of daily rain based on multiscale data. The predictor should be able to estimate the total amount of rain based on parameters such as temperature, rainfall rate, and humidity, among others. This task is formally called regression. The regressors used in this study are described below.

3.1 Predictors

Multilayer Perceptron (MLP)

Artificial Neural Networks (ANN) are inspired by the way biological nervous systems work [12]. ANN is a machine learning architecture and algorithm used to extract patterns, detect tendencies or perform data classification. The artificial neuron has multiple inputs that can trigger the neuron according to the pattern to which it was trained to respond and the trigger function. There exist multiple configurations of these networks using different training methods and activation functions [12]. MLP is a feedforward ANN model that maps a set of input data onto an appropriate output [12]. MLP utilizes a supervised learning technique called backpropagation for training.

Bagging

This is a machine learning ensemble meta-algorithm that attempts to reduce variance and helps to avoid overfitting. To generate different base learners, bagging adopts bootstrap distribution [13]. This is a method for generating multiple versions of the same predictor, using these versions to create an aggregated predictor [14]. These multiple versions are created using bootstrap replicates of the training dataset and using it as new training dataset.

Stacking

Stacking is a scheme for minimizing the generalization error rate for one or more classifier. It works by deducing the biases of the classifiers with respect to a provided learning dataset [15]. It is a different way of combining multiple models. Unlike bagging, stacking is used to combine different models. To achieve its goal, the stacking algorithm splits the training dataset in two: one is used to train the base models and the second one to test these models. At the end, using the outputs of these base models as inputs and the correct ones as outputs, a higher-level model is trained.

Random Forest (RF)

This ensemble learning method is the combination of tree predictors. Each one depends on the values of a random vector of samples obtained independently from the training samples and with the same distribution for all the trees [16]. It consists of a collection of tree-structured classifiers or regressors, outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

3.2 Data Collection

In this study, different multiscale climatological databases were pre-processed to estimate the amount of daily rain. To achieve a predictor strong enough to estimate with the highest accuracy using multiscale data, two free access and one non-free access databases were used as training data. Data sources are as follows.

TRMM data
The Tropical Rainfall Measuring Mission (TRMM) is a joint NASA and JAXA mission launched in 1997 to study rainfall [17]. The TRMM product used in this study was the 3B42 3-Hourly Rainfall Data with a spatial resolution of 25 km by 25 km in a global belt extend from 50° South to 50° North latitude [18] from 2001 to the present. The 3B42 estimates are produced in four stages described below [19]:

- Microwave precipitation estimates are calibrated and combined.
- Infrared precipitation estimates are created using the calibrated microwave precipitation.
- Microwave and IR estimates are combined.
- Rescaling to monthly data is applied by summing all 3-hourly data produced.

GOES data
The Geostationary Operational Environmental Satellite Program (GOES) is a joint mission of NASA and the National Oceanic and Atmospheric Administration (NOAA) [20]. In this study, GOES-13 was the satellite used to gather the climatic data. To obtain the rainfall data from the GOES data, the model proposed in [21] called the auto-estimator was used. With this model, average rainfall estimates are produced every half an hour if there is data available, from July 2014 to the present. This product has a 4 km by 4 km spatial resolution.

Meteoblue data
Initially developed at the University of Basel, based on NOAA and the National Centers for Environmental Prediction (NCEP) models [22]. It delivers local weather information for any point in the world, with a spatial resolution of 5 km by 5 km, providing temperature, wind direction and speed, relative humidity, solar radiation, cloud cover and precipitation, available hourly from 2012 to the present. The data used on this study was acquired by the RICCLISA program from Meteoblue.

Modis NDVI data
The Moderate Resolution Imaging Spectroradiometer (MODIS) vegetation indices, produced at 16-day intervals at multiple spatial resolutions, provide consistent spatial and temporal comparisons of vegetation canopy greenness, a composite property of leaf area, chlorophyll and canopy structure [23] using daily data to produce the composite product. In this study, MOD13A2 16-Day 1 km NDVI data was used.

Digital Elevation Map (DEM)
The Shuttle Radar Topography Mission [24] of NASA produced a DEM with a spatial resolution of 30 m by 30 m with global coverage and free access.

3.3 Data Pre-processing

Initially, we processed the TRMM and GOES databases to give it the correct georeferenced information, in our case the World Geodetic System 1984 (WGS 84) and changed the HDF format to Geotiff format for both cases. To achieve these changes an automatic program was developed using OpenCV and Gdal Libraries. Once this process was completed, we proceeded to extract all the information with the same date from the three databases. Due to the temporal coverage in the Meteoblue data we decide to work with a temporal window, from January 1, 2012 to December 31, 2016.

In the second phase, to fill the gap between January 1, 2012 and July 2014 in the GOES data, three imputation methods were used to fill in the missing values (MV): Predictive Mean Matching (PMM) [25]; HotDeck [26], and Expectation Maximization (EM) [27] in the R software tool. To achieve this, TRMM data was used. due to the close relationship between this data. The GOES data is used in the process to obtain TRMM data. The summary of this process can be seen in Table 1.

Table 1. Results from the imputation methods

Data produced	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	MV
Original	0.0006	0.0054	0.0315	0.4209	0.3277	10.5148	873
PMM	0.000584	0.006157	0.035935	0.458953	0.372143	10.5148	–
HotDeck	0.000584	0.006229	0.036811	0.410911	0.358121	10.5148	–
EMImputation	−3.57588	0.00252	0.09893	0.43213	0.80643	62.08591	–

As can be seen in Table 1, the PMM and HotDeck methods give similar results comparable with the Original data. The EMImputation method produces different results, which must be discarded due to the negative values that have been used to impute the missing values on the dataset. Negative values make no sense for rainfall rate (mmh^{-1}). Finally, the PMM and HotDeck imputed GOES variables were used on this study.

Once this was complete, we selected the data that was going to be used from Meteoblue. As we mentioned before we had multiple variables but only those with a direct or indirect relationship reported on literature were used.

As can be seen in Table 2, different authors have reported a relationship between these variables and precipitation, however Meteoblue wind direction, wind speed, relative humidity, solar radiation, and cloud cover did not report any type of relationship in the literature, so these were discarded in the final dataset.

Table 2. Selected variables and their relationship to rain

Variable	Rain relationship
AMSL	The results from the observation data indicate that the amount of precipitation increases with elevation, following the characteristics of the topography [28].
NDVI	The results show that NDVI depends on the time and amount of precipitation [29].
Humidity	The results showed a high correlation coefficient between rainfall and the occurrence of relatively humidity [30].
Temperature	Temperature and precipitation have a positive or negative correlation depending on spatial location and other climatic variables [31].

Finally, the Oceanic Niño Index (ONI) [32] produced by the NOAA was added to the dataset. This index is the measure of the intensity of the El Niño Southern Oscillation (ENSO) that has a great impact on the Andean Zone. This index has been produced monthly since 1950 to the present. The final dataset has the following variables.

Table 3 shows all the selected variables and the objective variable, with its corresponding name on the dataset, the unit and the source. In this case, Actual Total Precipitation is the objective variable. Max, Min and Prom temperature and the Actual Total Precipitation was extracted from the Meteoblue data.

Table 3. Selected variables and their corresponding units

Variable name	Unit	Source
Year	Numeric (dimensionless)	[22]
Month	Numeric (dimensionless)	[22]
Day	Numeric (dimensionless)	[22]
MaxTemp	Numeric (°C)	[22]
MinTemp	Numeric (°C)	[22]
MeanTemp	Numeric (°C)	[22]
Relative_Humidity	Numeric (%)	[22]
TRMM	Numeric (mm h^{-1})	[18]
NDVI	Numeric (dimensionless)	[23]
ONI	Numeric (dimensionless)	[32]
AMSL	Numeric (m)	[24]
GOES	Numeric (mm h^{-1})	[20]
Actual_Total_Precipitation	Numeric (mm)	[22]

4 Results

RF, Bagging and Stacking predictors in Tables 4 and 5 were used to evaluate the consolidated dataset imputed with PMM and HotDeck. To apply this test, WEKA tool was used and all the predictors were configured using the default parameters, cross validation process was used on the first test. The predictor used were Random Forest (RF), Bagging with Random Forest (Bagging-RF) and Stacking with MLP as meta-predictor and three RF as first level predictors (Stacking–MLP RFx3). This test was done using the PMM and the HotDeck imputed dataset.

Table 4. Results of the first test using the PMM imputed dataset

Predictor	Correlation	MAE	RMSE
RF	0.63	2.78	5.3
Bagging-RF	0.63	2.82	5.3
Stacking–MLP RFx3	0.5	3.4	5.9

Table 5. Results of the first test using the HotDeck imputed dataset

Predictor	Correlation	MAE	RMSE
RF	0.61	2.84	5.4
Bagging-RF	0.62	2.84	5.4
Stacking–MLP RFx3	0.5	3.3	5.8

As can be seen in Table 4 the correlation coefficient in two of the three predictors suggests a strong [33] correlation between the selected variables and Total Precipitation. The MAE and RMSE suggest that the predicted values are close to the real values. The high RMSE value could be due to some points where the predicted value has a large difference from the real value. The lowest correlation result was obtained using the Stacking-MLP RFx3 configuration although the RF predictor has a good performance when using alone or in the Bagging architecture. In this case using the MLP as meta-predictor does not produce an improvement over the previous results. In Table 5 the results are similar to those obtained with the PMM method but the PMM imputed dataset results are slightly better, except for the results of the Stacking-MLP RFx3 configuration. Other predictors such as C4.5, SMOReg, and M5R not shown in this paper present results that improved on those shown here, but in those cases, negative Total Precipitation was predicted. As mentioned before, negative values clearly make no sense in this case.

In light of the above, a second test was done using the PMM imputed dataset. This data set was divided into two. All data prior to 2016 was used as training data and 2016 data used as test data. In Table 6 we can see that the correlation is moderate. The MAE has an increase between 0.1 and 0.2 for RF and Bagging RF, but then RMSE is identical. The Stacking-MLP RFx3 for this case has the same correlation and RF but the MAE and RMSE are higher than the other configuration. For this reason, once again it has the worst results. Although all results suggest that this multiscale approach can produce good results in the prediction.

Table 6. Results of the second test using the PMM imputed dataset

Predictor	Correlation	MAE	RMSE
RF	0.48	2.95	5.3
Bagging-RF	0.49	2.93	5.3
Stacking–MLP RFx3	0.48	3.4	5.5

In all tests, the configuration with the highest correlation coefficient and lowest errors must be selected. In our case, the RF or the Bagging-RF configurations should be selected. Figures 1, 2 and 3 show the predicted value vs. the real value for the three machine learning models using the PMM imputed dataset for the first test case.

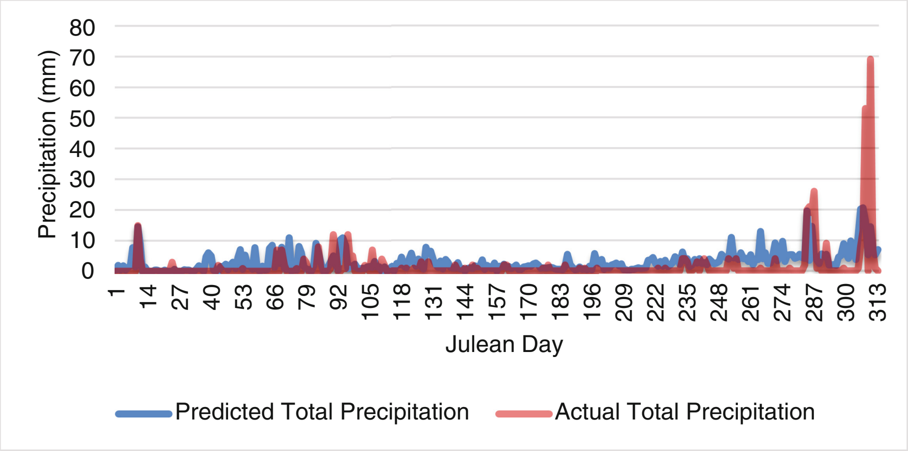


Fig. 1. RF Predicted Total Precipitation vs. Actual_Total_Precipitation

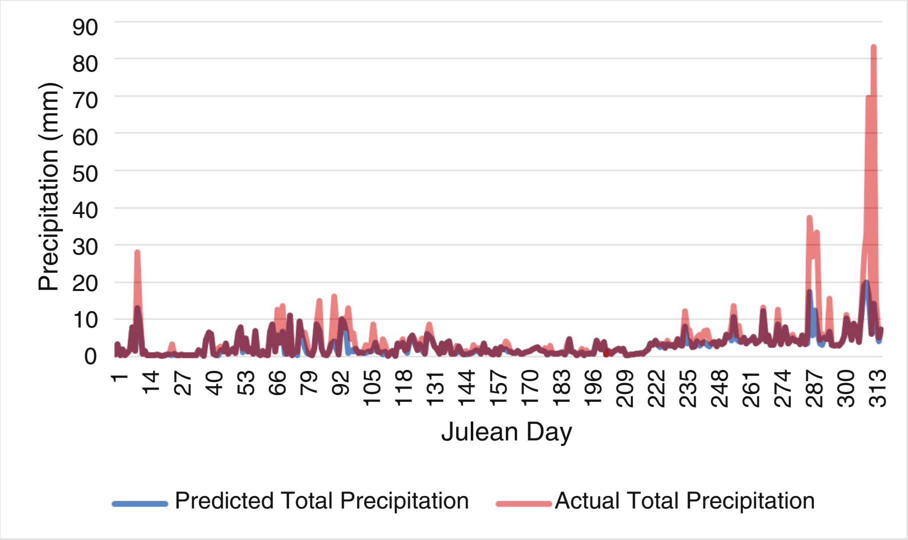


Fig. 2. Bagging-RF Predicted Total Precipitation vs. Actual_Total_Precipitation

As can be seen in Figs. 1, 2 and 3, predicted precipitation and real precipitation have a very similar behavior according to the correlation results obtained before. The three models predict the amount of rain correctly in most cases, except for those where the *Total_Precipitation* has high values. On these high peak values, the *predicted Total_Precipitation* fails to achieve the expected value on all three predictors. But the overall behavior is good. More information is probably needed to adjust the predictions when the amount of rain has high values or to apply a correction after the prediction model to adjust the predicted value in these conditions. Overall behaviors of the three

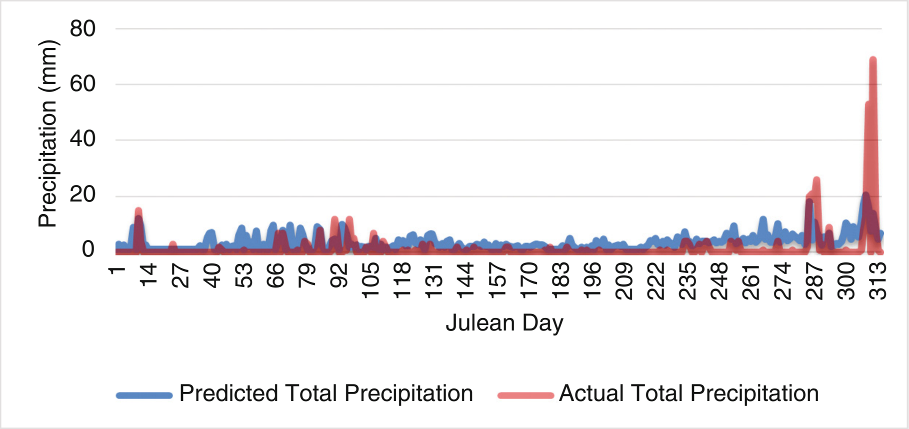


Fig. 3. Stacking-MLP RFx3 Predicted Total Precipitation vs. Actual_Total_Precipitation

configurations are consistent with the correlation coefficients shown in Tables 4 and 5, MSE and RMSE values are the results of this underestimation of the high precipitation peaks.

Finally, a comparison was carried out of the *predictedTotal_Precipitation* from the Bagging-RF model that has the best results on the performed test, the *Total_Precipitation* from *Meteoblue*, and the accumulated rainfall measure by a meteorological station within the study area. The results are showed in Figs. 4 and 5.

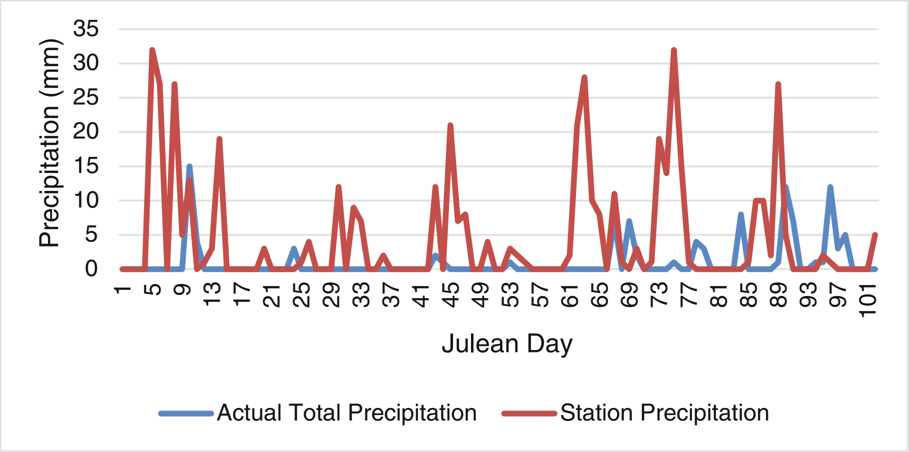


Fig. 4. Actual Total Precipitation vs. Station Precipitation.

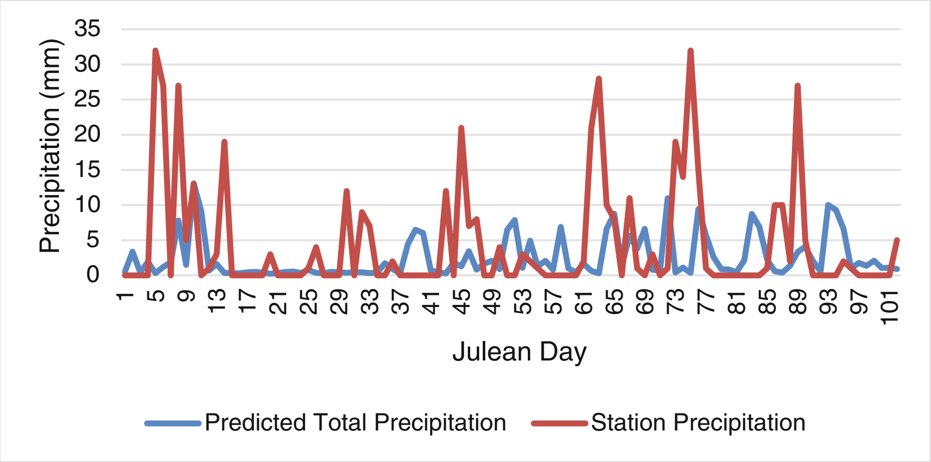


Fig. 5. Bagging-RF Predicted Total Precipitation vs. Station Precipitation.

Figure 4 shows the comparison between the *Meteoblue* (Actual Total Precipitation) data and the meteorological station data from 100+ days in 2016. As can be seen the data has different behavior over the days. There are points where the station registers no rain and the *Meteoblue* data registers rain, and vice versa, although there are some points where the data is consistent between model and reality despite the spatial resolution being different. Finally, a correlation test was carried out using SPSS software. The Pearson correlation coefficient for this data was −0.002. Non-parametric correlations were performed obtaining similar results: 0.029 for Kendall's Tau correlation and 0.032 for Spearman's Rho correlation. In other words, the Meteoblue data and the Station data have no relationship according to Pearson correlations, and a very weak correlation for Spearman's Rho and Kendall's Tau correlations.

Figure 5 shows the comparison between Bagging-RF *predictedTotal_Precipitation* and the meteorological station data. As can be seen, the predicted data behavior seems like a response to some local rain events. The predictor has partially reproduced the accumulated precipitation product of local events that are hard to detect from low spatial resolution models or by satellite measurement. But the model underestimates total precipitation in this case, except for some points where the behavior is similar. In this case, the Pearson correlation value was −0.012, −0.043 for Kendall's Tau correlation and −0.060 for Spearman's Rho correlation. Correlation values slightly increased with the predicted precipitation, but the direction of the correlation changed. These results are consistent with the behavior observed in Fig. 5 where multiple non-precipitation points in the meteorological station data have a coincidence with precipitation points in the predicted data.

5 Conclusions

This paper explains the importance of using multi-spatial scale data in predicting the amount of rain. A reliable prediction tool would allow the decision-making processes related to rain to be more accurate for watershed management authorities. These organizations and government entities may in future be able to take action on how to control the impact of rain events across whole regions according to predictions by these models. Based on the above, this study compares different Multiple Classifier Systems on a dataset composed of available multiple climatological data sources. Experimentally, the RF configuration has the best results on the prediction process using the cross-validation method.

In the final test, the Bagging-RF configuration was best, according to the results shown in Tables 4 and 5. This test has a more realistic approach than the cross-validation one. The 2016 data were not used on the training process, so the predictors would make the prediction according to previous data learning. The results were also interesting due to the ability of Bagging-RF to partially reproduce local rain events, although it underestimates the real amount of precipitation registered at the station. It still however represents a first approximation to a forecast able to reproduce local events without losing the ability to forecast large areas.

The results suggest that more data on highest spatial scales requires to be taken into account to be able to achieve the forecast of the amount of rain in local and large rainfall events, although these local events were able to be successfully detected. Future work would focus on the improvement for the prediction of the amount of rain in these two scenarios.

Acknowledgements. The authors are grateful to the Telematics Engineering Group (GIT) and the Optics and Laser Group (GOL) of the University of Cauca, The University of Cauca, Meteoblue, RICCLISA Program, and the AgroCloud project for supporting this research, as well as the AQUARISC program for the PhD support granted to Cristian Valencia-Payan. We are further grateful to Colin McLachlan for suggestions related to the English text.

References

1. Gagne II, D.J., McGovern, A., Xue, M.: Machine learning enhancement of storm-scale ensemble probabilistic quantitative precipitation forecasts. Weather Forecast **29**(4), 1024–1043 (2014)
2. Han, H., Wen, Y., Chua, T.-S., Li, X.: Toward scalable systems for big data analytics: a technology tutorial. IEEE Access **2**, 652–687 (2014)
3. Ahmed, B.: Predictive Capacity of Meteorological Data - Will it rain tomorrow, September 2014
4. Bendre, M.R., Thool, R.C., Thool, V.R.: Big Data in Precision Agriculture Through ICT: Rainfall Prediction Using Neural Network Approach, pp. 165–175. Springer, Singapore (2016)
5. Gupta, U., Jitkajornwanich, K., Elmasri, R., Fegaras, L.: Adapting K-means clustering to identify spatial patterns in storms. In: 2016 IEEE International Conference on Big Data (Big Data), pp. 2646–2654 (2016)

6. Ang, M.R.C.O., Gonzalez, R.M., Castro, P.P.M.: Multiple data fusion for rainfall estimation using a NARX-based recurrent neural network – the development of the REIINN model. IOP Conf. Ser. Earth Environ. Sci. **17**(1), 12019 (2014)
7. Lee, J., Hong, S., Lee, J.-H.: An efficient prediction for heavy rain from big weather data using genetic algorithm. In: Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication - ICUIMC 2014, pp. 1–7 (2014)
8. Hernández, E., Sanchez-Anguix, V., Julian, V., Palanca, J., Duque, N.: Rainfall Prediction: A Deep Learning Approach, pp. 151–162. Springer, Cham (2016)
9. Tao, W.-K., Lang, S., Zeng, X., Li, X., Matsui, T., Mohr, K., Posselt, D., Chern, J., Peters-Lidard, C., Norris, P.M., Kang, I.-S., Choi, I., Hou, A., Lau, K.-M., Yang, Y.-M.: The goddard cumulus ensemble model (GCE): improvements and applications for studying precipitation processes. Atmos. Res. **143**, 392–424 (2014)
10. Tao, W.-K., Lau, W., Simpson, J., Chern, J.-D., Atlas, R., Randall, D., Khairoutdinov, M., Li, J.-L., Waliser, D.E., Jiang, J., Hou, A., Lin, X., Peters-Lidard, C.: A multiscale modeling system: developments, applications, and critical issues. Bull. Am. Meteorol. Soc. **90**(4), 515–534 (2009)
11. Mohri, M., Rostamizadeh, A., Talwalkar, A.: Foundations of Machine Learning. MIT Press, Cambridge (2012)
12. Stergiou, C., Siganos, D.: Neural Networks. https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html. Accessed 28 July 2017
13. Efron, B., Tibshirani, R.: An Introduction to the Bootstrap. Chapman & Hall, New York (1994)
14. Breiman, L.: Bagging predictors. Mach. Learn. **24**(2), 123–140 (1996)
15. Wolpert, D.H.: Stacked generalization. Neural Netw. **5**(2), 241–259 (1992)
16. Breiman, L.: Random Forests. Mach. Learn. **45**(1), 5–32 (2001)
17. Precipitation Measurement Missions. https://pmm.nasa.gov/trmm. Accessed 28 July 2017
18. Tropical Rainfall Measuring Mission. https://trmm.gsfc.nasa.gov/3b42.html. Accessed 28 July 2017
19. Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P., Stocker, E.F.: The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeorol. **8**(1), 38–55 (2007)
20. Jenner, L.: GOES Overview and History (2015)
21. Vicente, G.A., Scofield, R.A., Menzel, W.P., Vicente, G.A., Scofield, R.A., Menzel, W.P.: The operational goes infrared rainfall estimation technique. Bull. Am. Meteorol. Soc. **79**(9), 1883–1898 (1998)
22. Meteoblue. https://content.meteoblue.com/en/about-us. Accessed 28 July 2017
23. MODIS Vegetation Index Products. https://modis.gsfc.nasa.gov/data/dataprod/mod13.php. Accessed 28 July 2017
24. Shuttle Radar Topography Mission. https://www2.jpl.nasa.gov/srtm/. Accessed 28 July 2017
25. Vink, G., Frank, L.E., Pannekoek, J., van Buuren, S.: Predictive mean matching imputation of semicontinuous variables. Stat. Neerl. **68**(1), 61–90 (2014)
26. Andridge, R.R., Little, R.J.A.: A review of hot deck imputation for survey non-response. Int. Stat. Rev. **78**(1), 40–64 (2010)
27. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. Ser. B **39**(1), 1–38 (1977)
28. Sasaki, H., Kurihara, K.: Relationship between precipitation and elevation in the present climate reproduced by the non-hydrostatic regional climate model. SOLA **4**, 109–112 (2008)

29. Purevsuren, T., Hoshino, B., Ganzorig, S., Sawamukai, M.: Spatial and Temporal Patterns of NDVI Response to Precipitation and Temperature in Mongolian Steppe
30. Umoh, A.A.: Rainfall and relative humidity occurrence patterns in uyo metropolis, Akwa Ibom state, South-South Nigeria. IOSR J. Eng. **3**(8), 27–31 (2013)
31. Trenberth, K.E., Shea, D.J.: Relationships between precipitation and surface temperature. Geophys. Res. Lett. **32**(14), July 2005
32. Climate Prediction Center. http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml. Accessed 28 July 2017
33. Yeager, K.: LibGuides: SPSS Tutorials: Pearson Correlation

Characterization in the Visible and Infrared Spectrum of Agricultural Crops from a Multirotor Air Vehicle

Julian Andrés Bolaños(✉), Liseth Campo, and Juan Carlos Corrales

University of Cauca, Popayán, Colombia
{julianbolanos,liscampo,jcorral}@unicauca.edu.co

Abstract. Plant health has a direct impact on the quality and quantity of agricultural products. Due to this fact, farmers must monitor crop conditions frequently, however the current tools for achieving this are complex and inaccessible. Therefore, this article proposes a method for the characterization of agricultural crops that allows a monitoring of the plants using photographs in the visible and infrared spectrum acquired from a multi-rotor air vehicle, using low cost cameras and free use software. The characterization is performed by identifying the Normalized Difference Vegetation Index (NDVI) in the photographic mosaics of the crops. This index provides information about plant health, consequently it is calculated and represented on a NDVI map, where the status of a crop is analyzed. The highest values of NDVI represent healthy plants and the lowest, plants with problems, water or others. The proposed method allows the monitoring of agricultural crops in a temporary and spatial form letting to a producer adopt measures that help the optimization of resources.

Keywords: Infrared spectrum · Unmanned Aerial Vehicles · NDVI · Plant health

1 Introduction

Colombia is a rich country in raw materials and enjoys tropical conditions to carry out agricultural practices, these conditions allow that all kinds of products grow. According to this context, studies are executed from engineering to achieve solutions that improve the production and quality of the products, these solutions are oriented towards Precision Agriculture (PA), which is based on the development of excellent agricultural practices in order to optimize farming techniques. Precision farming has evolved over the last 10 years up to implement technological advances such as Geographic Information Systems (GIS), Global Positioning Systems (GPS), satellite or terrestrial sensors, software tools, data processing, among others; these methods help the farmer to know in detail crop conditions and to carry out actions of containment, prevention and sustenance [1, 2]. One important phase in precision agriculture is information processing, which is the basis for crop characterization, and is basically the knowledge of their physical and physiological properties. Among the relevant methods for crop characterization is aerial photogrammetry, which includes georeferenced mosaics, digital terrain and surface models and vegetation indexes that correspond to mathematical combinations between

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_3

the different spectral levels found in a photography [3]. Generally, photogrammetric analysis are performed on satellite images [2]; however, this technology involves high implementation costs and low resolution. This difficulty generates interest in evaluating other alternatives such as the use of Unmanned Aerial Vehicles (UAV). A UAV has the advantage of carrying high resolution sensors such as lasers, heat sensors, infrared cameras, multispectral cameras, RGB (visible spectrum) cameras and ultraviolet cameras; outstanding among which are cameras of visible and infrared spectrum for their price and availability. Through the use of cameras in the visible and infrared spectrum, fundamental vegetation indexes are calculated such as NVDI (Normalized Difference Vegetation Index) which can be useful to determine the vegetative state of the plants [3]. With a correct agricultural crop characterization, it is possible to interpret the data acquired in the practices of precision agriculture, thus optimizing fertilizers application and an effective control in time and space. A correct interpretation translates into economic benefits due to fertilizers savings and a reduction of the treatments performed in the influenced areas.

The direct observation of the plants does not allow to know objectively the state of them, due to that fact it is necessary to use a system able to recognize different bands of the electromagnetic spectrum that reveal the crop health, and allow monitoring the plant growing. These requirements can be met by performing an agricultural crop characterization using photographs in the visible and infrared spectrum. In the market there are technologies dedicated to this task, however, their high cost does not allow them to be easily accessible to farmers, these situations motivate to the development of low-cost technology that could give benefits to the farmer, and also promote agricultural activities [8, 9].

The present proposal is based on works by Bendig et al. [2] who studied the biomass in an agricultural crop from the vegetation indexes NDVI, SAVI, MSAVI, OSAVI, GnyLi, GRVI, MGRVI, and RGB VI that are calculated through images in the visible and near infrared spectrum obtained from an octacopter UAV. The objective is to find relations between the physical factors of the crops and the captured images. The work developed in [4] is based on results comparisons taking into account the size of the pixels, it is sought to evaluate the accuracy of the vegetation indexes NDVI, GNDIV, NGRDI, and TGI obtained from images in the visible spectrum and near infrared. The method is based on free software with the disadvantage that increasing the number of images and the size of the crop, increases exponentially the effort and work required to obtain the results, Wang et al. [5] uses the vegetation indexes SAVI and RVI to evaluate nitrogen concentration in leaves of wheat and rice plants, the efforts made are based on analyzing the behavior of the different spectral bands, and it is concluded that vegetation indexes are appropriate to determine the nitrogen absorption in the leaves. Li et al. [6] is desired to find the height of the canopy as well as the area of biomass within a maize crop using images obtained from a low-cost unmanned aircraft. The location are 84 km north of Beijing, China, at 30 m above sea level. A flight at a height of 150 m and at a temperature of 9.1 °C is made, capturing RGB images with 80% longitudinal overlap and a 40% lateral overlap, vegetation indexes found in the range of the visible spectrum are used. Information on canopy height and biomass are presented in a georeferenced map obtained from vegetation indexes. The work in [7] describes algorithms for the detection of the plant fraction in a crop from multispectral images captured at different times to measure the spectral reflectance during plant growth, using the vegetation indexes NDVI, GNDVI, and VARI to later relate them with the existing vegetal fraction, the results are presented as graphs of percentage of vegetal fraction related to each one of the indexes of vegetation studied.

In the review, most of the software tools used for image processing and subsequent obtaining of vegetation indexes are privative products. It can also be evidenced that the majority of the projects found are focused on the calculation of the vegetation indexes by multispectral cameras of high cost, therefore, the use of an accessible commercial camera is proposed. Throughout this work, free-use processing tools are accessible to any farmer in the region. One of the most important approaches is obtaining a reliable NDVI from a commercial camera, throught the approximation of the resulting bands and the use of suppressive filters of visible light allowing to obtain information of infrared radiation which is useful to determine the vegetative state of plants.

The Sect. 1 in this article presents a brief introduction regarding vegetation indexes, later on Sect. 2 the tools used are presented and finally, Sects. 4 and 5 describes the processing stage accompanied by the results. The conclusions are presented in Sect. 4.

2 Background

For a better interpretation of the article the concepts handled in this research are described below.

NDVI Calculation

The visible and infrared bands of the electromagnetic spectrum are directly related to different physiological factors on plants, such as the process of photosynthesis, humidity, chlorophyll, etc. Normally the spectral reflectance in a plant varies according to the wavelength and its physiological state, therefore, it is possible to make mathematical combinations between these spectral bands to highlight the vegetation of its surroundings, the result of this process is denominated index of vegetation and is useful to have an idea of the physiological processes of plants [3]. According to [10–12] the NDVI is suitable to carry out a process of crops monitoring in order to determine the quality of the growth of the plants, thus is possible to determine affectations by pests or diseases, or if there is a problem of water stress in the plants. In [13–15], it is found that is possible to approximate the values of NDVI using a commercial camera by removing the infrared cutting filter and adding a filter that is capable of filtering the red band or the blue band of the visible spectrum, taking advantage of this space to form an image in false color. The relationship involved for NDVI in this case is represented in Eq. (1).

$$NDVI = \frac{NIR - VISIBLE}{NIR + VISIBLE} \tag{1}$$

In which the band corresponding to the NIR represents the one that removes the special filter used, and the visible band corresponds to the remaining band. In particular, NDVI with the red band blocker filter is calculated with Eq. (2) and the blue band with Eq. (3).

$$NDVI = \frac{RED - BLUE}{RED + BLUE} \tag{2}$$

$$NDVI = \frac{BLUE - RED}{BLUE + RED} \tag{3}$$

3 Technology and Method

The aircraft used was the Phantom 3 Standard quadcopter which has an approximate operating range of 500 m and an autonomy of approximately 25 min operated by an application called Litchi from a mobile device. This application allows to create missions manually from the Waypoints function, and to deploy them in the aircraft according to the desired configuration. The Raspberry Pi camera has a resolution of 5 Megapixels. It has two band-blocking filters which are the Roscolux #19 and the Roscolux #2007.

This article presents some experiences and considerations for precision agriculture applications using UAV platforms, low cost cameras and free license processing software. The experiments are focused on achieving a reliable NDVI from a commercial camera that captures infrared light. In particular, the proposed method focuses on:

A. The flight planning and images capturing by a commercial camera carried by an unmanned aerial vehicle.
B. Stitch of images obtained.
C. Image processing to obtain a reliable NDVI.
D. Analysis and testing over time to monitor agricultural crops in a real environment.

A. Flight Planning and Images Capturing

The flight planning and images capturing are the processes by which the photographic information is collected in the field, this processes are important to obtaining quality results. The flight planning takes into account the prior phases to obtaining information in the field, among these are the analysis of aspects such as target area, height, possible obstacles and others. This project considers calculations such as the maximum flying speed, the distance between flight lines and the distance between captures, in order to optimize resources and improve outcomes [16–18]. The capture process is the one that is carried out taking into consideration the information obtained in the planning process. In this work the Mission Planner software is used, which will help to establish the routes for autonomous flights. Finally, in the camera module, repetitive captures are implemented with a certain configuration from time to time.

B. Stitching of images obtained

After the capture step it is important to process the images obtained, the software used is Image Composite Editor which is freely used for non-commercial purposes. This process consists of forming a single image from the individual images captured at different points in the field and preferably at the same height.

C. Image processing to obtain a reliable NDVI

The Fig. 1 shows the proposed diagram to obtain NDVI values. Python language is used due to its versatility and all implementations are done using the Linux operating system. The proposed method allows obtaining as output a NDVI map which corresponds to a representation of the NDVI index in graphic form.

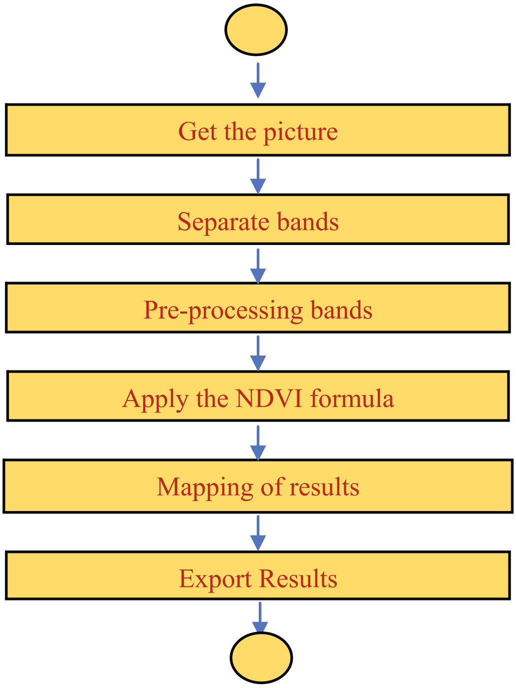


Fig. 1. Image analysis process.

- **Approximation of visible spectrum bands**

In this work, a normal RGB camera identical to the Raspberry Pi NoIR camera is used, with the only difference that it has the IR filter. The following tests are performed in a controlled environment in order to compare the bands such as the response of what should look like pure red, pure blue and pure green. Tests were performed on coffee plants:

(a) Verification and comparison of the resulting bands from the two cameras
(b) Approximation of the resulting visible spectrum bands in the two cameras.

(a) Verification and comparison of the resulting bands of the two cameras

The objective of this test is to establish similarities and differences between the visible band used to calculate the NDVI with the NoIR camera and the visible band of the RGB camera (Fig. 2).

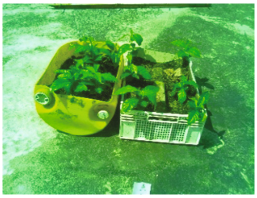


Fig. 2. Original image (RGB)

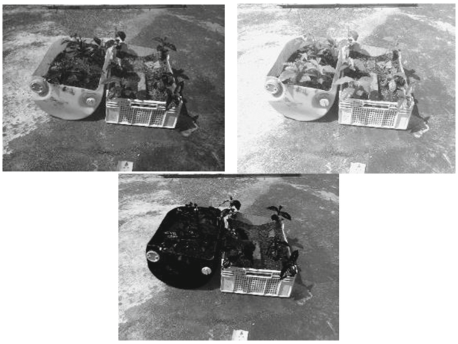


Fig. 3. Red, Green and Blue RGB image bands

Figure 3 presents the behavior of the plants around the visible spectrum, it can be seen that in red and blue, these do not have a high reflectance. These results allow to have a reference to obtain the bands that correspond to the visible spectrum when using the NoIR camera. The first step is a decomposition of the images obtained with the NoIR camera using both filters. This can be seen below (Figs. 4 and 6):

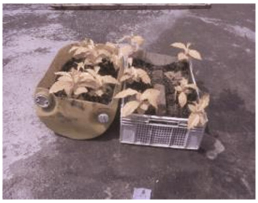


Fig. 4. Image NoIR #2007 Filter

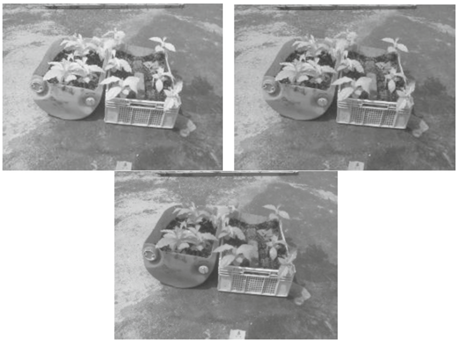


Fig. 5. Bands image NoIR #2007 filter (Red, Green and Blue)

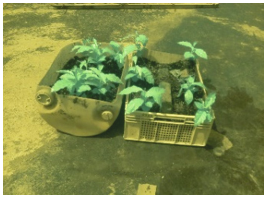


Fig. 6. Image Noir #19 Filter

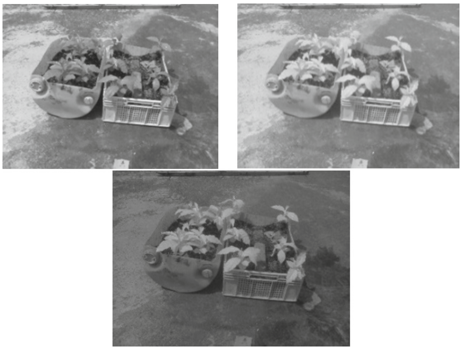


Fig. 7. Bands Image NoIR #19 filter (Red, Green and Blue)

In Figs. 5 and 7, it can be seen that the bands used to store the information of the visible spectrum do not have similarity to the bands originated by the RGB image, which leads to analyze the processing of the bands to bring those results closer.

(b) Approximation of the visible spectrum resulting bands in the two cameras

As it can be seen, there are notable differences between the visible bands resulting from the RGB camera, and visible bands resulting from the NoIR camera dependent on the filter to be used. This is due to any used filters these bands are contaminated by infrared information, which causes this difference. Taking into account this information, a mathematical process is carried out to achieve an approximation between the visible bands resulting from the Noir camera and RGB camera. In order to approximate the bands, a mathematical manipulation of the visible bands resulting from the capture with the NoIR camera is made and it is based on those obtained with the RGB camera under the same conditions.

It is known that the visible band information is affected by infrared information that can not be filtered, and from this fact it is proposed to remove that information by subtracting the pure infrared band obtained through the use of the filter, to the visible band of the same photography. Specifically the proposed equation is:

$$VISresult = A * (B * VISobtaining - C * INFRAREDobtaining) \tag{4}$$

Where, VISresul is Resultant visible band, VISobtaining is Visible Band captured, INFRAREDobtaining is Captured infrared band and A, B and C are Multiplying decimal constants.

In order to determine the values of A, B and C, the resulting visible band must be compared with the visible band obtained from the RGB camera, this comparison process is performed by the linear correlation index that defines the correlation between two random variables, in this case two images. The values of this index range from −1 to 1, of which the value of least similarity is 0, which means that the two images have nothing in common. This index can be calculated using the OpenCV functions that are already predefined in the library [20].

In general, it is proposed the design of an algorithm that allows to modify that values in a determined range until obtaining a value for each one of them that allows to calculate the maximum of possible correlation and to verify the results. The flow diagram used is presented in the Fig. 8.

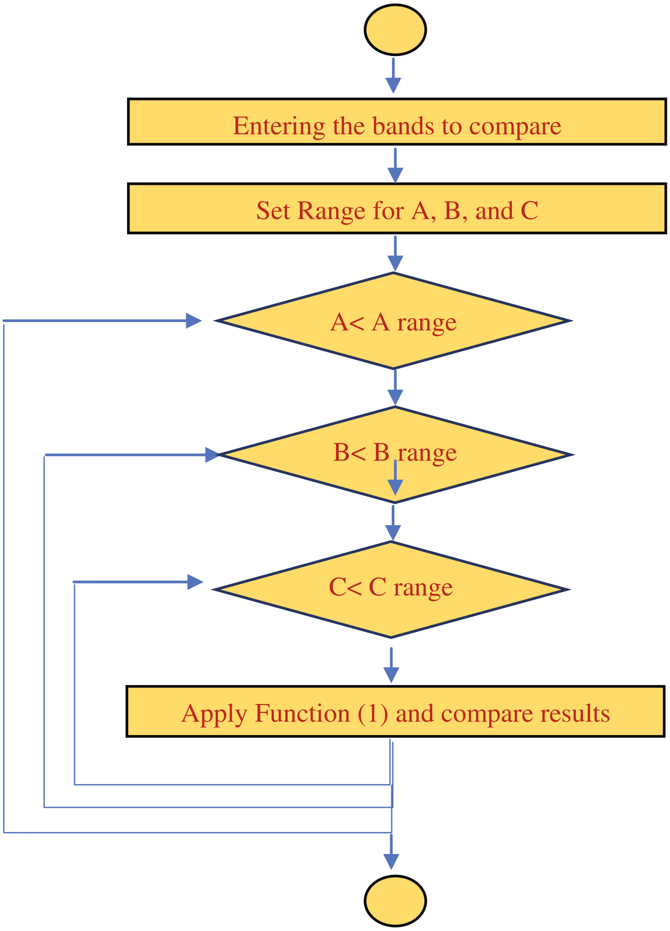

Entering the bands to compare
Set Range for A, B, and C
A< A range
B< B range
C< C range
Apply Function (1) and compare results

Fig. 8. Flow diagram band approach.

The correlation index obtained allow to define values for A, B and C suitable for each of the filter configuration used in the photographs aligned in the test of the combined bands. The process is performed with three different white balance settings.

D. Analysis and testing over time to monitor agricultural crops in a real environment

In order to test the proposed method, captures are made in a coffee crop in the municipality of Timbío, department of Cauca, at 1,782 m above sea level [21], delimiting an area of 1 ha of flat terrain without obstacles higher than 50 m. The objective of these studies is to demonstrate the capacity of the proposed method for calculating the NDVI that allows to carry out a monitoring of agricultural crops in a time lapse of three weeks.

The testing process begins by collecting the images in the crop, considering that in order to make the samples comparable, it is necessary to keep the initial configuration as much as possible. After collecting the images, the stitching process of pictures is executed, and the tool to be used is Image Composite Editor. The Mosaics are shown in the Fig. 9 using the process described in [22]. In the Figs. 9, 10 and 11 three configurations established for the camera are presented using the visible band eliminator filters whose characteristic curve can be found in [19]. In total, three captures were made, with one week of time difference between them.

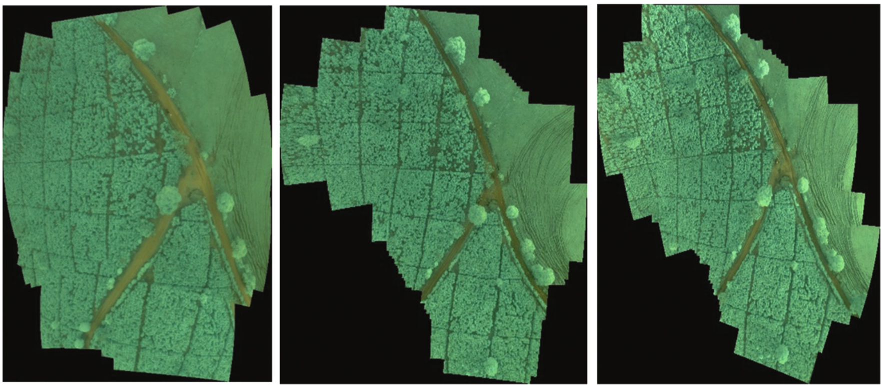


Fig. 9. Mosaics with #19 filter and AWB 0.54,1 configuration

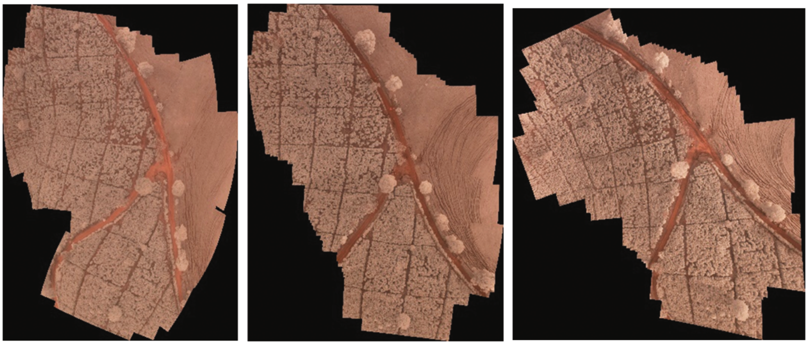


Fig. 10. Mosaics with #19 filter and AWB 1,1 configuration

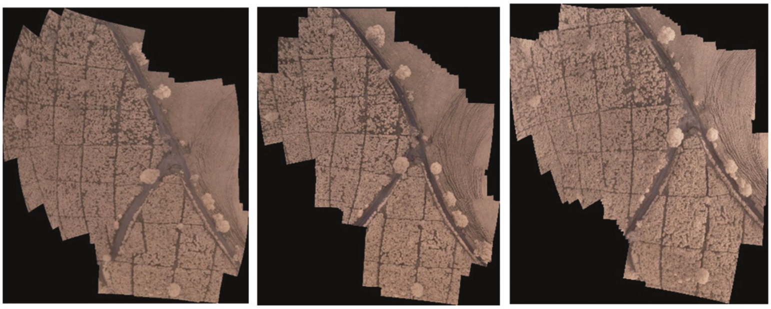


Fig. 11. Mosaics #2007 filter and AWB 1,1 configuration

- **Filter #19 AWB 0.54,1**
  This configuration allows the capture of images for the NDVI calculation using the filter #2007
- **Filter #19 AWB 1,1**
  This configuration is an alternative to using filter #19 with which NDVI calculations are also performed, the process to be executed was the same as the previous configuration.
- **Filter #2007 AWB 1,1**
  This setting is for NDVI calculation using filter #2007.

NDVI with #19 filter AWB 1,1 configuration
In Fig. 12 the NDVI generated in the three corresponding tests are presented with the configuration of the filter #19 Awb 1,1. These tests correspond respectively to the results of each of the images captured in each of the tasks, when the process is finished, these results are generated automatically.

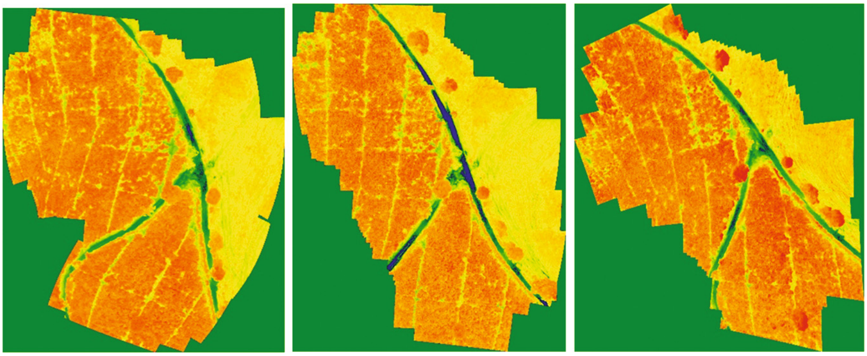


Fig. 12. NDVI with #19 filter and AWB 11 configuration

NDVI with #19 filter AWB 0.54,1 configuration
Figures 13 and 14 below.

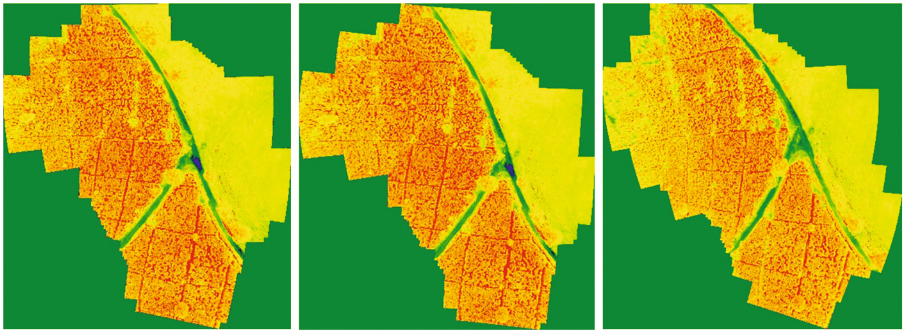


Fig. 13. NDVI with #19 filter and AWB 0.54,1 configuration

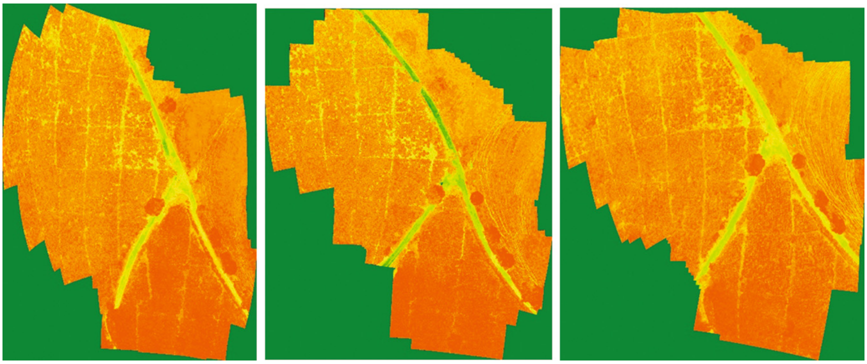


Fig. 14. NDVI with #2007 filter and AWB 11 configuration

NDVI with #2007 filter AWB 1,1 configuration

Nomenclature

One of the most important sources of radiation is the sun, the surfaces absorb or reflect this radiation and that depends on its wavelengths, or the characteristics of such surfaces as their composition or roughness [23, 24]. Healthy vegetation, for example, absorbs almost all the wavelengths that make up the visible spectrum, this is because of photosynthesis, but due to the molecular structure of the leaves, it completely reflects the wavelengths in the near infrared band. Pure water reflects low energy at all wavelengths, different from ice and snow, which have a high reflectance in all bands. In the case of soil, reflected energy depends on the presence of organic matter, texture and roughness, among others [24, 25].

The interpretation of the results shown is centered in describing the meaning of the colors that compose the resulting maps. According to [26–29] NDVI is an normalized index, which means that it has infinite values that are between −1 and 1, in which the vigorous plants with good health reflect all the energy of the sun in the band of the near infrared, whereas absorb about 90% in the red and blue bands, for the process of photosynthesis, reflecting part of the green band, due to the fact that the human eye perceives them in that color, which leads to have a high NDVI, greater than 0.5, whereas plants with some type of anomaly or problem reflect energy equally in all the bands described, the visible spectrum being comparable with the near infrared, which means that they have a lower NDVI, between 0.1 and 0.3 [28, 29]. For this reason, in Fig. 15 the continuous NDVI map is established where the NDVI values are mapped, with the blue color on the left end representing −1 and the red color on the right end representing 1.

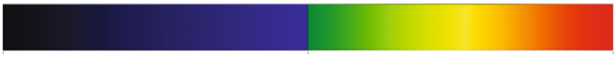


Fig. 15. Colormap

Figure 16 shows the classification that is made to achieve an interpretation to the results shown according of the defined nomenclature.

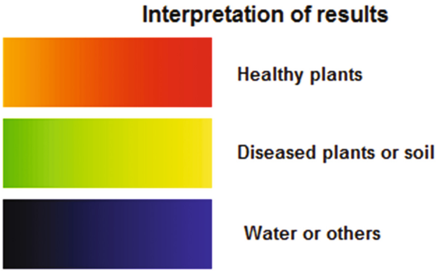


Fig. 16. Interpretation of results

From the results presented in Figs. 18, 19 and 20 it can be said that although there are differences between them, all three keep the above characteristics. However, the one that best represents the expected information is the one in Fig. 12, with this configuration a suitable representation is obtained for the values corresponding to the soil and the vegetation, achieving a good contrast between the different NDVI values. It is important to emphasize the temporal variation between the three consecutive captures at different times. Although in general terms the crop condition remained constant, a slight variation can be observed between them that may belong to changes in the physiology of the plants.

Considering the nomenclature and the analysis above, it can be affirmed that in the three defined configurations of the camera, there is a relation between what is obtained and what is expected according to the literature, this mainly refers to the fact that for vegetation there are high values of NDVI that approach 1 and are classified as healthy or vigorous plants, furthermore it can be appreciated that for the areas that do not correspond to the vegetation there are low NDVI values of around 0.3 or even negative values. As the tests were performed with one week of difference in each one and the environmental conditions were constant, a great difference in time between the results is not appreciated. However, it can be seen in Fig. 12 that belongs to the capture with a white balance of 1,1, represents the camera configuration that presents a superior performance compared to the others. This configuration maintains a differentiated contrast in comparison to the other captures, this means to establish it as the configuration to be used for future works.

4 Conclusions

In this article, a method for the characterization of agricultural crops using a multirotor air vehicle is proposed, it is possible to obtain a reliable NDVI to determine and monitor the health of plants in any type of agricultural crop through a low cost commercial camera. The proposed method for the monitoring of agricultural crops is economically convenient since the hardware and software processing tools used for processing are free to use and it is possible to monitor agricultural crops over a given time interval using the proposed method for monitoring agricultural crops. It is necessary to maintain environmental conditions to obtain valid results. The NDVI map obtaining is based on the infrared information from crops, and it is necessary for the analysis of the evolution of plants individually an experimented farmer as future work. This paper is a basis for future work focusing on low cost solutions.

Acknowledgements. The authors are grateful to the University of Cauca and its Telematics Engineering Group (GIT), the Colombian Administrative Department of Science, Technology and Innovation (Colciencias), AgroCloud project of The Interinstitutional Network of Climate Change and Food Security of Colombia (RICCLISA) for supporting this research.

References

1. Ahmad, A.S., Manan, A.: Aerial mapping using high resolution digital camera and unmanned aerial vehicle for Geographical Information System. In: 2010 6th International Colloquium on Signal Processing and Its Applications (CSPA), pp. 2–6 (2010)
2. Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M., Bareth, G.: Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. Int. J. Appl. Earth Obs. Geoinf. **39**, 79–87 (2015)
3. Gonzaga Aguilar, C.: Aplicación de índices de vegetación derivados de imágenes satelitales Landsat 7 ETM+ y ASTER para la caracterización de la cobertura vegetal en la zona centro de la provincia de Loja, Ecuador (2014)
4. Hunt, E.R., Daughtry, C.S., Mirsky, S.B., Hively, W.D.: Remote sensing with simulated unmanned aircraft imagery for precision agriculture applications. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. **7**, 4566–4571 (2014)
5. Wang, W., Yao, X., Tian, Y., Liu, X., Ni, J., Cao, W., Zhu, Y.: Common spectral bands and optimum vegetation indices for monitoring leaf nitrogen accumulation in rice and wheat. J. Integr. Agric. **11**, 2001–2012 (2012)
6. Li, W., Niu, Z., Chen, H., Li, D., Wu, M., Zhao, W.: Remote estimation of canopy height and aboveground biomass of maize using high-resolution stereo images from a low-cost unmanned aerial vehicle system. Ecol. Ind. **67**, 637–648 (2016)
7. Gitelson, A., Kaufman, Y., Stark, R., Rundquist, D.: Novel algorithms for remote estimation of vegetation fraction. Remote Sens. Environ. **80**, 76–87 (2002)

8. Mundial, B.: Agricultura y desarrollo-rural (2016). http://datos.bancomundial.org/tema/agricultura-y-desarrollo-rural?locations=CO. Accessed 20 Nov 2016
9. Mundial, B.: Indicadores de desarrollo del mundo (2015). http://wdi.worldbank.org/table/4.2. Accessed 10 July 2016
10. D. P. D. A. D. R. Negro: "NDVI," Área de planificación y evaluación de los recursos hídricos, Rio Negro, Argentina (2014)
11. ArcGis: Función NDVI (2016). http://desktop.arcgis.com/es/arcmap/10.3/manage-data/raster-and-images/ndvi-function.htm. Accessed 20 Jan 2017
12. Piura, U.: Biblioteca Universidad de Piura. http://www.biblioteca.udep.edu.pe/BibVirUDEP/tesis/pdf/1_59_185_24_508.pdf. Accessed 10 June 2016
13. Dworak, V., Selbeck, J., Dammer, K.-H., Hoffmann, M., Zarezadeh, A.A., Bobda, C.: Strategy for the development of a smart NDVI camera system for outdoor plant detection and agricultural embedded systems. Sensors **13**(2), 1523–1538 (2013). Multidisciplinary Digital Publishing Institute
14. José, B., Nicolás, M., Danilo, C., Eduardo, A.: Multispectral NDVI aerial image system for vegetation analysis by using a consumer camera. In: 2014 IEEE International Autumn Meeting on Power Electronics and Computing (ROPEC), pp. 1–6 (2014)
15. Ghazal, M., Al Khalil, Y., Hajjdiab, H.: UAV-based remote sensing for vegetation cover estimation using NDVI imagery and level sets method. In: 2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 332–337 (2015)
16. Rodrigo Andres, N.R.: Fotografia aerea (2005). http://cybertesis.uach.cl/tesis/uach/2005/bmfcin415f/doc/bmfcin415f.pdf. Accessed 1 Oct 2016
17. Abraham_Thomas: Aerial Photography. University of the Western Cape. http://www.cdioinstitute.org/papers/Day1/AERIAL%20PHOTOGRAPHY_Abraham_Thomas.pdf. Accessed 10 Jan 2017
18. Diaz, A., Cisternas, M.A.: Fotografías aéreas aspectos técnicos. https://wikigeomatica.wikispaces.com/file/view/fotosaereas.pdf. Accessed 10 Dec 2016
19. Rosco: Roscolux Filters. http://us.rosco.com/en/products/catalog/roscolux. Accessed 20 Nov 2015
20. OpenCV: OpenCV Image Matching. http://docs.opencv.org/2.4/doc/tutorials/imgproc/histograms/template_matching/template_matching.html. Accessed 20 Jan 2017
21. Timbío: Información General Timbío. http://timbio-cauca.gov.co/informacion_general.shtml
22. Microsoft Research: Image Composite Editor, February 2015. https://www.microsoft.com/en-us/research/product/computational-photography-applications/image-composite-editor/. Accessed 20 Jan 2016
23. Andreo, V.: Remote Sensing and Geographic Information Systems in Precision Farming (2013)
24. HS University: Spectral Reflectance. http://gsp.humboldt.edu/OLM/Courses/GSP_216_Online/lesson2-1/reflectance.html. Accessed 15 Jan 2017
25. University of Oldenburg: Spectral Reflectance Properties. http://www.seos-project.eu/modules/remotesensing/remotesensing-c01-p05.html. Accessed 4 Dec 2016
26. Velasquez, L.C., Argueta, J., Mazariegos, K.: Implementation of a low cost aerial vehicle for crop analysis in emerging countries. IEEE J., 21–27 (2016)
27. Ortiz, B., Shaw, J., Fulton, J.: Basics of crop sensing. Alabama Cooperative extension system, pp. 1–3 (2011)

28. Gitelson, A.A., Gritz, Y., Merzlyak, M.N.: Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Physiol. **160**(3), 271–282 (2003)
29. Gates, D.M., Keegan, H.J., Schleter, J.C., Weidner, V.R.: Spectral properties of plants. Appl. Opt. **4**(1), 11–20 (1965)

Cover Missions Planning for Unmanned Aerial Vehicles

Jhon Fredy Trujillo[✉], Yesica Villamil, Liseth Campo, and Juan Carlos Corrales

University of Cauca, Popayán, Colombia
{jftrujillo,yavillamil,liscampo,jcorral}@unicauca.edu.co

Abstract. Unmanned Aerial Vehicles are starting to be used for photogrammetric sensing, in several applications, such as precision agriculture, topography and search and rescue. This, thanks to the advantages they offer over platforms like satellites and manned airplanes, such as: greater spatial and temporal resolution and also low operating costs. However, they have a limiting flight time making necessary to obtain paths that optimize the resources. In this article, we present the study of three heuristic algorithms for coverage path planning, Wavefront, Spanning-tree based coverage and a new algorithm based on expansion by weight functions, called Sunshine with the aim of finding the best strategy to find coverage routes and their subsequent use in Precision Agriculture (PA). These methods are validated by means of simulation at different wind speeds in order to analyze how the energy consumption and the total path traveled differ according to this.

Keywords: Coverage path planning · Unmanned aerial vehicles · Wavefront · Spanning tree coverage · Precision agriculture

1 Introduction

Coverage path planning is defined as the task of calculate a path for an agent which ensures that it can pass through all points in an area or volume of interest, avoiding obstacles and taking into consideration the specific constraints of the environment. This problem has been mainly addressed to terrestrial robots in applications such as cleaning, agriculture, demining, among others [1]. The coverage works on a particular area using Unmanned Aerial Vehicles (UAVs), involves some additional restrictions. Typically, UAVs have a shorter mission time compared to terrestrial robots, in some cases their autonomy is less than one hour, for that reason the available resources should be used efficiently. Also, the payload affects directly the work cycle. In addition, the landing and take-off points must be defined [2]. Depending on the application, the problem restrictions may change slightly. These platforms are mainly divided by the arrangement of their wings, which can be fixed wings and, rotary wings or multi-rotor. Within the category of rotating wings are quadcopter, which have four motors that control the movements of the vehicle, changing the power in each one of them. These ones are affordable and easy to use for research in this field.

In recent years the quadrotors are starting to be introduced into aerial photography activities, allowing the collection of information about the physical properties of any

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_4

object through cameras and specialized sensors which can be easily integrated into these platforms. A field with great potential for these is the Precision Agriculture (PA), which is a modern trend that allows the management of crops and soils specifically, that is according to their needs, achieving that each area is given the most suitable agronomic treatment [3]. The PA requires the acquisition of an accurate and large amount of information. The use of these vehicles provides some advantages over platforms such as satellites, such as: a higher spatial resolution, achieving resolutions of the order of centimeters (1–5 cm) that allow to detail a particular crop. A higher temporal resolution, which facilitates deployments either periodically or every time it is necessary, instead of on-demand deployments. Operating costs are also significantly reduced, making them affordable for small and medium-sized farmers [4].

In works like [5] the development of a system of image collection for the estimation of the biomass in wheat and barley crops is presented, using two different types of cameras, a multispectral camera and a Near Infrared (NIR) camera. The images processing is done after the scheduled flight mission has finished at a constant height of 140 m. Another work involving the use of quadcopters for image collection and later calculating the Normalized Difference Vegetation Index (NDVI) of plants in a potato crop is presented in [6] using multispectral and RGB cameras mounted on a quadcopter that operated at 40 m Height. The flight mission is set up through a ground station defining crossing points trough of the crop coordinates.

But, a general problem exists in the proposals found in literature and it is about the way the coverage path is found. This path, is obtained using not efficient algorithms, like Zig Zag algorithm [7, 8], which do not take care about the battery limits of the quadcopter. The use of better algorithms generates better paths which optimize the flight time and the traveled distance, for that reason, three algorithms are showed and tested in this paper, which are wavefront, Spanning-tree based coverage, and a new proposal of an algorithm based on expansion by weight functions, called Sunshine. These algorithms are compared, showing the advantages and disadvantages of each one with values about their performance in a simulated environment with different kind of winds.

The organization of the document is as follows: After this brief introduction, Sect. 2 introduces the problem of CPP, and implemented algorithms. Section 3 presents results and conclusions.

2 Coverage Path Planning (CPP)

In the task of finding a path that covers an area of interest, two main approaches can be considered in the implementation of algorithms for the path planning: local (online) and global (offline) planners. Local schemas retrieve information from sensors to create a view of the environment in real time. It provides robustness and flexibility but it also demands more CPU processing which means higher power consumption. For that reason, local schemas are not optimal for aerial robotics. This paper approach global schemes, which use stationary information, since the environment has been studied already and is known a priori [1].

To ensure the planner generates a path that completely covers free space, many planners implicitly or explicitly decompose the working region. This can be in cells or grids. The exact decomposition approaches using cells always give an accurate representation of the robot's working space, where the robot can cover the cells with forward and backward motions. On the other hand, grid-based methods divide the environment into a collection of uniform cells, this provides an approximate representation of the workspace [9]. Figure 1 illustrates how grid division is performed, where shaded cells are an obstacle and normal cells are a potentially used cells. The centroid of each grid is considered a waypoint where the vehicle must pass.

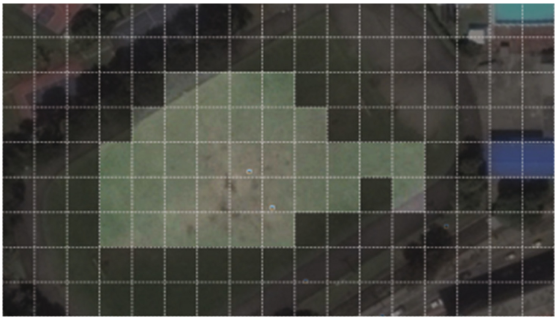


Fig. 1. Grid representation.

Most of the implemented paths in UAVs follow simple patterns that are easily configurable in ground stations such as zigzag. In this work, three offline algorithms based on approximate decomposition were implemented. This type of decomposition is usually used for aerial photography, where the dimensions of the acquired samples are required to be homogeneous. It should also be noted that this approach can deal with areas represented by non-convex polygons, which in some cases decomposition in cells cannot handle and must be transformed into convex spaces.

2.1 Wavefront Expansion

This path planning algorithm is in charge of create a distance function that starts in the point of arrival along the complete area that wants to be cover, thus giving a value in relation to the distance between each point and the goal. One way of looking at it is similar to the effect of throwing a rock into a lake, by this, a wave front is generated from the point of impact to its surroundings, hence its name [10]. With this algorithm, a matrix is obtained where each one of its cells have a weight, which increments as it gets far from the arrival point; this distance is measured in relation to the amount of cells that must pass to arrive at the goal.

The Fig. 2 shows an example of weight propagation using wavefront in a matrix with an arrival point G and four obstacles in four corners. In this example, the propagation is done with only four neighbors, that is, in each cell the only possible movements are North, South, East and West. However, a spread using 8 neighbors is also possible, as the Fig. 3 shows, considering the movements to North-East, North-West, South-East and South-West. The propagation in this case is observed as follows.

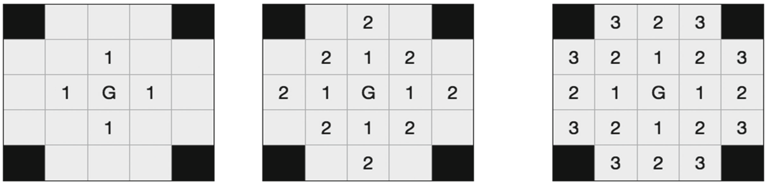


Fig. 2. Example of wavefront propagation.

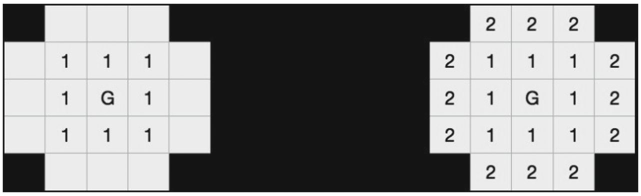


Fig. 3. Example of eight neighbors wavefront propagation.

As can be seen, propagation with eight neighbors is faster obtaining the final matrix, and it gives greater freedom in the movements than can make. Nevertheless, this approximation with a larger number of neighbors is not always possible.

With the resulting matrix, the next step is obtaining a coverage path for solution to the raised matrix. For this, it is necessary to move and get away from the arrival point as much as possible. This is achieved by always moving towards the neighbor with greater weight under some selection criteria like giving priority in the counterclockwise direction, as shown on Fig. 4. In case it finds a dead end, it proceeds to use a backtracking strategy to a cell where is possible to keep the aforementioned logic.

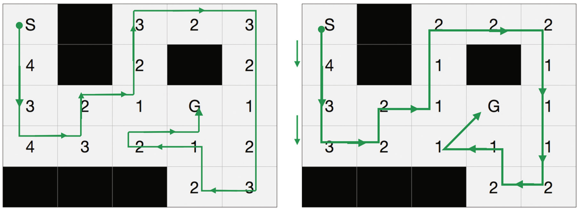


Fig. 4. Example of coverage path using wavefront with four and eight neighbors.

2.2 Spanning Tree Coverage

Spanning Tree Coverage (STC) is an algorithm that seeks to create an optimum coverage path which has a null number of revisits, forming a path that envelops an expansion tree generated by means of a scanning strategy [11]. This start from an array where each square cell has a D-size side. Each cell can describe the location of an obstacle or a possible point to visit. Subsequently, the cells are grouped into larger 2D-size nodes, thus creating an array of nodes, where each node contains four smaller cells, as shown on Fig. 5. The behavior of the array of nodes is very similar to its predecessor, being able to describe the location of obstacles or points to visit, but this time, a node becomes an obstacle as long as at least one cell within it is considered an obstacle.

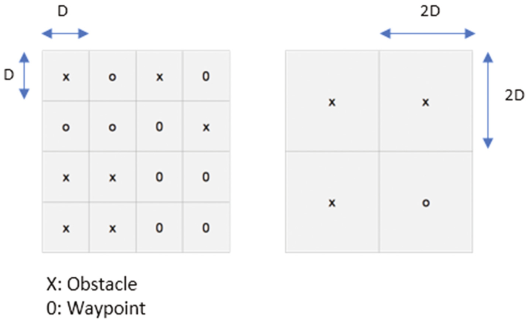


Fig. 5. Matrix of cells and grouping in matrix of nodes.

If a node is considered an obstacle, all of its cells immediately become obstacles. A node can have one of three possible states which depend on the states of the cells within the node, which are: *non-visited* in which no cell within the node has been visited, *partially visited* in which at least one cell has been visited, and *completely visited* in which all the cells inside the node have been visited. The coverage path is found as follows. The main objective in this algorithm is the exploration of new nodes. A starting point is placed inside the cell array and a scan is started giving priority to *non-visited* nodes. This algorithm evaluates the possibility of visiting an unseen cell that belongs to a *non-visited* node, if not possible, it continues the exploration within the current node until a cell that is neighbor of an unseen cell that belongs to a *non-visited* node is find. Scanning within the current node is done by extending the current movement of the vehicle, i.e. if it is moving from bottom to top, the scan must continue that movement and then prioritize in the counterclockwise direction. In case of not finding the possibility to visit a node *non-visited* after visiting the current node completely, proceed to visit the last node *partially visited*.

In the Fig. 6, the logic to find new nodes and scanning within a node is observed. A coverage path is found when the starting node goes into a *completely visited* state, this is achieved by following the previous logic as many interactions as necessary.

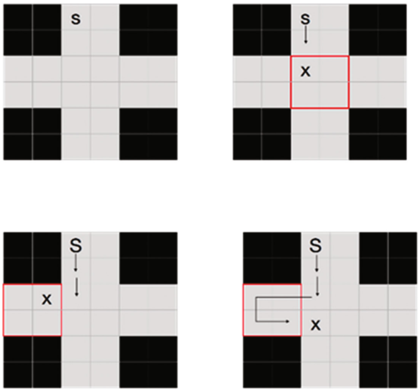


Fig. 6. Exploration of new nodes using STC.

Thus, through the algorithm a coverage path which has no revisits is found, the Fig. 7 illustrates this coverage path with a green line and the spanning tree with a red one.

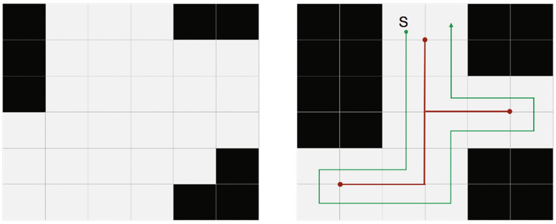


Fig. 7. Matrix and coverage path using STC.

2.3 Path Transform: Sunshine

Sunshine, is an algorithm that takes as its starting point the algorithm Path Transform, described in the article "Planning paths of complete coverage of an unstructured environment by a mobile robot" [12], which is created as an extension of the wavefront algorithm before mentioned [10]. Wavefront algorithm generates a weight function in relation to the distance from the current point to the point of arrival, instead, the Path Transform algorithm generates a weight function which is the result of summing the values given by a wavefront expansion and a cost function. In sunshine, the cost function is generated by obtaining the inverse of the distance to the nearest obstacle multiplied by a nonconformity factor. The expression that defines the weight value of each cell is described in the following matrix equation:

$$S(c) = wavefront(c) + \propto \frac{1}{obstacle} \qquad (1)$$

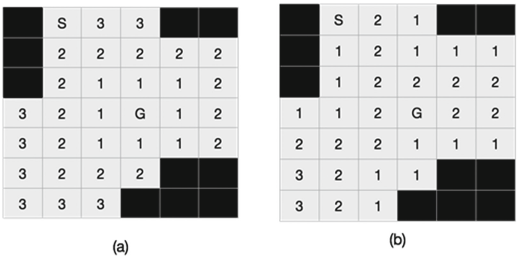
Where,

Wavefront(c), corresponds to the weight function that relates the distance of the current cell to the arrival cell.

Obstacle, distance from obstacle to arrival point.

$\propto$, the nonconformity factor determines the effort with which the resulting path will skirt the obstacles in the matrix.

Subsequently, the resulting matrix is born from the sum of the two previous, which are shown in Fig. 8(a) and (b).



■	S	3	3	■	■
■	2	2	2	2	2
■	2	1	1	1	2
3	2	1	G	1	2
3	2	1	1	1	2
3	2	2	2	■	■
3	3	3	■	■	■

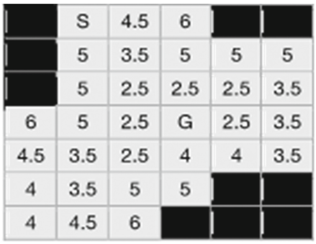
(a)

■	S	2	1	■	■
■	1	2	1	1	1
■	1	2	2	2	2
1	1	2	G	2	2
2	2	2	1	1	1
3	2	1	1	■	■
3	2	1	■	■	■

(b)

Fig. 8. Wavefront expansion (a) and wavefront expansion from obstacles (b).

Now with the resulting matrix (Fig. 9), this algorithm continues to obtain the coverage path as follows. Always move to the neighbor without visiting with the highest value (in this example the movement is done with 8 neighbors) according to a selection criterion like give priority in the counterclockwise direction and in case of find an exit alley, it is necessary to make use of a backtracking strategy until get to a cell with a neighbor without visiting.



■	S	4.5	6	■	■
■	5	3.5	5	5	5
■	5	2.5	2.5	2.5	3.5
6	5	2.5	G	2.5	3.5
4.5	3.5	2.5	4	4	3.5
4	3.5	5	5	■	■
4	4.5	6	■	■	■

Fig. 9. Matrix with Sunshine expansion.

As can be seen in the Fig. 10, the path obtained seeks to skirt obstacles while moving away from the point of arrival. An increase in the nonconformity factor changes the way in which the obtained route is behaved, bordering with more strength the present obstacles, and in the same way, when decreasing it the obtained path looks to a smaller extent to skirt the present obstacles.

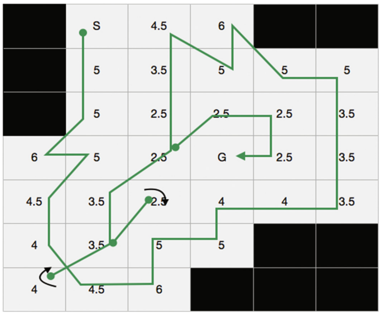
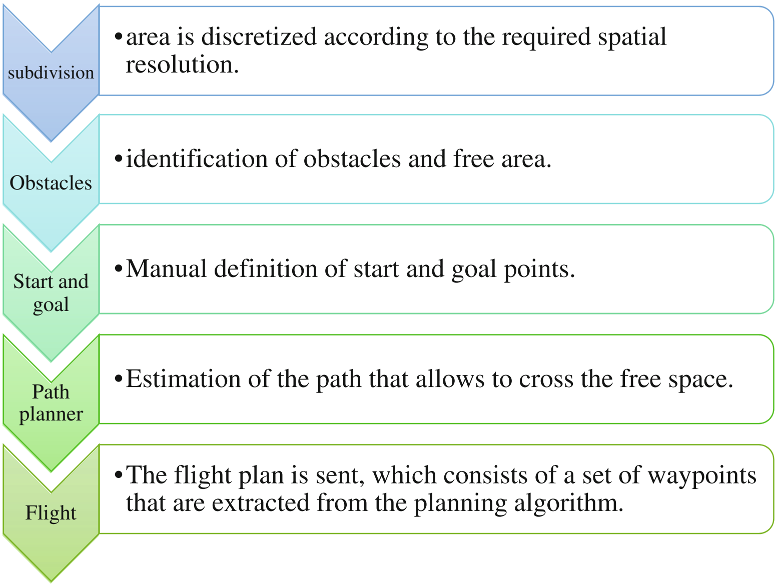


Fig. 10. Coverage path using Sunshine.

3 Testing Coverage Path

The case study is presented at the University Sports Campus (CDU), located at the University of Cauca, Popayán. Figure 11 shows the workflow of the proposed method. Initially the area map of interest is received, and the desired spatial resolution is defined. This is done by calculating the number of squares, in which the map will divided. Then, the possible obstacles are identified by calculating the histogram of each grid. Start and goal points are defined manually. This done, the implemented path-to-coverage scheduling algorithms calculate the path the vehicle should follow. Finally, the flight is done with the path obtained.



- area is discretized according to the required spatial resolution.
- identification of obstacles and free area.
- Manual definition of start and goal points.
- Estimation of the path that allows to cross the free space.
- The flight plan is sent, which consists of a set of waypoints that are extracted from the planning algorithm.

Fig. 11. Workflow

The simulation was carried out using the SITL (Software In The Loop) [13], which allows to run the ArduPilot autopilot on any platform. The program allows to configure a variety of parameters, corresponding to both the vehicle and battery, speed and flight height and the environment as the wind speed and its direction.

Testing Protocol

The test plan follows the next steps:

1. Get the coverage path for the selected zone with each algorithm.
2. Obtain a flight mission with each one of the coverage paths, defining waypoints which contain info about georeferenced coordinates and high of flight.
3. Configure the test environment, using SITL platform, defining parameters of the platform and the environment.
4. Set each flight mission into the test environment and run the test them using wind speed values of 2 m/s, 6 m/s and 8 m/s.
5. Recollect the logs of each flight mission and obtain info about battery spent, distance traveled, numbers of turns, cells visited and revisited.

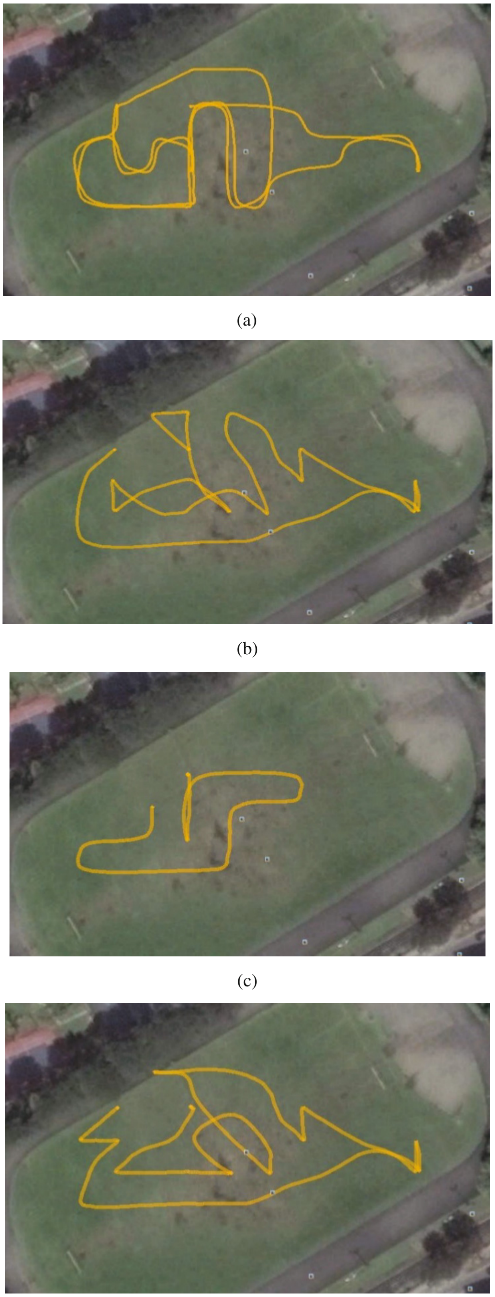
Testing Metrics

Three algorithms were implemented by comparing the following parameters that have been identified to take them into account in a coverage mission based on aerial robots.

- Nv Number of turns: Number of times the vehicle rotates through the z-axis.
- Cr Cells revisited: Number of times the vehicle covers a previously occupied cell.
- Cv Cells visited: Number of cells visited just one time.
- Pt Total length of the path traveled from the take-off point to the goal.
- T Coverage time to completion area of interest.
- Bt Battery percentage spent to complete the mission.

3.1 Obtained Results

In this simulation, it has been tested three different algorithms of route planning for coverage varying the wind speed, in order to compare the behavior of these as well as the variation of the mission time and the battery discharge, the velocities of wind were chosen 2, 6 and 8 m/s and the direction of the wind has been set from north to south, which is fixed in all simulations. The vehicle used was a quadcopter configured at a maximum travel speed of 5 m/s and a take-off/landing speed of 3 m/s. It is expected that the simulation results are reliable because this environment of simulation offers a complete simulation of an unmanned aerial system, it will be able to verify, plan tasks, manage faults and perform the subsequent validation. Figure 12 shows the path followed by the quadcopter in the simulation for each algorithm.



(a)

(b)

(c)

(d)

Fig. 12. Path for (a) Wavefront 4. (b) Wavefront 8. (c) Spanning tree coverage. (d) Sunshine.

Table 1 presents the number of revisits and turns that the platform performs for each of the implemented algorithms.

Table 1. Number of revisits and turns.

Algorithm	Cr	Cv	Nv 45°	Nv 90°	Nv 135°	Nv 180°
Wavefront 4	25	15	0	32	1	3
Wavefront 8	3	36	14	7	8	2
Spanning tree	0	23	0	10	0	0
Sunshine	3	36	10	7	9	2

For the simulation process the wind speed was varied, taking values of 2, 6 and 8 m/s and measuring the variation of the path traveled, the total time to complete the mission and the percentage battery discharge. The results obtained are presented in Table 2.

Table 2. Path, battery and time of mission comparison for different wind values.

Wind	Algorithm	Pt (km)	T (min)	Bt (%)
2 m/s	Wavefront 4	0,610	9:44	47
	Wavefront 8	0,449	7:06	40
	Spanning tree	0,204	4:44	23
	Sunshine	0,452	8:02	41
6 m/s	Wavefront 4	0,612	10:09	50
	Wavefront 8	0,453	8:22	42
	Spanning tree	0,206	4:45	24
	Sunshine	0,462	9:27	43
8 m/s	Wavefront 4	0,619	13:14	51
	Wavefront 8	0,454	9:38	43
	Spanning tree	0,221	5:22	25
	Sunshine	0,467	9:51	44

Below are graphs of the total path traveled by the vehicle versus battery discharge for each one of the algorithms at different wind speeds in order to analyze how the energy consumption and the total path length vary according to this (Figs. 13, 14, 15 and 16).

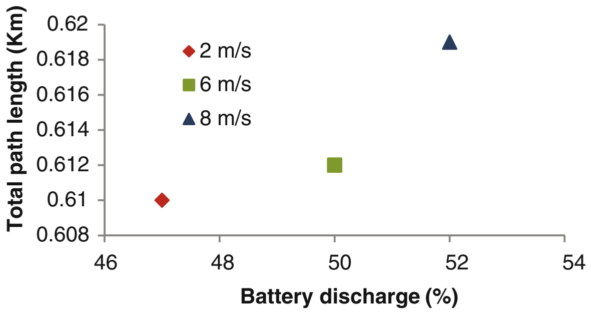


Fig. 13. Total path traveled by the vehicle versus battery discharge for wavefront 4 with wind speeds of 2, 6 and 8 m/s.

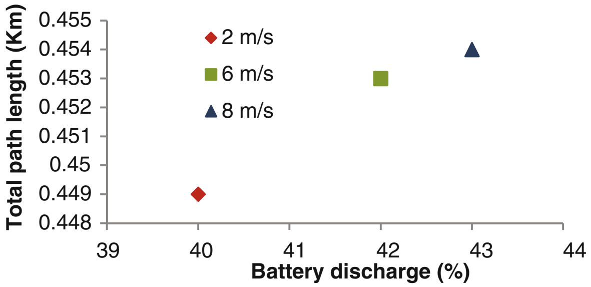


Fig. 14. Total path traveled by the vehicle versus battery discharge for wavefront 8 with wind speeds of 2, 6 and 8 m/s.

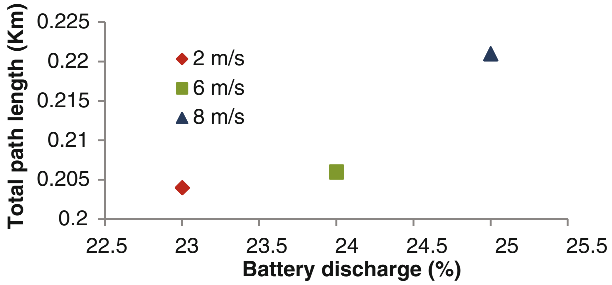


Fig. 15. Total path traveled by the vehicle versus battery discharge for Spanning tree coverage with wind speeds of 2, 6 and 8 m/s.

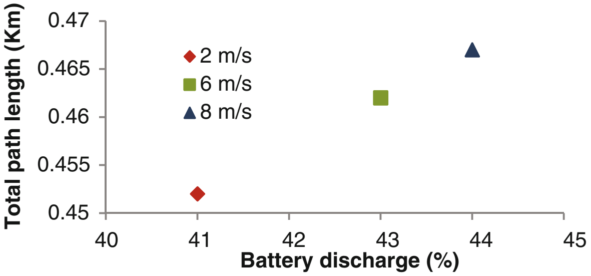


Fig. 16. Total path traveled by the vehicle versus battery discharge for Sunshine with wind speeds of 2, 6 and 8 m/s.

4 Conclusions

In this article, the problem of finding a suitable coverage path for the photogrammetric survey of an interest area through a quadcopter using four algorithms, wavefront 4 neighbors, wavefront 8 neighbors, STC and Sunshine is addressed. The need to find optimal coverage paths is presented as a solution to the limiting of the great energy discharge that these vehicles have. Better algorithms mean less flight time, less distance traveled, fewer turns, and lower energy use. This is evident by comparing directly the wavefront algorithm of 4 neighbors with its counterpart of 8 neighbors, showing in the last one, a lower energy use quite considerable in all cases. In this comparison, wavefront algorithm of 8 neighbors and Sunshine algorithm have very similar values of flight time, distance traveled and energy use, being in the latter a little larger due to a greater distance traveled and a greater number of turns. But, in case these improvements are not sufficient, STC can be used, an algorithm which presents a tremendous reduction in energy waste, being in some cases almost half of the waste presented with other algorithms, at the cost of losing one considerable coverage. The conditions of the environment and the terrain also affect the final yield of the quadcopter, being the wind factor the one to evaluate in this article. Here, we observe how the intensity of the wind is able to affect the evaluation parameters mentioned above, being at higher intensity a worse performance, thus turning the wind as a factor to be taken into account in the flight missions. For these reasons, it is necessary to find a unified strategy capable of finding an optimum coverage route that takes into account the variables of the environment and geography, in order to minimize the negative effects present in each flight plan.

Acknowledgement. The authors are grateful to the University of Cauca and its Telematics Engineering Group (GIT), the Colombian Administrative Department of Science, Technology and Innovation (Colciencias), AgroCloud project of The Interinstitutional Network of Climate Change and Food Security of Colombia (RICCLISA) for supporting this research.

References

1. Galceran, E., Carreras, M.: A survey on coverage path planning for robotics. Robot. Auton. Syst. **61**(12), 1258–1276 (2013)
2. Zhang, C., Kovacs, J.M.: The application of small unmanned aerial systems for precision agriculture: a review. Precis. Agric. **13**(6), 693–712 (2012)
3. McBratney, A., Whelan, B., Ancev, T., Bouma, J.: Future directions of precision agriculture. Precis. Agric. **6**(1), 7–23 (2005)
4. Pajares, G.: Overview and current status of remote sensing applications based on Unmanned Aerial Vehicles (UAVs). Photogramm. Eng. Remote Sens. **81**(4), 281–329 (2015)
5. Honkavaara, E., Kaivosoja, J., Mäkynen, J., Pellikka, I., Pesonen, L., Saari, H., Rosnell, T.: Hyperspectral reflectance signatures and point clouds for precision agriculture by light weight UAV imaging system. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. I-7, pp. 353–358 (2012)
6. Meneses, V.A.B., Téllez, J.M., Velasquez, D.F.A.: Uso de drones para el analisis de imágenes multiespectrales en agricultura de precisión. limentech, Ciencia y Tecnología Alimentaria **13**(1) (2015)
7. Vasquez-Gomez, J.I., Gomez-Castañeda, C., De Cote, E.M., Herrera-Lozada, J.C.: Multirotor UAV coverage planning under wind conditions. In: Mechatronics, Electronics and Automotive Engineering (ICMEAE), pp. 32–37 (2016)
8. Gonçalves, J.A., Henriques, R.: UAV photogrammetry for topographic monitoring of coastal areas. ISPRS J. Photogramm. Remote Sens. **104**, 101–111 (2015)
9. Colomina, I., Molina, P.: Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS J. Photogramm. Remote Sens. **92**, 79–97 (2014)
10. Valente, J., Sanz, D., Del Cerro, J., Barrientos, A., de Frutos, M.Á.: Near-optimal coverage trajectories for image mosaicing using a mini quad-rotor over irregular-shaped fields. Precis. Agric. **14**(1), 115–132 (2013)
11. Guruprasad, K.R., Ranjitha, T.D.: ST-CTC: a spanning tree-based competitive and truly complete coverage algorithm for mobile robots. In: Proceedings of the 2015 Conference on Advances in Robotics (2015). Article No. 43
12. Zelinsky, A., Jarvis, R.A., Byrne, J.C., Yuta, S.I.: Planning paths of complete coverage of an unstructured environment by a mobile robot. In: Proceedings of International Conference on Advanced Robotics, vol. 13, pp. 533–538 (1993)
13. Ardupilot. http://ardupilot.org/dev/docs/sitl-simulator-software-in-the-loop.html. Accessed 13 July 2017

Impact of Climate Change on Soil Organic Carbon Content on Agricultural Soils of Mexico

Jesús D. Gómez-Díaz(✉), Alejandro I. Monterroso-Rivas,
Lizeth M. Lechuga-Gayosso, Antonio R. Arce-Romero,
and Patricia Ruiz-Gracia

Departamento de Suelos, Universidad Autónoma Chapingo,
Carretera Mexico - Texcoco Km 38.5, 56230 Chapingo,
Estado de México, Mexico
dgomezl059@yahoo.com.mx, aimrivas@correo.chapingo.mx,
liz_260291@hotmail.com, antonio.arce93@gmail.com,
cometapatl8@hotmail.com

Abstract. Changes in soil organic carbon (SOC) content for Mexico's agricultural land were estimated using the carbon accounting model of the IPCC guidelines (IPCC 2006) and the FAO soil organic matter decline model (Ortiz et al. 1994) for the short- and medium-term time horizons for a RCP of 8.5 W/m^2 of radiative forcing with the HADGEM and GFDL models. All models show a significant decrease in the surface with higher content of organic matter while the soils with lower organic matter content increase considerably. The variation between the models is relatively low among them, associated to both establishing an increase in temperature and decrease in precipitation which are the fundamental factors that define the mineralization of the SOC in the methods used.

Keywords: Climate change · Soil degradation · Mineralization of soil organic matter

1 Introduction

One of the main concerns of our time is food security to meet the growing demand for food as the number of people on the planet increases. The increase in population and rapid infrastructure development are expected to increasing pressures on ecosystems for food production, leading to rapid rise in the release of CO$_2$ into atmosphere and decline in atmospheric carbon sequestration (Mohanraj et al. 2011). Facing this scenario will increase the pressure on natural resources and thereby the emission of greenhouse gases (Fearnside 2004).

Soil Organic Carbon (SOC) is very important in the global C cycle, since it constitutes the largest terrestrial reservoir of this element (Chabra et al. 2003). Soils store at least three times more carbon (in the form of Soil Organic Carbon) than that found in the atmosphere or living plants (Schmidt et al. 2011). This reservoir of organic carbon is sensitive to changes in climate and the environment, although the interrelationships

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_5

between SOC and climate are not yet fully understood, but one way to explore these relationships is to use numerical models associated with the cycle of the organic carbon of the soil and the climate considering other factors of the natural environment and those associated with the use and management of the soils. With these results, management options that include best management practices can be evaluated to reverse the degradation of soils and increase their productive capacity. The C stored in the soil is highly reactive and dynamic (Dumanski 2004). The cultivation of virgin lands or the conversion of natural ecosystems to agriculture leads to the loss of soil C with significant greenhouse gases (GHG) emissions into the atmosphere. This loss of C is accentuated by deforestation, burning of biomass, drainage of wetlands, tillage and removal of crop residues (Lal 2002).

The influence of soils on the structure and operation of ecosystems depends largely on the organic C (Cleve and Powers 1995). Understanding the long-term changes in SOC is important because it directly affects soil quality (Lal 2002), by influencing soil porosity and water retention and serving as the reservoir of nutrients for plants, especially N, P, S and K (Bandarenayeke et al. 2003). Consequently, promote carbon sequestration by soil is an effective strategy to reduce atmospheric CO_2 and improve soil quality (Lal et al. 1998; Lal 1999). Failure to follow proper practices, loss of C is accentuated and thereby decreasing the productive capacity of soils. Because of the importance of SOC, precise estimations are required in different types of land uses, especially data in long time series of different land uses and management to evaluate and monitor soil C stocks over time (Beets et al. 2002).

Regional assessment of change in soil organic carbon stocks due to land-use change are essential for supporting policy and management decisions related to greenhouse gas emissions and mitigation through carbon sequestration in soils (Mishra et al. 2012). Soils at a particular location can be source or sink of atmospheric CO_2, depending on previous and existing land use and management (Singh and Lal 2005). The impact of agricultural management on SOC stocks varies, depending in part on climate conditions, which influence the plant and soil processes driving soil organic matter dynamics (Ogle et al. 2003). The possibility of a reduction in global soil carbon under climate warming has been raised by several studies, in this situation, increased heterotrophic respiration because of rising temperatures during the 21st century exceeded enhanced biospheric uptake as a result of elevated CO_2 levels (Mishra et al. 2012). As a result, the rate of increase of atmospheric CO_2 and hence the rate of climate change was accelerated (Jones et al. 2003).

In Mexico, of the total area of the country, 13.6% is reported as agricultural land corresponding to 26.7 million ha, of which an average of 22.1 million ha are planted every year, where 6.1 million are irrigated and 16 million of ha of rainfall (SAGARPA-SIACON 2014). In most of the agricultural land one of the common land management practice include the removal of crop residues or crop residues burning, so, those practices let these soils as net CO_2 emitters into the atmosphere and thus contribute to climate change. And also, the reduction of the soil productive capacity. That is why it is important to carry out studies that estimate the dynamics of organic carbon in the agricultural soils of the country where food and other important products are obtained. To seek food security for the country's population, specifically considering climate change scenarios it is important to implement a strategy that promote the increase of soil productivity that is linked to soil carbon sequestration. Knowing the estimates of the decrease of the organic matter of the agricultural soils and the eco-climatic conditions where they are located, allows to establish general guidelines for the rehabilitation of the same ones through the implementation of good management practices adjusted to the land sustainable management, and promote the mitigation and adaptation to climate change.

2 Methodology

2.1 Database of Precipitation, Temperature and Monthly Potential Evapotranspiration for the Climate Influence Areas of the Country for the Base and the Climate Change Scenarios

Data bases for regional climatic variables scale 1: 250,000 were generated for the agricultural areas of Mexico, with updated information reported on the different meteorological stations of the country in the CLICOM system (WMO 2007). The monthly and annual averages of the temperature and precipitation were obtained for all the selected meteorological stations, these values were used for the elaboration of the isohyets and isothermals maps and the databases in the Climatic Influence Areas as described by Gomez and Monterroso (2012) in the project on Delimitation of the Arid, Semi-Arid and Subhumid Zones of the Country at regional level. The changes in temperature, precipitation and monthly potential evapotranspiration associated to climate change were estimated using the HADGEM and GFDL General Circulation Models (GCM), for the short and medium-term horizons (2015–2039, 2045–2069, respectively) for the Representative Concentration Pathways (RCP) of 8.5 W/m^2 radiative forcing. Subsequently, the change estimated for each described climate change situation were applied to the average precipitation and temperature and relative humidity data of the baseline scenario in the Climate Influence Areas. Cartographic representation of climate variables under the climate change scenarios estimated for the agricultural soils of Mexico at scale 1: 250 000 was done with the use of Arc Gis.

2.2 Estimation of the Mineralization Dynamics of Soil Organic Matter in the Agricultural Land of Mexico Considering the Climate Change Scenarios

The National Institute of Statistics and Geography (INEGI) database that reports the soil organic matter content for the agricultural lands of Mexico was used, and the reported data of soil organic matter content were overlapped with the climatic databases of the Climate Influence Areas, where the data obtained from precipitation, temperature and potential evapotranspiration for the base scenario and under climate change scenarios. The carbon accounting model established in the IPCC guidelines (2006) and the FAO model of climate change were applied for all the areas using the soil and climate variables information to estimate the decrease of soil organic matter in the land degradation process (Ortiz et al. 1994). In both cases it was considered that in agricultural practices there is no incorporation of crop residues into the soil, nor the application of organic amendments such as compost or manure.

2.2.1 The Carbon Accounting Model of the IPCC (2006), Is Represented in the Following Formula

$$\Delta C_{mineral} = \frac{\left(SOC_0 - SOC_{(0-T)}\right)}{D} \tag{1}$$

$$SOC = \sum\nolimits_{c,s,i} \left(SOC_{REF} * F_{LU} * F_{MG} * F_I * A\right) \tag{2}$$

where

$\Delta C_{mineral}$ annual change in organic carbon stocks of mineral soils, Mg of C ha^{-1} year^{-1}.

SOC_0 Stocks of organic carbon in the soil in the last year of an inventory period, Mg of C.

$SOC_{(0-T)}$ Stocks of organic carbon in the soil at the beginning of an inventory period, Mg of C.

D 24 years for the short-term time horizon (from 2015 to 3039).

D 30 years for the medium-term time horizon (from 2039 to 2069).

c represent the climatic zones.

s soil types.

i management systems, in this case is agricultural land.

SOC_{REF} Reference carbon stocks, Mg of C ha^{-1}.

F_{LU} Factor of change of carbon existences for the systems of land use.

F_{MG} Factor of change for the specific land use.

F_I Factor of change of the carbon stock for the contribution of organic matter.

A Land area of the stratum estimated.

The SOC of reference was the one reported by INEGI for the different soil units considering the natural vegetation with which the agricultural areas of the country are associated, that is the natural vegetation prior to the change of land use to agricultural activities. The SOC_0 was the one that was determined in the inventory made by INEGI and reported for the agricultural areas of the country.

The estimation of the SOC content for the country's agricultural soils was performed for the two-time scenarios, for the short and the medium term.

2.2.2 FAO Model of Soil Organic Matter Decline in the Land Degradation Process (Ortiz et al. 1994)

As already noted, in the agricultural land there are generally no inputs of organic matter, so that in calculating biological degradation from the reduction of soil organic matter by annual mineralization of this material, applying the approach presented by (Ortiz et al. 1994), the mineralized organic matter was subtracted from the previous soil organic matter content, so the calculation for the time series considered was obtained. The SOC of each time series, to 2039 and 2069 for each of the models (HADGEM and

GFDL), for the RCP 8.5 W/m^2, were estimated. At the end, four country-level estimates of SOC were generated for Mexican agricultural soils.

The model considers the following factors:

(a) Climatic Index

In general, the climatic index used to determine the rate of mineralization of soil organic matter in which resistant humus is considered was the Koepf index (1953). With a rate of decomposition as expressed in the following equation and reported in Mg of Organic Matter per hectare per year:

$$K_2 = \frac{1}{12}\sum\nolimits_{12}^{1} e^{0.1065*t}\left(\frac{P}{PET}\right) \tag{3}$$

where:

K_2	Climatic index of annual biological degradation.
P	Mean monthly precipitation mm.
PET	Monthly Potential Evapotraspiration.
t	Mean monthly Temperature °C.
	When P > PET, P/PET = 1; and for t < 0, t = 0.

With the soil organic matter contents estimated for each time period and climatic change scenarios, the cartographic representation of these values for the agricultural soils of Mexico at a scale of 1: 250,000 was done using Argis.

(b) Edafic index

It considers that the decrease of the soil organic matter depends on the texture of the soil, the nature of the organic matter, the pH, the percentage of calcium carbonate and the water content in the soil above the field capacity.

For regional studies such as that proposed in this study, the edaphic index is determined by evaluating the effect of texture and calcium carbonate content (Ortiz et al. 1994). The estimation of this index is obtained from the relationships shown in Table 1.

Table 1. Factors associated with the texture and $CaCO_3$ content to estimate the edaphic index

Textural class	Coarse	Medium	Fine
Factor associated with the soil texture	1.5	1.0	0.5
Percentage of $CaCO_3$	0	5	15
Factor associated with $CaCO_3$ content	1	0.8	0.6

For this index, we consider the textural class reported by INEGI for each of the agricultural soil units, as well as the calcium carbonate content reported for these soils.

The mineralization value of organic matter per hectare per year was estimated as follows:

$$\text{Mineralization of soil organic matter} = K_2 \times \textit{Factor asscociated with texture} \times \textit{Factor asscociated with } CaCO_3 \textit{ content} \tag{4}$$

The value of soil organic matter content for a given year was equal to: (Organic matter content of soil in Mg ha^{-1} in the previous year) – (Organic matter of the mineralized soil in the year in question in Mg ha^{-1}).

The results were compared with those obtained using the formula proposed by the IPCC (2006) in which it also assumes that there are no inputs of organic matter in the agricultural land. Cartographic representation of these values was made for agricultural soils of Mexico at scale 1: 250,000 with the use of Arc Gis.

3 Results

In general, the agricultural soils of Mexico have low SOC contents, this is associated to the fact that the climatic conditions and soil properties do not favor the accumulation of organic matter due to the low production of biomass in the dry regions and by the management practices of using the crop residues to feed livestock or burning crop residues. The higher levels of organic carbon in the country's agricultural soils are associated with Andic soil properties and in the areas of accumulation of organic materials from which soils are developed. The variation of the SOC of Mexican agricultural land estimated from the application of the carbon accounting model established in the IPCC (2006) guidelines for the short and medium term, and the carbon mineralization model determined using the FAO approach described by Ortiz et al. (1994) for a RCP 8.5 W/m^2 with the GFDL and HADGEM models. Table 2 shows the short term scenarios and Table 3 the medium term scenarios.

For the short-term time scenario compared to current conditions, the lowest SOC class (0–50 Mg ha^{-1}) have considerable surface increases, for the IPCC guidelines it increases by 10.97% of the total of the area of the country's agricultural soils, whereas for the GFDL and HADGEM models with the FAO methodology the increase is 29.13% and 28.69%, respectively. For the 50–100 Mg ha^{-1} class, the IPCC guidelines increase by 15.2% of the total land area of the country's agricultural land, while for the GFDL and HADGEM models it is 2.03% and 2.29%, respectively. The results of the IPCC guidelines in the class of 100–150 Mg ha^{-1} of SOC have practically no change in the surface and for the subsequent classes with the highest SOC content there are decreases in the surface. For the GFDL and HADGEM models from the class of 100–150 Mg ha^{-1} of SOC and the following with higher SOC content, there are considerable decreases.

Table 2. Soil Organic Carbon content of Mexican agricultural land estimated from IPCC guidelines (2006) and mineralization with GFDL and HADGEM models for RCP 8.5 W/m^2 using FAO methodology (Ortiz et al. 1994) for the short-term scenarios.

Classes by content of SOC (Mg Ha^{-1})	Actual condition	IPCC 2015–2039	GFDL 2015–2039	HADGEM 2015–2039
	Area in thousands of hectares Percentage of total agricultural land			
0–50	11,753 (36.2)	15,313 (47.2)	21,205 (65.4)	21 061 (64.9)
50–100	4,652 (14.3)	9,584 (29.5)	5 312 (16.4)	5 397 (16.6)
100–150	4,479 (13.8)	4,578 (14.1)	2 871 (8.9)	2 559 (7.9)
150–200	6,583 (20.3)	1,428 (4.4)	1 300 (4.0)	1 665 (5.1)
200–250	1,687 (5.2)	737 (2.3)	302 (0.9)	305 (0.9)
250–300	996 (3.1)	431 (1.3)	161 (0.5)	170 (0.5)
>300	2,297 (7.1)	376 (1.2)	1 295 (4.0)	1 289 (4.0)

Table 3. Soil Organic Carbon content of Mexican agricultural land estimated from IPCC guidelines (2006) and mineralization with GFDL and HADGEM models for RCP 8.5 W/m^2 using FAO methodology (Ortiz et al. 1994) for the medium-term scenarios.

Classes by content of SOC (Mg Ha^{-1})	Actual condition	IPCC 2039–2069	GFDL 2039–2069	HADGEM 2039–2069
	Area in thousands of hectares Percentage of total agricultural land			
0–50	11,753 (36.2)	27,738 (85.5)	28,200 (86.9)	28 138 (86.7)
50–100	4,652 (14.3)	3,755 (11.6)	2 166 (6.7)	2 221 (6.8)
100–150	4,479 (13.8)	608 (1.9)	742 (2.3)	738 (2.3)
150–200	6,583 (20.3)	4 (0.0)	218 (0.7)	0 (0.0)
200–250	1,687 (5.2)	99 (0.3)	265 (0.8)	264 (0.8)
250–300	996 (3.1)	238 (0.7)	269 (0.8)	509 (1.6)
>300	2,297 (7.1)	5 (0.0)	585 (1.8)	577 (1.8)

For the medium-term scenario, the decrease of the SOC is greater compared to the current conditions, the lower class of SOC (0–50 Mg ha^{-1}) for the IPCC guidelines the increases of the surface is 49.27% of the total area of the country's agricultural soils, whereas for the GFDL and HADGEM models with the FAO methodology the increase is 50.69% and 50.5%, respectively. For all classes of more than 50 Mg ha^{-1} of SOC, both the IPCC and the GFDL and HADGEM models show considerable decreases.

Figure 1 shows graphically the surface area (thousands of ha) of the different SOC classes of the agricultural soils of Mexico for the current condition and the different scenarios of climate change estimated with the GFDL and HADGEM models for the CPR of 8.5 for short and medium term with the FAO methodology and for those estimated time scenarios applying the IPCC guidelines (2006).

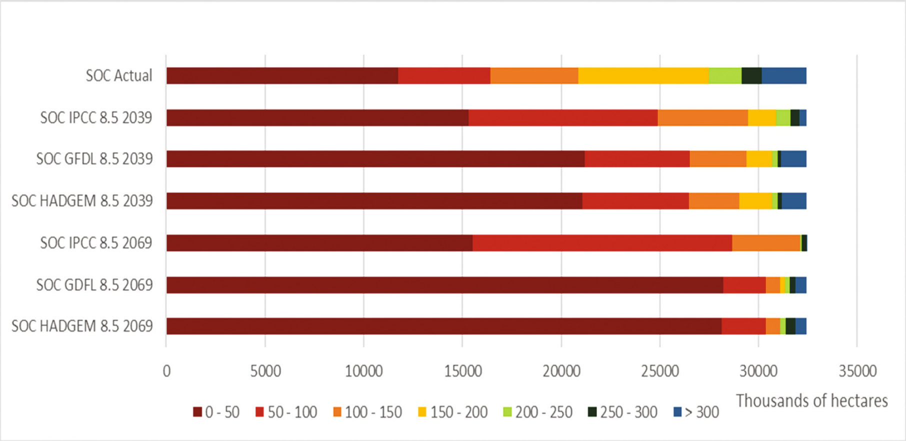


Fig. 1. Area in thousands of hectares of the SOC-content classes of Mexican agricultural soils for the current condition and short- and medium-term climate change scenarios estimated with the GDFL and HADGEM models for RCP 8.5 and with the IPCC guidelines.

Figure 2 shows the maps with the distribution of SOC content classes of Mexican agricultural soils for current conditions and short and medium-term climate change scenarios for the two models and the IPCC guidelines. Soils in agricultural regions with dry climates generally have low organic carbon contents, as well as those with coarse textures, the greatest changes in SOC decrease occur in soils with medium and coarse textures and with subhumid and humid climates in where the mineralization of organic matter is greater do to the increase of the soil temperature and the availability of soil moisture. The mayor changes expressed in the different maps is in the medium-term scenario for estimates with the FAO method described by Ortiz et al. (1994), for a RCP of 8.5 with the GFDL and HADGEM models, and is accentuated in these agricultural areas.

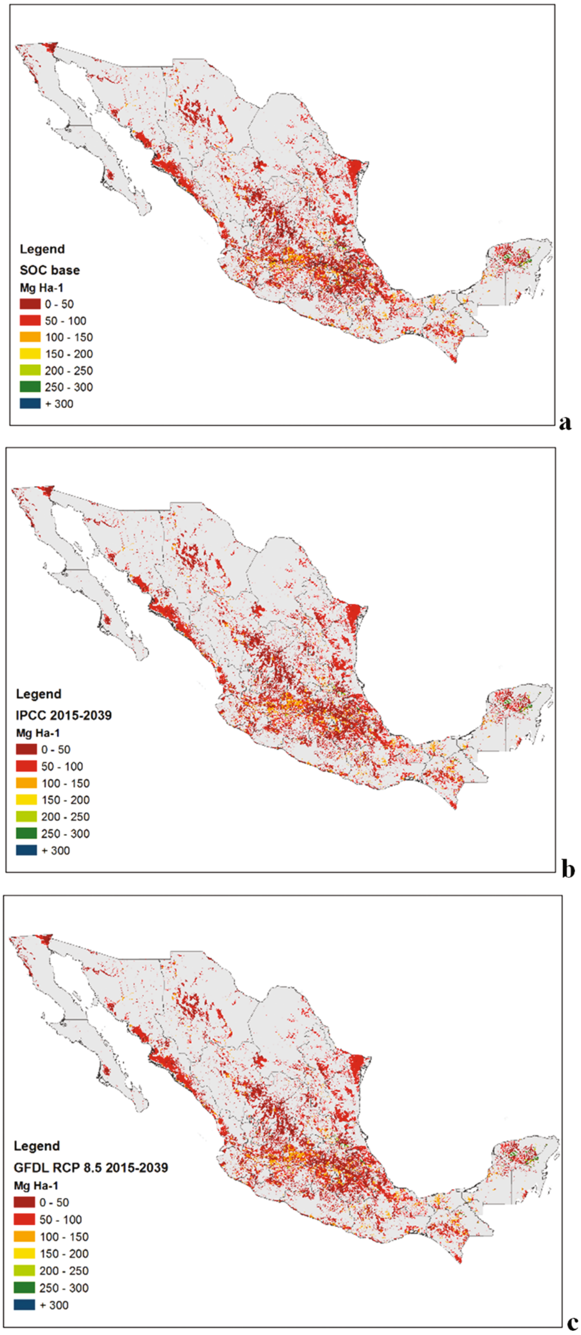


Fig. 2. Distribution of SOC classes for Mexican agricultural soils (a) current scenario, (b) IPCC 2015–2039; (c) GFDL 2015–2039; (d) HADGEM 2015–2039; (e) IPCC 2045–2069; (f) GFDL 2045–2069; (g) HADGEM 2045–2069

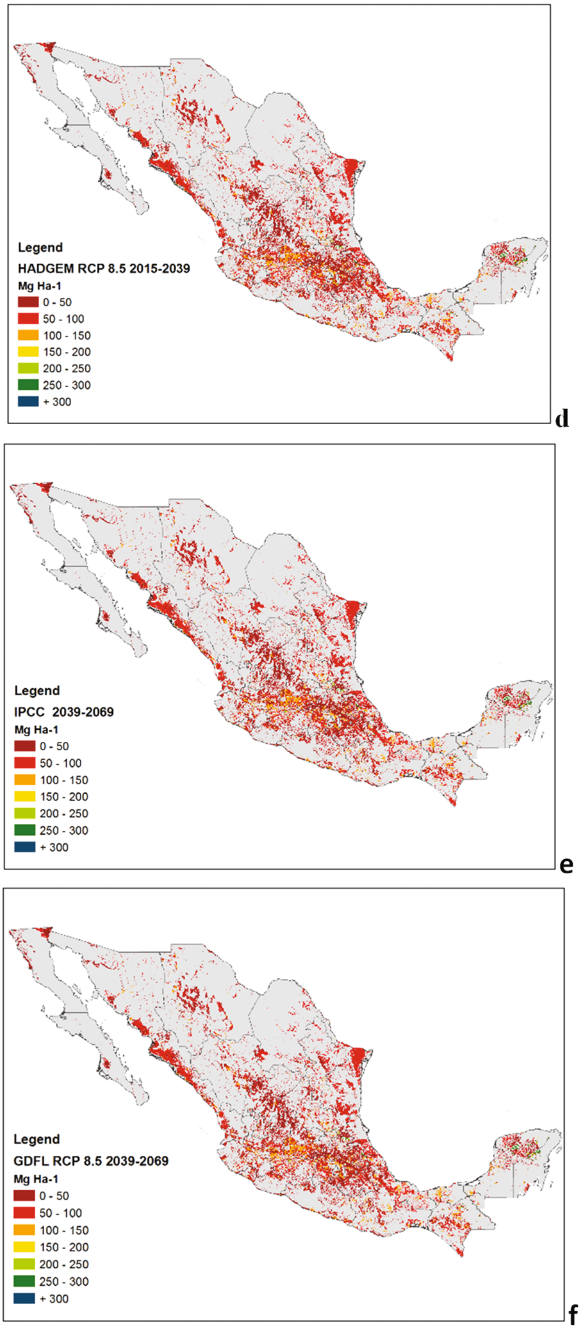


Fig. 2. (*continued*)

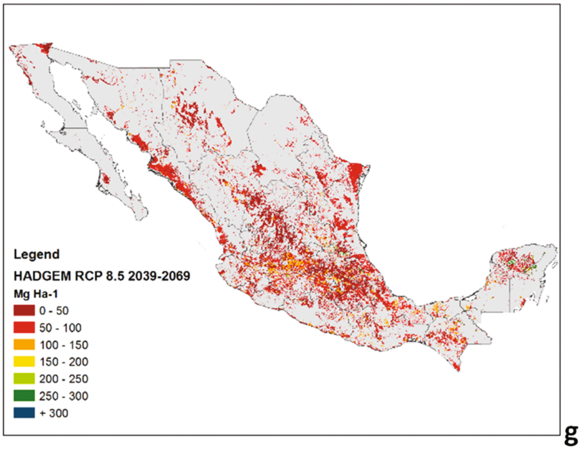


Fig. 2. (*continued*)

4 Conclusions

In the climate change scenarios for the application of the carbon accounting model established in the IPCC (2006) guidelines for the short and medium term, and also for the carbon mineralization model using the FAO method for a RCP of 8.5 W/m^2 also for the short and medium-term scenarios using the GFDL and HADGEM models, very high reductions are estimated in the content of the SOC and increase as the time period is considered higher. In the short term, there is a greater reduction in the FAO model estimates with the two GCM than in the IPCC guidelines, but in the medium term there is no difference between the estimates with the two models used in the FAO approach and the IPCC guidelines.

The variation in applying the GFDL and HADGEM models is relatively low among them, associated with both establishing an increase in temperature and decrease in precipitation, specially in the HADGEM, which are the fundamental factors that define the mineralization in the method used. It is important to note that the decrease of the SOC is slower the less the amount of this component in the soil, thus can be associated that the SOC is in the most recalcitrant forms of organic matter and is more difficult to decompose by the soil microorganisms.

References

Bandaranayake, W., Qian, Y.L., Parton, W.J., Ojima, D.S., Follett, R.F.: Estimation of soil organic carbon changes in Turfgrass systems using the Century Model. Agron. J. **95**, 558–563 (2003)

Beets, P.N., Oliver, G.R., Clinton, P.W.: Soil carbon protection in podocarp/hardwood forest, and the effects of conversion to pasture and exotic pine forest. J. Env. Pollut. **166**, 563–573 (2002)

Chabra, A., Palria, S., Dadhwal, V.K.: Soil organic carbon pool in Indian forests. Forest Ecol. and Manag. **173**, 187–199 (2003)

Cleve, K.V., Powers, R.F.: Soil carbon, soil information and ecosystem development. In: McFree, W.W., Kelly, M. (eds.). Carbon Forms and Functions in Forest Soils. SSSA, U.S.A, pp. 155–199 (1995)

Dumanski, J.: Carbon sequestration, soil conservation, and the Kyoto protocol: summary of implications. Clim. Change **65**, 255–261 (2004)

Fearnside, P.M.: Are climate change impacts already affecting tropical biomass? Glob. Environ. Change **14**, 299–302 (2004)

Gómez-Díaz, J., Monterroso-Rivas, A.I.: Actualización de la delimitación de las zonas áridas, semiáridas y sub-húmedas secas de México a escala regional. Reporte final de proyecto de investigación Fondo Semarnat-Conacyt. Texcoco, Universidad Autonoma Chapingo, México (2012)

IPCC (Intergovermental Panel on Climate Change): Climate Change. The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC 2006, 1st edn. p. 1056. Cambridge University Press (2006)

Jones, C.D., Cox, P.M., Essery, R.L.H., et al.: Stong carbon cycle feedbacks in climate model with interactive CO_2 and sulphate aerosols. Geophys. Res. Lett. **30**, 32:1–32:4 (2003)

Lal, R.: Soil management and restoration for carbon sequestration to mitigate the accelerated greenhouse effect. Prog. Env. Sci. **1**(4), 307–326 (1999)

Lal, R.: Why carbon sequestration in agricultural soils. In: Kimble, J.M., Lal, R., Follett, R.F. (eds.) Agricultural Practices and Policies for Carbon Sequestration in Soil, pp. 21–30. Lewis Publishers, Boca Raton (2002)

Lal, R., Kimble, J.M., Follet, R.F., Cole, C.V. (eds.): The potential of U.S. cropland to sequester carbon and mitigate the Greenhouse effect. J. Soil Water Conserv. **54**, 374–381 (1998)

Mohanraj, R.J., Jothivel, S., Dhanakumar, K.S.: Carbon stock in Kolli forest, Eastern Ghats (India) with emphasis on aboveground biomass, litter. Woody Debris Soils **4**, 61–65 (2011). iForest

Mishra, U., Torn, M.S., Masanet, E., Ogle, S.M.: Improving regional soil carbon inventories: Combining the IPCC carbon inventory method with regression kriging. Goderma **189–190**, 288–295 (2012)

Ogle, S.M., Breidt, F.J., Eve, M.D., Paustian, K.: Uncertainty in estimating land use and management impacts on soil organic carbon storage for US agricultural lands between 1982 and 1997. Glob. Change Biol. **9**, 1521–1542 (2003)

Ortiz, S., Anaya, G., Berg, W.: Evaluación, Cartografía y políticas preventivas de la degradación de la tierra. Primera Edición. Comisión Nacional de las Zonas Áridas. Colegio de Postgraduados. Universidad Autónoma Chapingo, México (1994)

SAGARPA-SIACON: Producción Agropecuaria y Pesquera. Servicio de información agroalimentaria y pesquera, México (2014)

Schmidt, M.W., Torn, M.S., Abiven, S., Dittmar, T., et al.: Persistence of soil organic matter as an ecosystem property. Nature **478**(7367), 49–56 (2011)

Singh, B.R., Lal, R.: The potential of soil carbon sequestration through improved management practices in Norway. Environ. Dev. Sustain. **7**, 161–184 (2005)

WMO: Base de Datos Climatológica CLICOM. World Meteorological Organization, México (2007)

Fertilization Strategies in Sugar Cane Crops in the Cauca Valley: A Review

Álvaro Pachón[1], Gonzalo Llano[1](✉), Luis Munera[1], Camilo Barrios[1], Claudia Lubo[1],
Julián Borrero[2], and Gonzalo Calderon[2]

[1] Grupo de Informática y Telecomunicaciones i2T, Universidad Icesi,
18th street #122-135, Valle del Cauca, Cali 760001, Colombia
{alvaro,gllano,lemunera,claudia.lubo}@icesi.edu.co,
groovy.kmilo@gmail.com

[2] ALC Construcciones y Riegos SAS, 52th AN street #2GN-79, Valle del Cauca, Cali, Colombia
jborrero.agro@yahoo.com, conservaslacalma@gmail.com

Abstract. Our article analyzes several research approaches about fertilization strategies, considering the soil-plant-climate ecosystem. We used the systematic literature review methodology, aiming to answer the research questions raised in our study. Yield and productivity of sugar cane crops is closely related to quality and efficiency of fertilization strategies. We approach fertilization strategies in sugar cane crops analyzing old and new methodologies, recognizing their respective strengths and weaknesses, and looking for new improvements on them to prevent the environmental impact caused by crop fertilization.

Keywords: Fertilization strategies · Climate change · Sugar cane crops

1 Introduction

In agriculture, air and soil provide the nutrients required by plants during their germination and growth [1]. When the soil is balanced, i.e. when nutrient contents in soil are adequate, crops have a higher yield. However, when some of the nutrients do not exist, are not available or present in an excessive amount, an imbalance of nutrients occurs, and crop yield is decreased [1, 2].

For this reason, agriculture in the last decades has focused on improving soil preparation techniques, developing optimal irrigation techniques, and defining more efficient fertilization strategies, with support from *Information and Communication Technologies* (ICT) [3]. Fertilization strategies are efficient if their design considers both the physical, chemical and biological soil characteristics, and the plant's nutritional needs and its resistance to pests and diseases [4]. Fertilizers restore nutrients to the soil, making them available to crops. Soil loses nutrients either by erosion, absorption by plants, retrogression or water runoff [5]. It is essential to study and analyze *fertilization strategies* considering the following criteria: (A) optimal balance of nutrients required by

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_6

plants and (B) soil-specific characteristics, in terms of its nutritional and water requirements [1, 2, 5]. However, few research studies on fertilization strategies consider the above requirements—especially in sugar cane crops—[6].

Soil fertilization strategies have changed over time to include the environmental impact of fertilizers. This leads to productive and sustainable agriculture, in which water and fertilizers are used in a rational way, with maximum efficiency, but respecting and protecting the environment [7, 8].

The previous context shows the actors involved in agronomic management. We see the production system as a soil-plant-climate management system with a holistic approach, considering the soil as a fundamental element. Achieving the soil largest nutritional potential, requires that the fertilization strategy is articulated with the system, and with other elements, e.g. environmental impact and sustainability [8, 9]. Recent studies on crop fertilization practices around the world show that 30%–50% of crop yield is attributable to applied nutrients [10, 11]. Sugar cane is one of the agricultural products requiring the largest amounts of nutrients, especially nitrogen, phosphorus and potassium. Optimal nutrient levels are essential for achieving the biomass yield, e.g. sugar concentration per unit of mass [12].

Research studies from Brazil demonstrate that the upper portion (stem) of sugar cane contains the following nutrients per 100 tons of stems: **N** (nitrogen) with 100 to 154 kg, **P2O5** (phosphorus pentoxide) with 15 to 25 kg, **K2O** (potassium oxide) with 77 to 232 kg, and **S** (sulfur) providing 14 to 49 kg. Among Brazilian crops, sugar cane ranks as third in fertilizer consumption, after soybean and corn [7, 10]. In 2007 and 2008, approximately 8.3×10^6 ha of sugar cane were cultivated, requiring about 3.4×10^6 tons of NPK (nitrogen, phosphorus, and potassium), representing 13.8% of the total fertilizer used in crops in Brazil [10]. Most of the required nutrients have a specific function to improve plant performance. Nitrogen is important to stimulate plant growth and development. Phosphorus is particularly important for root growth, shoot early growth, increased early productivity, and good internode length. Potassium—like nitrogen—stimulates a strong development of the cane, provides a good internode length, greater circumference, and yield [13]. Elements like magnesium, sulfur, and iron increase photosynthetic activity, leading to good development and high yield. Calcium ensures plant resistance and promotes root, stem and leaf growth.

However, usage of large amounts of nitrogen-based fertilizers in agriculture has increased pollutants causing high $\mathbf{N_2O}$ (nitrous oxide) emissions into the atmosphere. This calls for rational and optimal use fertilizers, especially the ones containing nitrogen [14, 15]. Regarding water usage for sugar cane crops, the authors of a study conducted in Brazil concluded that when plants have greater water availability either by irrigation or rain, their efficiency in fertilizer use increases, leading to a reduction of 25% in the use of nitrogen-based fertilizers [16]. However, this behavior is not evident in all the developed studies and trials, leading to conclude that other factors related to soil, climate, and water, interact and affect effectiveness of used fertilizers [16].

Application of nitrogen-based fertilizers in agriculture is one of the main sources of $\mathbf{N_2O}$ emissions to the atmosphere. $\mathbf{N_2O}$ is one of the GHG (greenhouse gases) with the greatest effect in climate change (298 times worse than $\mathbf{CO_2}$) [18]. Consequently, research on fertilization strategies should be aimed towards integrating productivity with the environmental sustainability of agricultural practices.

Research suggests that lower doses of fertilization could: (A) reduce GHG emissions related to agricultural production, (B) promote climate change mitigation, and (C) contribute to improved crop productivity in terms of cost [19]. For this reason, methodologies such as carbon footprint measurement, which seeks to quantify the total amount of GHG emissions generated during life cycle of the crops, have been developed.

Undoubtedly, the correct definition and usage of a fertilization and irrigation strategy is a practice offering better expectations about the quality and productivity of the sugar cane crops. If not defined and used effectively, it will adversely affect production and income. Hence, we must define a methodology identify and delimit variables and parameters affecting soil and crops, as well as availability and accessibility of nutrients [19].

The need to use fertilization methods and strategies to mitigate effects of GHG, fluctuating sugar prices, and constant increase of production costs lead the sugar industry to define and develop fertilization and irrigation techniques, supported by ICT. This can help to improve cultivation processes, increase sustainable production of sugar cane, strengthen producer's resilience, and reduce GHG emissions [3, 14]. One technological advance achieved so far in terms of fertilization strategies is *site specific crop management* (SSCM), consisting in adequate irrigation and fertilization for a given site, as required. Here, ICT plays a fundamental role as a tool for planning and managing land usage, thus becoming an important enabler for productive and sustainable fertilization strategies [20].

Statistical models, sensor networks, global positioning systems, and geographic information systems can be used for specific characterization of each sector, and pattern recognition in soil spatial and temporal data [19]. The obtained information allows to define an adjusted fertilization strategy, aiming to increase productivity. In spite of advances aimed at improving crop productivity, inadequate fertilization strategies remain as an unsolved problem in agriculture.

2 Research Methodology

In order to obtain a complete and adequate survey of research work related to fertilization strategies for sugar cane crops aided by ICT, we approached the analysis using the systematic literature review (SLR) methodology [21]. This methodology bases analysis on available evidence, featuring a review of quantitative and qualitative aspects of articles and/or primary studies, aiming to summarize existing information about a given topic. SLR consists of five steps: define research questions, perform a literature review, select studies, classify articles, and extract/aggregate data.

We expect to find deficiencies in fertilization processes carried out by sugar cane producers, thus opening the way to new tools allowing better use of resources, increased production rates, and reduced GHG emissions. Carbon footprint impact increases every day, impacting environment, human health, and economy. These impacts are expected to become more severe, unless concerted efforts to reduce emissions are undertaken.

Improved fertilization techniques will eventually lead to reduce the carbon footprint. Our goal is to assess effectiveness, standardization, and environmental impact of agricultural interventions in sugar cane crops on the geographic valley of Cauca River (GVCR).

Due to the large quantity of scientific information regarding fertilization, precision agriculture, sugar cane crops, carbon footprint of sugar cane crops, and soil analysis, we decided to sample such information. This selection of articles yields a sample allowing to recognize, evaluate, and interpret important information for development of our project. The sample consists of articles or scientific journals meeting chosen by using a non-probabilistic method known as *purposive sampling*, which avoids random selection, but uses selection criteria satisfying the aforementioned goal. Article search was performed manually in the following databases:

ScienceDirect: Database containing scientific, technical and medical research articles. (http://www.sciencedirect.com/)

SIDALC: The SIDALC alliance brings access to information produced in agricultural institutions and related information centers in the Americas. (http://www.sidalc.net/cgi-bin/wxis.exe/)

Colombian Society of Soil Science: This scientific society's library features publications, magazines and manuals on soil science. (http://www.sccsuelo.org/index.php/afiliaciones)

Cenicaña: A private nonprofit corporation founded in 1977 with support of the Association of Colombian Sugarcane Growers, *Asocaña*, and funded with donations from sugar mills and sugar cane growers/suppliers located in the Cauca River Valley. (http://www.cenicana.org/web/publicaciones).

Each of these organizations has a track record and recognition, either global or in Latin America. This assures us they are trusted sources with quality information. Table 1 presents conventions used for keywords used in the inclusion criteria, and Fig. 1 presents a graph of number of articles found per inclusion criteria. Search was limited to articles published from January 2000 to April 2017, matching topics A, B and E as described in Table 1. Full-text articles in English, Spanish or Portuguese were considered.

Table 1. Conventions used for the keywords of inclusion criteria.

Convention	Topic (in Spanish)
A	Caña de Azúcar
B	Región Tropical
C	Fertilización Tasa Variable
D	Agricultura de Precisión
E	Medio Ambiente/Huella de Carbono
F	Agricultura Específica por sitio

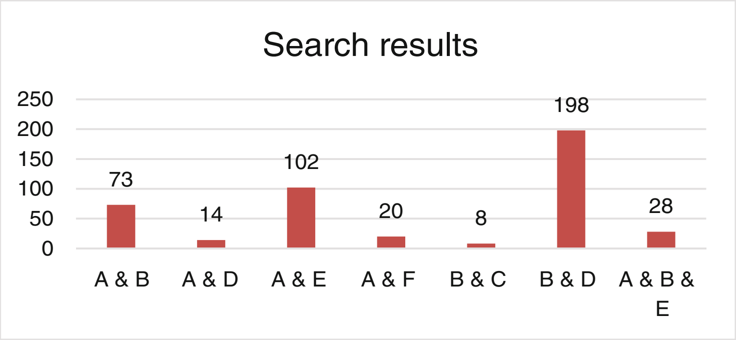


Fig. 1. Number of papers found when submitting the topic keywords described in Table 1.

2.1 Research Questions

To contribute to the state of art of fertilization strategies for sugar cane crops aided by ICT, our review defined the following research questions:

(A) Which are the traditional methodologies applied to the fertilization problem?
(B) What are the effects of such methodologies in terms of yield, in current practice of sugar cane cultivation?
(C) What are the recent trends in methodologies for solving the problem of site specific fertilization management, aided by ICT?

Which are the traditional methodologies applied to the fertilization problem? Nutritional management is one of the fundamental pillars allowing optimization of yield and productivity of sugar cane processing systems. For the farmer, this means TIC-aided fertilization methodologies and strategies should be integrated into the production processes. If this inclusion is to have a positive impact for the producer in the technical/ economical aspects, a production planning/scheduling process must exist, and such process must include a fertilization plan [22, 23]. Sugar agroindustry in Colombia has been implementing fertilization strategies aimed to achieve greater yield per area unit at a lower cost, allowing farmers to obtain larger profits. However, these traditional strategies are not supported in formal processes aided by ICT [3]. Moreover, they do not consider adverse environmental effects of nitrogen-based fertilizers [23]. Traditional methodologies for fertilization feature the following stages: (A) Soil sampling and analysis, (B) Fertilization diagnosis, (C) Fertilization plan design, (D) Fertilization plan execution and monitoring, and (E) Fertilization plan evaluation and analysis [24].

One of the crucial activities in the methodology is sampling and analysis of soil, due to specific nutritional requirements of each sugar cane variety [7, 25]. Fertilization and irrigation strategies for sugar cane have not been adequately addressed; hence it is common to find crops where the same fertilization composition and dosage is applied to different types of soils. Low availability or absence of micro-nutrients hinders growth and yield of the crops.

Studies show traditional fertilization methodologies are mostly based on empirical knowledge of the soil behavior by the farmer, and upon fertilization practices that do not consider soil variability and conditions. This means the farmer does not have historical records allowing to determine yield of a specific area, taking previously used fertilization strategies into account [26]. We conclude deficiencies in yield arise from traditional methodologies not considering the specific characteristics of the soil-plant-climate system, especially the spatial-temporal variability of soil. This is mostly caused fertilization strategies not being supported by standardized processes aided by ICT [22].

Which are the effects of such methodologies in terms of yield, in current practice of sugar cane cultivation?
At the production level, deficiencies of N, P, and S are not isolated from each other, but are combined in several ways; hence it is necessary to fully evaluate the effect of fertilization strategies on soil fertility [16, 26]. According to [6], fertilizer usage in Colombia shows a marked increase since 1950, with consumption rising from 20,000 tons/year to 742,200 tons/year in 1980, reaching 1.56 million tons in 2006. However, these increases in fertilizer consumption are not directly proportional to crop yield, because annual production of sugar cane had a growth period only between 1960 and 1980 period, when it raised from 88 tons/ha to 126.4 tons/ha. Between 1982 and 1992, annual production became stagnant at 120 tons/ha, and the crop growing time raised from 9 months in the 1960–1970 period, to nearly 13 months in recent years.

According to [27], excessive application of N acidifies the soil, stimulating intense microbial activity, that in turn decreases the availability of organic matter (OM). Eventually, carbon drops down to levels affecting bacteria populations, causing less mineralization and loss of N by lixiviation. As stated by [23], after 6 years of continuous application of stillage for fertilization, it is possible to observe contamination of underground aquifers with the stillage, showing lixiviation of fertilizer applications. This effect is more acute in sandy soils, since they have less retention capacity and they favor fertilizer infiltration.

Low organic carbon conditions due to fertilization combined with intensive farming can reduce the OM percentage below 1%, which is biologically negative for crop development. When soil is plowed, its vulnerability to erosion either by rainfall, wind, or irrigation is increased, generating cumulative damages that show in the long term, with a lower yield of sugarcane, as observed in the GVCR [23]. According to [25], farming methods and other soil management approaches, together with acceleration of organic carbon mineralization in soil, also influence $\mathbf{CO_2}$ emissions—the GHG of anthropogenic origin having the highest growth between 1970 and 2004 (80%) [18]. As stated in [25], conventional farming in sugar cane crops of Brazil releases 350.09 gm^{-2} of $\mathbf{CO_2}$ to the atmosphere, suggesting that 80% of the carbon that could potentially be stored in the first 20 cm of soil is lost [18]. Climate change is caused by greenhouse effect, which is caused by accumulation of GHG in the atmosphere. GHG absorb ground-emitted radiation and trap heat into the atmosphere, preventing energy from being radiated to space. This raises the temperature in the planet, triggering modifications in ecosystems, biodiversity loss, decrease in water availability, loss of productivity for certain crops and livestock activities, floods, storm surges, typhoons, erosion, and other phenomena [23] that also affect sugar cane production.

The term *carbon footprint* was coined as an indicator of the amount of greenhouse gases generated and emitted by operation of a company, or during the life cycle of a product, including not also its production, but also its consumption, end-of-cycle recovery, and disposal. Measurement and control of this indicator promotes improvement of energy efficiency, eco-efficiency, clean production, life-cycle analysis, and design of environmentally friendly products and services, as well as search for alternative energies [23]. In terms of sugar cane production, controlling GHG emission in crops through exact use of the fertilizer the soil can assimilate is positive from the economic and environmental points of view, because fertilization costs, soil damage and emissions are reduced, while increasing crop yield. However, because of high soil heterogeneity, determining the exact amount of fertilizer is still a task under development. According to [23], $\mathbf{N_2O}$ and $\mathbf{CO_2}$ emissions from sugar cane crops increase significantly when organic and mineral fertilizers are applied in combination. Also, during new cane growth after sowing, $\mathbf{N_2O}$, $\mathbf{CO_2}$, and $\mathbf{CH_4}$ emissions increase due to the high amount of crop residues on the soil surface.

Authors in [17] suggest that sugar cane production contributes more than 59% of the GHG emissions coming from sugar production. 19% to 50% of such emissions are due to use of fertilizers in the growing process. Although the belief is that sugar cane productivity is linked to the amount of fertilizer applied to soil [6, 7, 24] for the case of sugar production in Mexico, [27] evidences that fertilizer doses used in Brazil are lower than the ones employed in Mexico, yet sugar cane crop yield in Brazil (in terms of kg of sugar per kg of processed cane) is much higher than the one reported in Mexico. This case is similar to the information reported in [20] and [6] for sugar cane production in the GVCR.

What are the recent trends in methodologies for solving the problem of site specific fertilization management aided by ICT?

Analysis of our systematic literature review evidences that the soil-plant-climate management system can be modeled as a holistic system, where different components complement each other to offer the best results in terms of yield. This section examines how ICT can be used to support a sustainable agriculture system, which is defined by actions seeking to increase sustainable productivity, strengthen producers' resilience and reduce GHG emissions [3, 28]. Three ICT tools are key for implementing sustainable fertilization strategies: (A) Global Positioning Systems (GPS), (B) Geographic Information Systems (GIS), and (C) Remote Sensing (RS) techniques. GPS and GIS allow to stack multiple layers of information derived from various sources in a single environment. This is useful for building consensus on land planning and usage, especially when farmers have different perspectives and preferences over a certain territory [13, 28]. GIS lets the farmer get a more detailed vision of the soil and lets him store data about its characteristics. RS techniques are a valuable tool for monitoring different agronomic and weather variables characterizing the soil-plant-climate system. These sensors provide a huge amount of information about the variables of interest; however, it is necessary to obtain value from the information provided by such technologies. This is done by analyzing and filtering in the data, in order to obtain better information and facilitate decision making [29, 30]. Field sensors are a complement to the remote sensing techniques; their main task is to monitor soil characteristics and properties, especially dynamics related to vegetative cycles of the crops [30]. Continuous technological advances are fueling the big data revolution, and redefining the data concept [28]. We consider two types of innovations: remote data capture and cellular phones as potential data sources can affect farmers' decision-making capacities, when combined with other information sources. Currently in Colombia, usage of GPS and GIS in agriculture is growing. Farmers use them for fertilizer application prior to planting, and for sowing and crop yield measurement [3]. But measurements could also be performed during plant growth, in order to detect the need for applying agricultural supplies, to assess phenomena and learn from them, or to forecast yield upon sowing [31]. Then, it is clear that ICT is increasingly embedded in the business, and that cost/benefit ratios will improve as the technological developments continue [23].

Current economic and environmental conditions promote development of new management techniques for agricultural systems [26]. Rapid development of ICT has allowed to design alternative management systems for the crop fertilization problem, e.g. precision agriculture or site-specific agriculture. These management systems use GPS, GIS, simulation models, productivity heat maps, decision support systems (DSS), and artificial intelligence (AI) techniques, aimed to increase yield and the efficiency of the agricultural supplies, by accurately determining the spatial-temporal variability of conditions of the whole soil-crop-climate ecosystem [32]. In order to achieve high and sustained yield over time, it is essential to integrate fertilization management with other aspects of crop management (sanitary management, genotype selection, hydric balance, water cycle, pest control, etc.) and to apply good agronomic management practices (GAMP). The main GAMP are: direct planting, crop rotation with gramineous plants, and balanced fertilization. Application of these good practices allows entering into *ideal sustainable agriculture*, with higher, more stable yield, and minimum soil deterioration. There is clear evidence that well-rotated and fertilized soils improve their physical, chemical, and biological fertility, thus benefiting sustainability of productive systems. Recently there has been an increase in AI techniques bound to agriculture. The most common ones are neural networks, genetic algorithms, expert systems, and fuzzy logic. [33] discusses a model to determine management zones for fertilization, using an artificial neural network with multilayer perception capabilities and a backpropagation learning algorithm. This model was fed with data from a sugar cane field (234.5 ha) of the Normandía farm in Bugalagrande (at the north of the VGRC), in a soil developed in the basin physiography, characterized by its alkaline reaction, high saturation of exchangeable magnesium, and high in clay. The model was based in the spatial interaction among some soil properties involved in fertilization and crop productivity. The map of homogeneous areas generated by the model showed a coefficient of spatial autocorrelation of 0.94 with a significance level $P \leq 0.01$, which is adequate to delimit zones for fertilization management. The model results are useful to identify areas of high, medium, and low yield, which can be handled with an economical and/or an environmental approach. Also, authors identified that high levels of calcium (Ca2+) form insoluble and highly soluble magnesium (Mg2+) and sodium (Na+) cations, impairing the physical, chemical, and biological soil, affecting their productivity [33]. The model generated two homogeneous management zones: low and high productivity. The difference between these zones in terms of productivity can be explained by the pH level and the properties of the exchangeable cations Ca2+ , Mg2+ and Na+ [33]. Other example of application of spatial technologies for site-specific management in crop production is analyzed in [34], with the use of GPS and GIS systems. The contribution of an explicit spatial econometric methodology in the estimation of crop yield is assessed, in order to determine functions that are used to optimize fertilizer application. The specific case study is for nitrogen application to corn production in Argentina, where the implementation of a variable rate technology (VRT) requires methods that use inexpensive information, and are focused on the inputs and variability common to Argentine growing areas. The objective is to assess the economic value of application of spatial regression analysis to yield monitor data, as a means to optimize variable rate fertilizer strategies. Data in the case study are from on farm trials with a uniform N rate along strips, and a randomized complete block design to estimate site-specific crop response functions. Spatial autocorrelation and spatial heterogeneity are taken into account in regression estimation of the N response functions by landscape position, in the form of both a spatial autoregressive error structure and group-wise heteroskedasticity. Both uniform rate and VRT returns are computed from a partial budget model, and the results suggest that the N response differs significantly by landscape position, and that the VRA for N may be modestly profitable depending on the VRT fee level. Profitability depends crucially on the model specification used, with all the spatial models consistently suggesting profitability, whereas the non-spatial models do not [34]. The key benefit of an explicit spatial econometric methodology, such as the one employed in [34], is that any spatial structure in the data is exploited to yield more precise (and, in some cases, less biased) estimates for parameters in a yield response function. Since these parameters form the basis for all ensuing economic computations, an increased precision affects precision of estimates for yield, return and profitability. [35] presents a very interesting analysis of how to apply ICT in precision agriculture in India. The study allows to conclude that its results can be applied to developing countries, since precision agriculture (PA) uses all modern technologies such as GPS, RS, and VRT. Three components namely ‘single PA technology’ (for the user to select one or a combination) and ‘integrated PA technology’ have been identified as a part of adoption strategies of PA in developing countries [35]. A detailed strategy for adoption of PA in India and developing countries has also been proposed in [35, 36]. Soil analysis are the most efficient tool to know the availability of soil nutrients or variable soil properties in time and space. The purpose of soil analysis is to assess the adequacy, surplus or deficiency of available nutrients for crop growth, and to monitor change brought about by farming practice. When data of this stage of the process becomes available, most of the AI procedures come into play, because it is a vital input to make correct decisions about fertilization.

3 Conclusions

Deficiencies in crop yield are due to the fact that traditional methodologies do not consider specific characteristics generated by the variability of the soil-plant-climate system, especially the spatial and temporal variability of the soil. This happens because fertilization strategies are not supported in standardized processes, and do not rely on ICT tools or technologies.

Whilst soil analysis is not a perfect tool, it is the most effective and practical means of assessing its fertility. Analytical data should be used in conjunction with other knowledge such as soil type, structure, decision support systems, environmental impact, and crop yield as the basis for deciding on fertilizer and manure use. Also, ICT tools will help to store and gather data to make accurate decisions about fertilization. Crop farmers in Colombia need to embrace ICT tools to improve the way fertilization is currently being done. In time, they will see a better cost/benefit ratio and will be possible to track results of decisions made upon their crops.

It is important that farmers start taking action about the greenhouse effect that farming causes, because $\mathbf{N_2O}$ produced during fertilization will increase the production of GHG. Greenhouse effect is a natural and inevitable process where the atmosphere traps some of the sun's energy, warming Earth just enough to support life. But in a non-controlled fashion, it will cause changes in the weather that will affect the types of crops grown in different parts of the world. Changes in the amount of rainfall will also affect how many plants are grown. Effect of weather changes on plant growth may lead to some countries not having enough food.

References

1. Asociación internacional de la industria de los fertilizantes: Los fertilizantes y su uso, pp. 9–11 (2012)
2. Alvarado, A., Hernandez, F., Duran, N., Gutiérrez, M., Herrera, W., Molina, E., Sancho, F., Soto, G., Flores, C., Salas, R., Rodríguez, J.: Fertilidad en Suelos y Manejo de la Nutrición de Cultivos en Costa Rica. In: Méndez, G., Molina, E. (eds.) Memorias Laboratorio de Suelos y Foliares 2011, San José, Costa Rica, pp. 109–115 (2011)
3. Palmer, N.: Las TIC y la agricultura en el contexto del "crecimiento verde". Las TIC en la agricultura **2**, 1–9 (2012)
4. García, P.: El suelo, los nutrientes, los fertilizantes y la fertilización. Guía Práctica de la fertilización racional de los cultivos en España. Parte 1(2), 15–17 (2009)
5. Hodges, S.: Soil Fertility Basics. North Carolina State University, North Caroline
6. Delgadillo, V., Garcia, R., Forero, J.: Fertilising techniques and nutrient balances in the agriculture industrialization transition: the case of sugarcane in the Cauca river valley (Colombia), 1943–2010. Agr. Ecosyst. Environ. **218**, 150–162 (2016). Elsevier, 1–2
7. Hazelton, P., Murphy, B.: Interpreting soil test results, what do all the numbers mean?, 2nd edn. CSIRO Publishing, Melbourne (2007)
8. Frohmann, A., Herreros, S., Mulder, N., Olmos, X.: Sosteniblidad ambiental y competividad internacional. La huella de carbono en las exportaciones de alimentos. Naciones Unidad. Cooperación regional francesa para América del Sur (2015)
9. Abdel Wahab, D.M.: Introduction of fertigation in sugarcane production for optimization of water and fertilizers use. Agric. Sci. **5**, 945–957 (2016). 1–2
10. Cantarella, H., Rossetto, R.: Fertilizers for Sugarcane. Sugarcane bioethanol—R&D for Productivity and Sustainability, pp. 405–422 (2014)
11. McCray, J., Rice, R., Ezenwa, I., Lang, T., Baucum, L.: Sugarcane Plant Nutrient Diagnosis. In: Florida Sugarcane Handbook, pp. 1–10 (2016)

12. COLPOS: Manejo sustentable de la fertilidad del suelo y de la nutrición de la Caña de Azúcar (2016)
13. Plassmann, K., Norton, A., Attarzadeh, N., Jensen, M.P., Brenton, P., Edwards-Jones, G.: Methodological complexities of product carbon footprinting: a sensitivity analysis of key variables in a developing country context. Agr. Ecosyst. Environ. **13**(5), 393–404 (2010). Elsevier, 1–2
14. Nele, I., Grahmann, K., Cox, R., Grovaerts, B.: Eficiencia del uso de nitrógeno y optimización de la fertilización nitrogenada en la agricultura de conservación. Centro Internacional de Mejoramiento de Maíz y Trigo, pp. 1–12 (2015)
15. Di, H., Cameron, K.: Calculating nitrogen leaching losses and critical nitrogen application rates in dairy pasture systems using a semi-empirical model. N. Z. J. Agric. Res. **43**, 139–147 (2000). 1–10 (2017)
16. Cabrera, J., Zuaznábar, R.: Respuesta de la caña de azúcar a la fertilización nitrogenada en un experimento de larga duración con 24 cosechas acumuladas. Cultivos Tropicales **1**(31), 93–100 (2010)
17. Stewart, W., Dibb, D., Johnston, A., Smyth, T.: The contribution of commercial fertilizer nutrients to food production, vol. 97. American Society of Agronomy (2005)
18. Wesseh, K., Lin, B.: Climate change and agriculture under CO2 fertilization effects and farm level adaptation: where do the models meet? Collaborative Innovation Center for Energy Economics and Energy Policy (2017)
19. Molina, C., Molina, E., Molina, J.: Manejo agroecológico de caña de azúcar y sistemas silvopastoriles intensivos. Revista Tecnicaña **29**, 29–35 (2012)
20. McCray, M., Rice, R., Ezenwa, I.: Nutrición en caña de azúcar, Uso del análisis foliar. Revista Fertilizar **21**, 5–8 (2011)
21. Kitchenham, B., Brereton, P., Budgen, D., Turner, M., Bailey, J., Linkman, S.: Systematic literature reviews in software engineering – a systematic literature review. Information and Software Technology (2008)
22. Fernandez, O.: Variabilidad espacial y uso del manejo sitio específico de la fertilidad del suelo, en el cultivo de caña de azúcar. Universidad de los Andes – Venezuela (2017)
23. Luna, C.: Aumento de la Productividad de Caña de Azúcar por Unidad de Área Cultivada - El biocarbón una alternativa ecológica y rentable. Revista Tecnicaña, 1–16 (2005)
24. Barea, J., Azc, R.: Boletín técnico informativo octubre 2015 Nutrición del cultivo de caña de azúcar y uso eficiente de fertilizantes. Revista Condesuca - SAGRAPA, pp. 1–11 (2015)
25. Isaacs, T., Cassalett, C., Torres, J.: El cultivo de la Caña en la zona azucarera de Colombia, 1st edn. Centro de investigación de la Caña de azúcar de Colombia, Colombia (1995)
26. Ferraris, G., Toribio, M., Falconi, R., Moriones, F.: Efectos de diferentes estrategias de fertilización sobre los rendimientos, el balance de nutrientes y su disponibilidad en los suelos en el largo plazo. Informaciones Agronómicas de Hispanoamérica - IAH 19, pp. 11–17 (2015)
27. Silva-Olaya, A., Cerril, C., La Scala, N., Dias, C., Cerri, C.: Carbon dioxide emissions under different soil tillage systems in mechanically harvested sugarcane. Environmental Research Letters **8**, 015014 (2013). IOP Publishing
28. Rafea, A.: Managing Agriculture Knowledge: Role of Information and Communication Technology. American University in Cairo (2012)
29. Urbano-Molano, F.: Wireless Sensor Networks Applied to Optimization in Precision Agriculture for Coffee Crops in Colombia. Journal de Ciencia e Ingeniería **1**(5), 46–52 (2013)
30. Barreiro, P.: Sensores para la caracterización del suelo agrícola usados en agricultura de precisión. Revistsa Eumedia (2014)

31. Kostka, G., Polzin, C., Scharrer, J.: The future of sugar cane in (the) people's Republic of China and India – supply constraints and expansion potential. Appl. Energy **86**, 100–107 (2009). Elsiver
32. Martin, P., Hernandez, O., Scharrer, J.: Autonomous vehicle for dosage fertilizers on small and medium crops. In: ABCM Symposium Series in Mechatronics, vol. 6, pp. 650–660 (2014)
33. Erazo, E., Mosquera, C., Borrero, J.: Modelo de fertilización de suelos por sitio específico utilizando redes neuronales artificiales. In: Memorias del congreso Atalac – Tecnicaña, vol. 6, 773–784 (2012)
34. Anselin, L., Bongiovanni, R., Deboer, J.: A spatial econometric approach to economics of site specific nitrogen management in corn production. Am. J. Agr. Econ. **86**(3), 675–687 (2004)
35. Pinaki, M., Manisha, B.: Adoption of precision agriculture technologies in India and in some developing countries: scope, present status and strategies. Prog. Nat. Sci. **19**(6), 659–666 (2009)
36. Palaniswami, C., Gopalasundaram, P., Bhaskaran, A.: Application of GPS and GIS in sugarcane agriculture. Sugar Tech. **13**(4), 360–365 (2011). https://doi.org/10.1007/s12355-011-0098-9

Coffee Rust Detection Based on a Graph Similarity Approach

Gersain Lozada(✉), Geraldin Valencia, Emmanuel Lasso, and Juan Carlos Corrales

Grupo de Ingeniería Telemática, Universidad del Cauca, Campus Tulcán,
Popayán, Cauca, Colombia
{lominoli, geral, eglasso, jcorral}@unicauca.edu.co

Abstract. Diseases affecting agricultural sectors are often closely related to weather conditions and crop management. In this regard, different researches have focused on identifying patterns that lead to the incidence of these diseases. This research was carried out in order to detect favorable conditions for rust in coffee trees (*hemileia vastatrix*) based on a graph representation of the Agro-climatic information of the crops. Furthermore, we adapted 4 error-correcting graph pattern matching algorithms, classified taking into account the precision and the execution time, in order to find a similarity percentage between current conditions of a coffee crop and the graph patterns that describe coffee rust infection rates.

Keywords: Graph pattern · Coffee rust · Error-correcting graph matching · Edit distance · Disease · Agriculture

1 Introduction

Agroindustry in Colombia is a traditional production sector that needs the implementation of contingency measures in a timely manner for events that create a risk for crops. For coffee production, coffee rust is a disease that affects quality and production costs for farmers greatly. As a result, government organizations, academic and specialized research centers have focused on proposing techniques of chemical control (pesticides) and changes in agronomic properties of crop, which must be applied in an appropriate time in order for the disease to be treated before generating irreversible effects on coffee trees [1]. However, the constant use of these elements (pesticides) generate high health risks for farmers and also produce a deterioration in the ozone layer which represents a great environmental impact [2].

Consequently, some researches in the sector [1, 3–6] have focused their efforts on determining the relationships between weather conditions and agronomic properties of crops with episodes of epidemics of diseases as coffee rust to determine which variables are directly related to the disease [7] and make use of several techniques to generate predictive models from a knowledge base. In [7, 8], the climatic effects on the development of coffee rust are studied using various techniques of data mining to find coffee rust incidence patterns in an agroclimatic dataset. These approaches allow an approximation in the characterization and prediction of the disease. In spite of the

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_7

numerous approaches in this area, there have not been reported investigations that carry out an exploration and application of an error-correcting technique in predictive models of coffee rust yet.

The predictive models can be represented as graph patterns which contains all the variables related to the development of coffee rust and their relationships [9]. A graph-based information structure consists of a set of nodes (for the specific case of coffee rust, the nodes may represent variables of monitoring data) related through edges (they can represent the semantic context between variables measured in crops [10]), which have a dynamic nature with capabilities for handling large amounts of data. Thus, the graphs can be used for storage and analysis of different variables present in a crop environment. One of the graph analysis techniques is the graph pattern matching, used to find a correspondence between the nodes and the edges of two graphs that satisfies some (more or less stringent) constraints ensuring that similar substructures in one graph are mapped to similar substructures in the other [11].

For exact graph matching the mapping between the nodes have to be bijective, consequently a correspondence must be found between each node of the first graph and each node of the second one. On the other hand, the error-correcting graph matching does not need found a bijective relation between nodes, instead it uses the graph edit distance [12–14] which is one of the most flexible methods for this type of graph matching which defines the dissimilarity of the graphs as the amount of distortion (set of editing operations: Insertion, deletion or substitution of nodes or edges) which looks for the smallest number of operations needed to transform one graph into another [14, 15]. We use error-correcting graph matching, as it allows a closer approach to the problem domain due to the intrinsic flexibility of the technique and that considers a degree of proximity in the task of searching for graph patterns.

We present an adaptation made for 4 algorithms (A*, Beam, Hungarian, Volgenant-Jonker (VJ)), chosen based on the precision, runtime and availability of the source code. This adaptation allows to find the similarity between the graphs patterns and the current conditions of the crop represented as a graph giving as a result a probability of growth or decline of the coffee rust infection rate.

The outline of this paper is organized as follows: Sect. 2 describes the graph-based representation used; Sect. 3 refers to the selection and adaptation of graph matching algorithms, Sect. 4 exposes the obtained results and Sect. 5 presents the conclusions and future work.

2 Graph Based Representation

We used the graph patterns obtained in [16] and the structure of graphs shown in Fig. 1. It is necessary to take into account that the patterns represent three coffee rust infection rates which were calculated by means of the evaluation of the increase or decrease of the incidence percentage between the analyzed month and the following month. The three infection rates obtained from [16] were: TI1: reduction or latency, to negative or none rates of infection; TI2: moderate growth, to positive infection rates lesser or equal to 2% points (pp); and TI3: accelerated growth, infection rates higher than 2 pp. Thus, there are 7 graph patterns: 2 for TI1 (called P3 and P7), 3 for TI2 (called P1, P2 and P5) and 2 for TI3 (called P4 and P6).

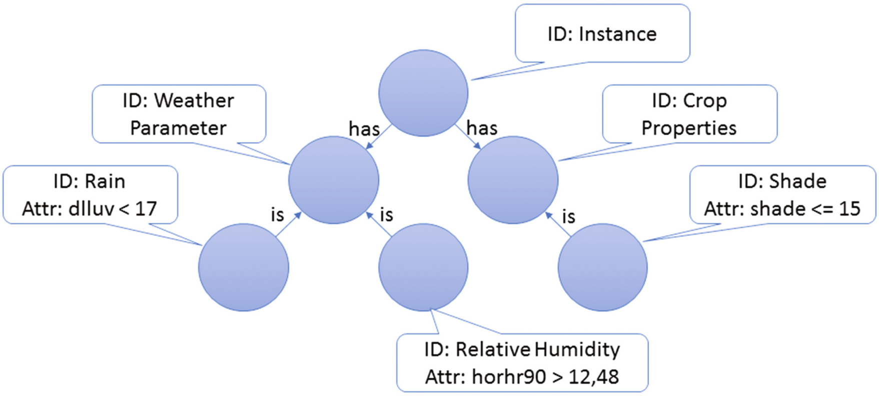

ID: Instance
ID: Weather Parameter
ID: Crop Properties
has
has
ID: Rain Attr: dlluv < 17
is
is
is
ID: Shade Attr: shade <= 15
ID: Relative Humidity Attr: horhr90 > 12,48

Fig. 1. Structure of graph patterns and infection rates

The graph pattern structure in Fig. 1 allows to identify the position of node entities, the crop information represented by them and their relationships. Further, each entity contains labels with the predictive variables shown in Table 1 (described in [16]) that allow to perform an estimation of coffee rust growth by means of the infection rates associated to each graph pattern displayed in Table 2. These crop variables and properties are related with a principal node that represents an Instance from this a graph instance that describes the weather and agronomic conditions in the crop for a specific month is created. In this research, we worked with a set of 98 graph instances which were extracted by Lasso et al. [16], using the data obtained from the monthly monitoring of coffee crops at Los Naranjos Farm belonging to the Supracafe Company in Cajibio, Colombia from 2011 to 2015 by Corrales et al. [17].

Table 1. Coffee rust predictive variables

Crop information	Predictive variable	Node entity type
Weather parameter	TMAX TMED TMIN	Temperature
	HORHR90 HORHRN90	Relative humidity
	PRE_ACUM	Rain
Crop property	SHADE	Shade

The principal components of the graph structure used and their meaning in coffee crops environment are:

Table 2. Coffee rust graph patterns

Infection rate	Graph patterns
TI1	P3, P7
TI2	P1, P2, P5
TI3	P4, P6

- **Nodes:** Represent Instances (entity related with crop monitoring monthly), Crop properties (as shade and density) and Weather Parameters (entities that represents weather monitoring variables as Rain, Temperature, Relative Humidity and Wind)
- **Edges:** Represent the relationships between nodes and their dependency.
- **Node Labels:** Represent the value of predictive variables related with coffee rust development.
- **Edges Labels:** Represent the context semantic of the nodes relationships.

The format used for graph representation is GXL (Graph Exchange Language) [18]. This one is based in XML, which allow graph storage, adaptability and flexibility, supports multiple edges and nodes and multiple labels per edge and node. Furthermore, in this research we used a framework called Graph Matching Toolkit (GMT) developed by Riesen [19]. GMT allows to find an optimal solution to the graph matching problem and also provides some algorithms for non-optimal but computationally less expensive solutions. In order to compute the graph edit distance, several modifications were made on this software tool, specifically, nodes and edges labels edit distance functions were changed by a function adapted to our application domain, besides a function to compute the edges inversion cost was added.

3 Graph Based Representation

The algorithms for similarity calculation were selected based on the recurrence in which they are presented in the different investigations where comparisons between them were made. Given the characteristics of the application domain it is necessary to have a high reliability in the information obtained from the algorithms; the considered selection criteria were: precision as the most relevant, the approach of labeled inexact graph matching and finally the source code availability that allows to reduce the uncertainty in the results and to increase the reliability.

3.1 A* Algorithm

It is a widely used method for calculating the exact value of the edit distance. It is a classic tree search algorithm which is based on exploring all the space of all possible mapping of nodes and edges of both attributed graphs [20]. A* algorithm [21–23] performs an exploration of the possible maps within two graphs, organizing the underlying search space as an ordered tree by mean of the creation of successor nodes iteratively linked by edges to the node currently considered in the search tree. Formally, for a node p in the search tree a function $g(p)$ to denote the cost of the optimal path from the root node to the current node p is used. Furthermore, a heuristic function $h(p)$ in order to determine the most promising node in the current search tree and denote the estimated cost from p node to a leaf node is used. The sum $g(p) + h(p)$ gives the total cost assigned to an open node in the search tree. Finally, the total minimum cost is obtained by $pmin \leftarrow argmin\{g(p) + h(p)\}$ that is the minimum path between the root node and the current node. The distance between two graphs is measured by applying a sequence of operations (node and edge insertion, deletion or substitution).

3.2 Beam Algorithm

This algorithm is based on the operating principle of A*, however, it provides the possibility of establishing the size of the search space and depending on it varies proportionally its accuracy.

The method follows a similar idea to the previously described technique, but instead of expanding all successor nodes in the tree, it limits the number of nodes to be processed that are maintained in the set of open nodes at all times. Whenever a new partial edit path is added, only the partial edit path p with the lowest cost $g(p) + h(p)$ is maintained and the remaining partial edit paths are deleted. This means that the entire search space is not scanned, only the nodes belonging to the most promising partials matching are expanded [24, 25].

3.3 Hungarian Algorithm

The famous Hungarian algorithm was proposed by Kuhn and it is known to solve the problem of assignment in time *O(n3)* where n is the size of the bipartite graph. In this research, an adaptation of the original algorithm is used, it allows to calculate the distance of edition of two graphs, which can be applied to any type of graph [13]. This method is based on a bipartite optimization procedure of nodes and edges mapping, where the nodes or edges of one graph are mapped in another [26].

One of the main problems in graph matching is that the standard algorithms for calculating graph similarity are exponential at run time depending on the number of nodes involved, for example, tree search algorithm. Therefore, these algorithms are applicable to small graphs only. For this reason, the Hungarian algorithm is used to calculate the editing distance, which calculates a complete matching of graphs, such that the sum of the weights of the edges in the matching is minimized [27]. In addition its use for the matching of bipartite graphs is desired because it provides a much faster process where precision is slightly affected [26].

3.4 Volgenant-Jonker Algorithm

This algorithm has been object of attention in the current literature due to its efficiency, it consists of three steps: a preprocessing method used to find the first partial solution, a dispersion stage to solve an instance with a reduced number of edges followed by a procedure that iteratively adds edges until obtaining an optimal solution and finally, a procedure to find the shortest path [28].

The VJ algorithm is used to solve the assignment problem, which leads to a faster computation of the graph editing distance running in a polynomial time. In addition, it provides a calculation of the suboptimal editing distance without considerably affecting accuracy; the reason for its suboptimality is that the edge information is taken into account only in a limited way during the process of finding the optimal allocation of the node between two graphs [28].

3.5 Proposed Adaptation to the Selected Algorithm

The proposed adaptation was made for the four selected algorithms (A*, Beam, Hungarian, VJ).

Adaptation of the Cost Functions. We considered the definition of directed graphs obtained in [29], where $G_p = \left(V_p, E_p, f_{vp}, f_{ep}\right)$ is defined as a graph; V_p is the set of nodes and E_p is the set of edges, so that the edges are directed and connect two nodes v_i and v_j if $(v_i,\ v_j) \in E_p$. In addition,v_i nodes with multiple directed edges $e_k = \left(v_i, v_j\right)$ can be found, where $K = 1...n$. Furthermore,$f_{vp}(v_i)$ is defined as the set of labels of a node v_i and $f_{ep}(e_i)$ is the label of the edge e_i.

We assume two graphs, where $Gp = \left(V_p, E_p, f_{vp}, f_{ep}\right)$ is called the pattern graph and $G_s = (V_s, E_s, f_{vs}, f_{es})$ is denominated graph instance, where the pattern contains ranges of climate and agronomic properties and graph instance maintains the parameters of the crop to be analyzed.

The Algorithm 1 shows the nodes label cost function that is based on ranges, it considers whether the evaluated value is within the range and also take into account the approximation by right or left to the range.

Nodes Label Cost Function

```
Node label cost function procedure $(V_p, f_{vp}, V_s, f_{vs})$
Start $cost, fit, nofit = 0$
if there is at least one $v_i \epsilon V_p$ and a $v_j \epsilon V_s$
    while to map all nodes $V_p$ in all nodes $Vs$
        if $v_i$ and $v_j$ are of the same type
            while to go through all $v_i$ and $v_j$ labels
                if the attribute name $f_{vp}(v_i)$ and $f_{vs}(v_j)$ are equal
                        evaluate value $f_{vs}(v_j)$ in the range $f_{vp}(v_i)$
                        compute the range bound $f_{vp}(v_i)$closest to $f_{vs}(v_j)$
                        $cost = cost + Abs$ ($f_{vs}(v_j)$ value $-$ closest bound of $f_{vp}(v_i)$)
                if $f_{vs}(v_j)$ is within the range $f_{vp}(v_i)$
                    $fit++$
                else
                    $nofit++$
                    $cost++$
                        End if
                    End if
                End while
            End if
        End while
    End if
Return cost
End
```

Algorithm 1. Nodes Labels Cost Function Adapted (author's own source)

The ranges are expressed as nodes labels and are evaluated using the nodes labels similarity function. The matching of these label seeks to find the closeness between a value given by an instance and the value closest to this within the range. However, it is not only taking into account this last parameter but a weight is defined in case the value of the instance is not in the indicated range [30].

Likewise, a edges labels cost function is performed, where only those labels of the same type of node are compared, in case the labels are not the same, it is assigned a weight, the pseudocode is shown in Algorithm 2.

Edges Labels Cost Function

```
Edge labels cost function Procedure $(V_p, E_p, f_{ep}, V_s, E_s, f_{es})$
Start $cost = 0$
If $e_i$ is an edge between ($v_i$, $v_i$') and $e_j$ is an edge between ($v_j$, $v_j{}'$)
    If $v_i$ and $v_j$ are of the same type and $v_i$' y $v_j$' are of the same type
        If $f_{ep}(e_i)! = f_{es}(e_j)$
            $cost++$
        End if
    End if
End if
Return cost
End
```

Algorithm 2. Edges Labels Cost Function Adapted (author's own source)

v_i represents a node of a pattern and $v_i^{'}$ is another node of the pattern, analogously to the nodes of instances v_j and $v_j^{'}$, just when the edge belongs to nodes of the same type in the pattern and in the instance the edge label cost is computed.

In addition, it was necessary to consider the inversion of an edge that occurs in the case that two edges one of each graph are connected by the same type of nodes, but have opposite directions.

Edges Inversion Cost Function

```
Edge inversion function procedure $(V_p, E_p, V_s, E_s)$
Start $cost = 0$
If $e_i$ is an edge ($v_i$, $v_i$') and $e_j$ is an edge ($v_j$, $v_j$')
    If $v_i$ and $v_j$' are of the same type and $v_i$' and $v_j$ are of the same type
        $cost = +0{,}5$
    End if
End if
Return cost
End
```

Algorithm 3. Edges Inversion Cost Function Adapted (author's own source)

Adaptation of the Similarity Function to the Selected Algorithms. In Fig. 2 it is possible to observe the adaptation made to the graph matching algorithms, as already mentioned previously. The algorithms evaluate the structural similarity by means of the edit distance, also these have a similarity function, which is conformed by the nodes labels cost function and the edges labels cost function.

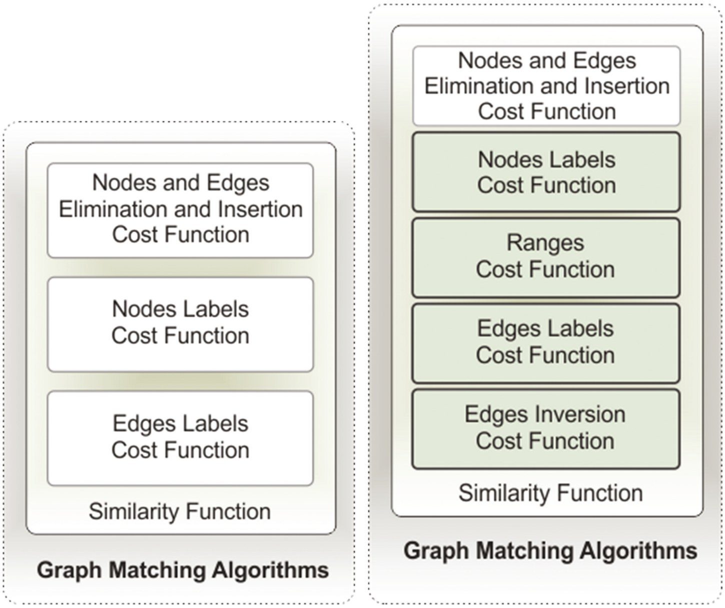


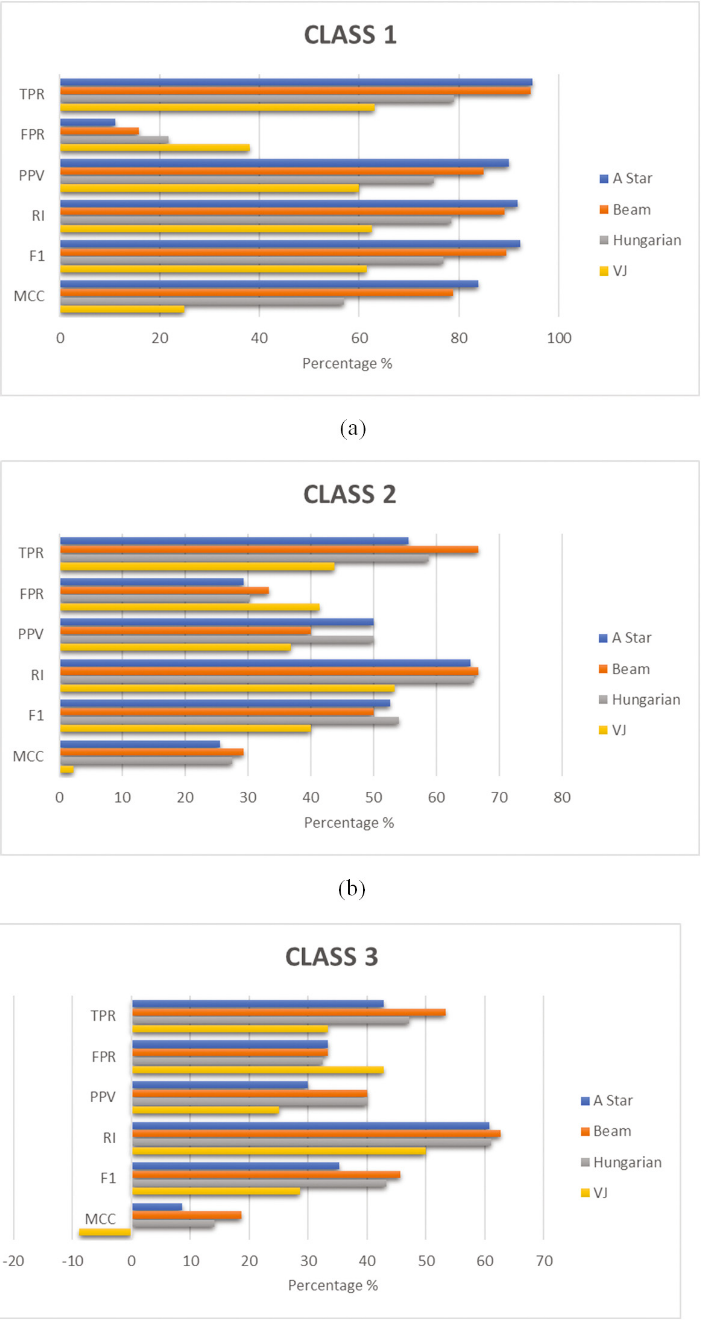
Fig. 2. Comparison between original algorithm and adaptation performed.

To achieve the objectives of this research, an edges inversion cost function is added to the similarity function, also a nodes labels function is adapted, this one allows to handle the labels expressed as ranges in order to find the minimum difference between the range bounds and the evaluated value. Besides, a ranges cost function that find if a evaluated value is in a defined range and an edges inversion cost function is added. With this adaptation, a greater flexibility in the found matches related to the ranges defined by the used patterns is sought.

The normalization of nodes and edges elimination and insertion cost function, range cost function, edges labels cost function and edges inversion cost function results is obtained dividing the cost of each of them by the number of possible comparisons, and the nodes labels cost function results is normalized dividing the result between the sum of the maximum differences of the instances nodes labels value and the patterns nodes label range. The results obtained in each of the previously mentioned functions are normalized so that the values are in a range of (0, 1), where 1 indicates that the graphs are exactly the same and 0 indicates that the graphs are totally different. After normalizing the cost functions results is assigned a weighting to each one of them, this weighting is obtained from the weights assigned to the cost functions and allow to take into account which results are more relevant, characterizing the application domain in a better way.

4 Results

In this section, we present an experimental evaluation of the algorithms discussed, which seek to find a classification of the most similar patterns with an instance evaluated in order to find the coffee rust infection rate in a crop. The results obtained by the



(a)

(b)

(c)

Fig. 3. (a) Precision Comparison of Class 1, (b) Precision Comparison of Class 2 and (c) Precision Comparison of Class 3.

prototype are compared with the evaluation provided by an expert, which are taken as a reference point for the experimental evaluation of the prototype, to determine the precision level of the algorithms and classify them based on the precision and runtime. It is important to mention that the patterns and instances that were used to evaluate the system were indicated in Sect. 2.

The expert analyzes the patterns and instances and generates a classification of the first three patterns most similar to the instances, where the three patterns are located in three classes depending on their similarity to the instance, being class 1 the one that containing the most similar pattern and the class 3 has the less similar pattern to an evaluated instance. From the comparison of the results obtained by the expert and those thrown by the algorithms, the precision and accuracy of the selected algorithms is established by means of the following parameters: True Positive Rate (*TPR*), False Positive Rate (*FPR*), Positive Predictive Value (*PPV*), Rand Index (*RI*), F-Measure (*F1*) and Matthews Correlation Coefficient (*MCC*). The results of algorithms precision and accuracy comparison are shown in Fig. 3.

In the most significant classification for the domain of the problem (Class 1) Fig. 3(a), the algorithm A* is the one that presents a better precision, accuracy and quality in the results, surpassing considerably the two bipartite suboptimal algorithms (Hungarian and VJ). In the other two classes Fig. 3(b) and (c), the Beam and Hungarian algorithm obtain the best results.

Finally, in the Fig. 4 it is observed that with weighting values of 70%, 20% and 10% for class 1, 2 and 3 respectively, the A* algorithm is the most precise and accurate one, followed by the Beam, Hungarian and VJ algorithms, being the VJ algorithm the one that gives the least precision. With the above it can be observed that the algorithm A* predicts the rate of infection of coffee rust better than the other algorithms.

On the other hand, to evaluate the runtime of the algorithms a set of synthetic graphs analogous to the patterns (SPG) and instances (GIS) in which the agroclimatic data and each one of its labels are represented was generated because the real dataset does not contain graphs with enough size to perform this test. These graphs have

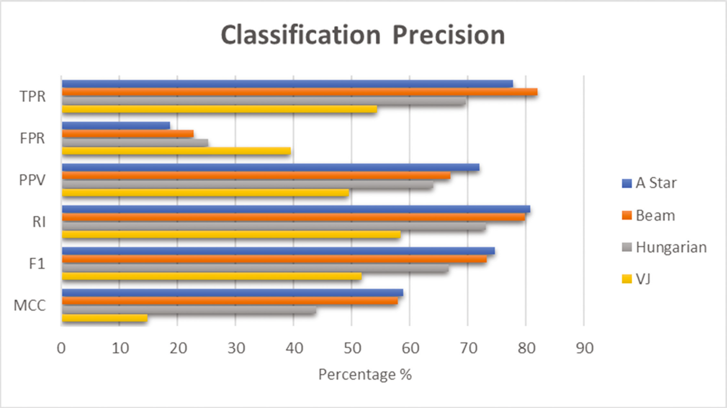
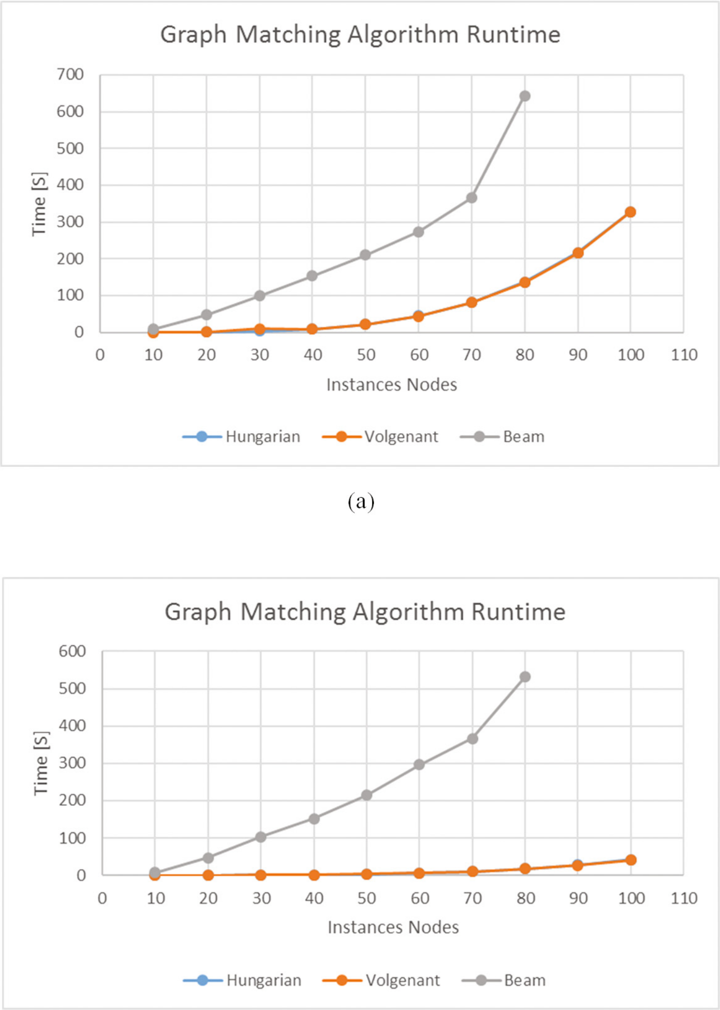


Fig. 4. Precision Classification.

different sizes and random values in a set range, with this, it is sought to determine the relationship between the number of nodes and edges with the runtime. The tests were performed on a computer with 8192 MB of RAM and an Intel Core i5-4210U @ 2.4 Ghz (4CPUs) processor. The aim of the runtime test is to analyze the behavior of the proposed adaptation on the algorithms with the purpose to find the one that have the better ratio between runtime and precision.

Initially, it is not considered A* algorithm for the analysis of the execution time, because this does not present a wide range in the number of nodes that can be matched



(a)

(b)

Fig. 5. (a). Graph matching algorithms runtime using directed biconnected graph (b) Graph matching algorithms runtime using directed connected graph.

before the memory overflows, which does not make it comparable with other algorithms. On the other hand, in Fig. 5(a) and (b) a comparison between the execution time of the remaining three algorithms is shown, the number of nodes and the edges of the (SIG) are varied and matched with the (SPG) of 10 nodes each, this last value was selected based on the size of the graphs actual patterns present in our application domain. In (a) the graphs are directed and biconnected, indicating that for each pair of nodes there are two edges that join them, in (b) the graphs are directed and connected; Which implies that in the tests carried out in (a) the number of edges processed is doubled with respect to (b). These tests are performed in that way in order to analyze how nodes and edges influence the runtime results obtained by each algorithm.

Finally, in Fig. 5(a) and (b) we can observe the behavior of the algorithms between runtime and the variation of nodes and edges, these charts indicate that the Beam algorithm has a longer execution time compared to the Hungarian and VJ algorithms. The number of edges significantly affect the results, especially in bipartite algorithms.

5 Conclusions and Future Works

The main contribution of this paper is the adaptation of a similarity function to the previous graph matching algorithms, the results of this adaptation show that the algorithm A* has a slightly higher precision, however the runtime increases considerably. The Hungarian and VJ algorithms perform the graph matching in a much faster way but the obtained results are unreliable. On the other hand, Beam algorithm presents a better performance in relation to coffee rust problem, because it maintains a good ratio between precision and runtime. Additionally, their results can not only be taken as a characterization of coffee rust in a crop, but also represent the incoming conditions of the disease in the next month.

The proposed adaptation allows us to better characterize and generate warnings about the occurrence of coffee rust in relation to other approximations, since the error-tolerant graph matching considers a degree of proximity in the search task of graph patterns for a crop disease; In addition, this type of pairing provides greater flexibility in the matches found corresponding to the ranges defined by the patterns used.

As future work, we will integrate a function to change the parameterization of the weights of the editing operations established by default. Furthermore, we will explore the behavior of the proposed adaptation for other types of crops in order to verify its scalability and to propose as a possible general solution to the problem of prediction of occurrence of pests and diseases in the agricultural sector.

Acknowledgments. The authors are grateful to the Telematics Engineering Group (GIT) of the university of Cauca, the Colombian Administrative Department of Science, Technology and Innovation (Colciencias), AgroCloud project of The Interinstitutional Network of Climate Change and Food Security of Colombia (RICCLISA) for supporting this research and Supracafé for providing the dataset used.

References

1. Rivillas, C., Serna, C., Cristancho, M., Gaitán, A.: Roya del Cafeto en Colombia: Impacto, Manejo y Costos del Control. In: Chinchiná Bol. Téc., no. 36 (2011)
2. Perez, A.L.T., Vigil, A.L.M., Morales, M.E.S.: Producción de café (Coffea arabica L.): cultivo, beneficio, plagas y enfermedades. In: Temas selectos de Ingenieria de alimentos, pp. 54–74 (2011)
3. Waller, J.M., Bigger, M., Hillocks, R.J.: Coffee pests, diseases and their management. CABI (2007)
4. Muller, R.A., Berry, D., Avelino, J., Bieysse, D.: Coffee diseases. In: Coffee: Growing, Processing, Sustainable Production: A Guidebook for Growers, Processors, Traders, and Researchers, pp. 491–545 (2004)
5. Avelino, J., Willocquet, L., Savary, S.: Effects of crop management patterns on coffee rust epidemics. Plant. Pathol. **53**(5), 541–547 (2004)
6. Kushalappa, A.C., Eskes, A.B.: Advances in coffee rust research. Annu. Rev. Phytopathol. **27**(1), 503–531 (1989)
7. Corrales, D.C., Ledezma, A., Hoyos, J., Figueroa, A., Corrales, J.C.: A new dataset for coffee rust detection in Colombian crops based on classifiers. Sist. y Telemática **12**(29), 9–23 (2014)
8. Perez-Ariza, C.B., Nicholson, A.E., Flores, M.J.: Prediction of coffee rust disease using bayesian networks. In: Proceedings of the Sixth European Workshop on Probabilistic Graphical Models, pp. 259–266 (2012)
9. Lasso, E., Corrales, J.C.: Expert system for crop disease based on graph pattern matching: a proposal. Revista Ingenierías Universidad de Medellín **15**(29), 81–98 (2016)
10. Lasso, E., Ortega, S.M., Corrales, J.C.: Enriquecimiento semántico e inferencia de relaciones en una red social en línea. Ing. Univ. **17**(2), 355–373 (2013)
11. Conte, D., Foggia, P., Sansone, C., Vento, M.: Thirty years of graph matching in pattern recognition. Int. J. Pattern Recognit. Artif. Intell. **18**(3), 265–298 (2004)
12. Riesen, K., Bunke, H.: Approximate graph edit distance computation by means of bipartite graph matching. Image Vis. Comput. **27**(7), 950–959 (2009)
13. Fischer, A., Suen, C.Y., Frinken, V., Riesen, K., Bunke, H.A.: Fast matching algorithm for graph-based handwriting recognition. In: International Workshop on Graph-Based Representations in Pattern Recognition, pp. 194–203 (2013)
14. Nikolic, M., Measuring similarity of graphs and their nodes by neighbor matching. In: ArXiv preprint arXiv:1009.5290 (2010)
15. Abu-Aisheh, Z., Raveaux, R., Ramel, J.Y., Martineau, P.: An exact graph edit distance algorithm for solving pattern recognition problems. In: 4th International Conference on Pattern Recognition Application and Methods (2015)
16. Lasso, E., Thamada, T.T., Meira, C.A.A., Corrales, J.C.: Graph patterns as representation of rules extracted from decision trees for coffee rust detection. In: Garoufallou, E., Hartley, R.J., Gaitanou, P. (eds.) Metadata and Semantics Research, pp. 405–414. Springer, Heidelberg (2015)
17. Corrales, D., Ledezma, A., Peña, A., Hoyos, J., Figueroa, A., Corrales, J.: A new dataset for coffee rust detection in Colombian crops base on classifiers. Revista S&T **12**(29), 9–23 (2014)
18. Holt, R.C., Winter, A., Schurr, A.: GXL: toward a standard exchange format. In: 2000 Proceedings of Seventh Working Conference on Reverse Engineering, pp. 162–171 (2000)

19. Lasinger, K.: Overview of existing software tools for graph matching. Technical report, Global Grid Forum Pattern Recognition and Image Processing Group, Institute of Computer Aided Automation, Vienna University of Technology (2003)
20. Serratosa, F.: Fast computation of bipartite graph matching. Pattern Recognit. Lett. **45**, 244–250 (2014)
21. Thamada, T.T., Neto, C.D.G., Meira, C.A.A.: Sistema de alerta da ferrugem do cafeeiro: resultado de um processo de mineração de dados. In: Embrapa Informática Agropecuária-Artigo em anais de congresso (ALICE). In: Congresso Brasileiro de Agroinformática, 9, 2013, Cuiabá. Agroinformática: inovação para a sustentabilidade do agronegócio brasileiro: anais. Cuiabá: Universidade Federal de Mato Grosso (2013)
22. Meira C.A.A., Rodrigues, L.H.A.: Mineração de dados no desenvolvimento de sistemas de alerta contra doençias de culturas agrícolas. In: Congresso Brasileiro de Agroinformática, vol. 5 (2005)
23. Neto, C.G., Rodrigues, L.H.A., Meira, C.A.A.: Modelos de predição da ferrugem do cafeeiro (Hemileia vastatrix Berkeley & Broome) por técnicas de mineração de dados. In: Coffee Science, vol. 9, no. 3, pp. 408–418 (2014)
24. Hart, P.E., Nilsson, N.J., Raphael, B.: A formal basis for the heuristic determination of minimum cost paths. IEEE Trans. Syst. Sci. Cybern. **4**(2), 100–107 (1968)
25. Neuhaus, M., Riesen, K., Bunke, H.: Fast suboptimal algorithms for the computation of graph edit distance. In: Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR), pp. 163–172 (2006)
26. Riesen, K., Neuhaus, M., Bunke, H.: Bipartite graph matching for computing the edit distance of graphs. In: International Workshop on Graph-Based Representations in Pattern Recognition, pp. 1–12. Springer, Heidelberg (2007)
27. Bellur, U., Kulkarni, R.: Improved matchmaking algorithm for semantic web services based on bipartite graph matching. In: IEEE International Conference on Web Services, ICWS 2007, pp. 86–93 (2007)
28. Fankhauser, S., Riesen, K., Bunke, H.: Speeding up graph edit distance computation through fast bipartite matching. In: International Workshop on Graph-Based Representations in Pattern Recognition, pp. 102–111. Springer, Heidelberg (2011)
29. Fan, W., Li, J., Ma, S., Tang, N., Wu, Y., Wu, Y.: Graph pattern matching: from intractable to polynomial time. VLDB Endow. **3**(1–2), 264–275 (2010)
30. Lasso, E.: Sistema experto basado en emparejamiento de patrones, Popayán, Cauca, Colombia (2016)

A Guideline for Building Large Coffee Rust Samples Applying Machine Learning Methods

Jhonn Pablo Rodríguez(✉), Edwar Javier Girón, David Camilo Corrales, and Juan Carlos Corrales

Department of Telematics Engineering, Engineering Telematics Group,
University of Cauca, Popayán, Colombia
{jhonnpablo,edwardgb,dcorrales,jcorral}@unicauca.edu.co
http://www.unicauca.edu.co

Abstract. Coffee rust has become a serious concern for many coffee farmers and manufacturers. The American Phytopathological Society discusses its importance saying this: "the most economically important coffee disease in the world," while "in monetary value, coffee is the most important agricultural product in international trade". The early detection has inspired researchers to apply supervised learning algorithms on predicting the disease appearance. However, the main drawback of the related works is the few data samples of the dependent variable: Incidence Rate of Rust, since the datasets do not have a reliable representation of the disease, which will generate inaccurate classifiers. This paper provides a guide to increase coffee rust samples applying machine learning methods through a systematic review about coffee rust in order to select appropriate algorithms to increase rust samples.

Keywords: Synthetic data · Incidence · Crops · Dataset

1 Introduction

Coffee rust has become a serious concern for many coffee farmers and manufacturers. The American Phytopathological Society discusses its importance saying this: "the most economically important coffee disease in the world," while "in monetary value, coffee is the most important agricultural product in international trade". Without a solution, the effects on the coffee industry may soon be reflected in price and availability [1].

For several years, the disease was managed through the combination of various techniques such as quarantine, cultural management, fungicides and resistant crops. Due to the effectiveness of chemical control and the relatively limited damage caused by the disease, particularly at high altitudes, Mesoamerican coffee farmers and technical authorities considered it manageable. This view prevailed until the epidemic between 2008 and 2013 along Mesoamerica, from Colombia to Mexico, including Peru, Ecuador and some Caribbean countries [2]. Coffee farmers were desperate to obtain an answer to this terrible situation since the intensity was higher than anything previously observed, affecting a large number of countries including: Colombia, from 2008 to 2011, affecting an average of 31% of coffee production compared with the production in 2007; Central

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_8

America and Mexico, in 2012–13, affecting an average of 16% of the production in 2013 compared with 2011–12 and an average of 10% in 2013–14 compared with 2012–13; and Peru and Ecuador in 2013 [2]. More specifically, in 2013, the Guatemalan government and the Guatemalan Nation Coffee agency declared a national state of emergency after a projection of nearly 15% crop loss in their region. The devastation has continued to spread due to higher temperatures in this region, which are making fungus growth at higher altitudes possible [3]. Higher temperatures may be linked to climate change. And several/many experts are worried about the persistence of these conditions (high temperatures) will not change in the near future. In this regard, several reports and experts proposed solutions related with early detection of the disease and the eradication of infected plants.

The early detection has inspired researchers to apply supervised learning algorithms on predicting the disease appearance. The data collected about conditions and soil fertility properties, physical properties and management of a coffee crop, can be used to forecast the rust infection rate. In the same way, weather conditions such as the minimum and maximum levels of temperature, humidity and rainy days can help to estimate the behavior of the disease. Several Colombian and Brazilian researches in supervised learning attempt to detect the incidence rate of rust (IRR) in coffee crops using Neural Networks, Decision Trees, Support Vector Machines, Bayesian Networks, K Nearest Neighbor, and Ensemble Methods [4–8]. However, the main drawback of the related works is the few data samples of the dependent variable: Incidence Rate of Rust, since the datasets do not have a reliable representation of the disease, which will generate inaccurate classifiers [5].

This paper provides a guide to increase coffee rust samples applying machine learning methods through a systematic review about coffee rust in order to select appropriate algorithms to increase rust samples. The paper is structured as follows: in Sect. 2, we describe the coffee rust disease and supervised learning concepts. Section 3 exposes the supervised learning approaches applied to coffee rust detection and the main challenges due to low accuracy of rust detecting models; Sect. 4 shows a systematic review of the approaches to generate synthetic data. Section 5, proposes a guideline for building large dataset of coffee rust based on systematic review of Sect. 4. Finally, the Sect. 6 presents the conclusions.

2 Background

2.1 Coffee Rust

Coffee rust is caused by the fungus Hemileia vastatrix, among the cultivated species C. Arabica is the most severely attacked. The disease causes defoliation, sometimes this one can lead to death of branches and crop losses. The first symptoms are small yellowish lesions that appear on the underside of the leaves, where the fungus has penetrated through the stomata. These lesions grow, coalesce and produce uredospores with their distinctive orange color. Chlorotic spots can be observed on the upper surface of the leaves. During the last stage of the disease, lesions become necrotic [2]. The progression of coffee rust depends on four factors that appear simultaneously [9]:

- **The host:** There are varieties of coffee plants susceptible and resistant to rust. Varieties such as Típica, Borbón and Caturra suffer severe rust attacks, while Colombia and Castillo varieties are highly resistant to rust.
- **Pathogenic organism:** Hemileia vastatrix lifecycle begins with the germination of uredospores in 2–4 h in optimal conditions. Within 24–48 h, infection is completed. Once the infection is completed, the underside of the leaf is colonized and sporulation will occur through the stomata [10].
- **Weather conditions:** Weather with constant precipitations, mainly in the afternoon and night with cloudy sky, high humidity in the plants and low temperatures are relevant factors for germination of rust. Spread of disease and its development is usually limited to the rainy season, while in dry periods the rust incidence is very low.
- **Agronomic practices:** This practice refers to properties of crop sowing (plant spacing, percentage of shade, etc.), application of fungicides and fertilizations on coffee crops with the aim to avoid several rust attacks.

In Colombia the incidence of rust is measured through the methodology developed by Centro Nacional de Investigaciones de Café (Cenicafé) [9], which is explained as follows:

Incidence Rate of Rust (IRR) is calculated for a plot with area lower or equal of one hectare. The methodology is composed by three steps:

1. The farmer must be standing in the middle of the first furrow, choose one coffee tree and pick out the branch with greater foliage for each level (high, medium, low); the leaves of the selected branches are counted as well as the infected ones for rust.
2. The farmer must repeat the step 1 for every tree in the plot until 60 trees are selected. It is worth mentioning that the same number of trees must be selected in every furrow (e.g. if the plot has 30 furrows, the farmer selects two coffee trees for each furrow).
3. Once the step 1 and 2 are finished, the Leaves of the selected Coffee Trees (LCT) are added as well as the Infected Leaves by Rust (ILR). Then, the Incidence Rate of Rust (IRR) must be computed using the following formula:

$$IRR = \frac{ILR}{LCT} \times 100 \tag{1}$$

Furthermore in Brazil, Procafé Foundation [11] proposes a methodology that enables measuring the incidence of rust through the following steps:

1. The farmer selects a random region of 4 m^2 from the plot.
2. From the region selected in the previous step, the farmer selects two coffee trees, located in different furrows, one in front of the other.
3. The trees are divided into three levels according to its height: high, medium, and low; and the branches are divided in quartiles by its size. One branch of the middle zone of the plant is chosen, then two leaves of that area are taken from the third or fourth part of the branch.

Finally, for each plot 25 regions are selected, from which 50 plants are chosen, therefore 100 leaves are collected. The samples are taken the first two days of each month.

2.2 Supervised Learning Techniques

Based on [4], in this section the main supervised learning algorithms, from coffee rust domain are explained:

A supervised learning process is based on the iteration of a training process from a dataset named training data. The training data consist of a set of examples. Each example is represented by a pair (x_i, y_i), where x_i is an attributes vector and y_i is the desired output value (also called class) of the example. A SL algorithm analyzes the training data and produces an inferred function which is called classifier (if y_i is discrete) or regression function (if y_i is continuous) [12]. In Table 1 an example of a dataset with three attributes is presented: Number of Days of Precipitation (NDP), average Daily Nighttime relative Humidity (DNH), average Daily Minimum Temperatures (DMT), which can take different values (discrete or continuous); and the desired output value [4].

Table 1. Four sample for training dataset

Attribute			Output value
NDP	DNH	DMT	IRR
2	96.1	14	65.23%
3	93.8	16	62.54%
4	95.7	15	57.32%
1	98.2	14	61.12%

The most commonly APPLIED algorithms are Decision Trees (DT), Bayesian Networks (BN), Artificial Neural Networks (ANN), Support Vector Machines (SVM) and K Nearest Neighbor (KNN) [4].

3 Supervised Learning Techniques in Coffee Rust Detection

This systematic review took into account the inclusion criterion: Colombian and Brazilian research using supervised learning algorithms. And as an exclusion criterion: investigations not greater than 8 years. Systematic review was based on the following research question:

Are there researches that address the coffee rust by supervised learning?

18 papers were found (2008–2016) from 4 sources of information: IEEE Xplore (1 paper), ScienceDirect (1 paper), Springer Link (6 paper) and Google Scholar (10 papers). We defined 2 search queries: “coffee rust prediction” and “coffee rust detection”.

The papers found in the systematic review, contain approaches for coffee rust detection in Colombian and Brazilian crops, using supervised learning algorithms. These researches are detailed below:

3.1 Supervised Learning for Coffee Rust Detection in Colombian Crops

The dataset built in the Colombian researches was obtained for 18 plots from experimental farm Los Naranjos (21° 35′08″ N, 76° 32′53″ W), of the Supracafé enterprise, located in Cajibío (Cauca) [26]. The samples were recollected among years 2011–2013, with 147 instances and 21 attributes: 6 of weather conditions, 5 soil fertility properties, 6 physic crop properties, 4 crop management; the class represents the Incidence Rate of Rust (IRR). This dataset was used in several works presented in Table 2:

Table 2. Related works for coffee rust detection in Colombian crops

Work	Year	Algorithm
[13]	2014	ANN, SVM, RT
[5]	2015	ANN, SVM, RT
[14]	2016	Two-level classifier ensembles using Back Propagation Neural Networks, Regression Tree M5 and Support Vector Regression
[15]	2014	SVM
[16]	2015	DT

Briefly, Colombian researchers using SVM (4 papers), ANN (2 papers) and DT/RT (4 paper), but the most used are SVM and ANN, since this class of algorithms deliver results accurate to the end user, these are less prone to overfitting than other methods, tolerance to the noise, Accuracy in general, Tolerance to irrelevant attributes, easy to understand and speed in its learning and classification [4].

3.2 Supervised Learning for Coffee Rust Detection in Brazilian Crops

Brazilian researchers built a dataset from physic crop properties and weather conditions. These data were collected in the experimental farm Procafé (South latitude 21° 34′00″ longitude West 45° 24′22″ and altitude 940 m) located in Varginha, Minas Gerais, during the years 1998–2011 [17]. The final dataset includes 182 instances. Table 3 are presented the works related with the Brazilian dataset:

Thus, the algorithms used in Brazil are DT, SVM, SVR, ANN, RF and BN, focusing in its priority to generate an easy interpretation model based on graphs like the DT and generate accurate results with SVM.

Table 3. Related works for coffee rust detection in Brazilian crops

Work	Year	Algorithm
[17]	2008	DT
[18]	2012	BN
[19]	2013	ANN, DT, Random Forest (RF), SVM
[8]	2015	Ensembles Methods with SVM, ANN, DT
[20]	2009	DT
[6]	2011	DT
[21]	2014	SVM, ANN, DT, RF
[22]	2011	SVR
[23]	2010	SVR

3.3 Discussion

The algorithms used in the last years are Support Vector Machines (SVM), Decision Trees (DT), Bayesian Networks (BN), Nearest Neighbor (KNN) and Artificial Neural Networks (ANN), but the algorithms SVM and ANN are the most used for the precision in the categorization of results, however the algorithms present deficiencies in their interpretation, since classifiers built by these algorithms do not generate a visual representation, in contrast to the DT and algorithms the BN, which allow the user to observe the classifier through a representation based on graphs [4]. On the other hand, algorithms such as: KNN and BN are good by their speed of learning (training phase) [24–26].

In accordance with the system review, Brazil and Colombia are the countries that address the coffee rust detection through supervised learning. However, the researchers found are limited due lack of data in measures of Infection Rate of Rust, due to the high costs and time invested for the collection of data rust infection. As a result, datasets cannot represent faithfully the total population, generating low accuracy in the results obtained by classifiers [5]. Section 4 describes approaches for resolving this kind of problem, through the generation of synthetic data.

4 Synthetic Data Generation

The few amount of samples do not let to the models to represent important characteristics of the population, therefore the models constructed are affected in its precision [5]. For this systematic review, the inclusion criterion was taken into account: research to increase the number of samples in a dataset. And as an exclusion criterion: proposed researches that do not have a benchmark analysis with traditional algorithms. Systematic review was based on the following research question:

Which are synthetic data approaches most used for lack of data?

We found 26 papers (2000–2015), considering 5 search queries: “Synthetic Data Generation”, “Imbalanced Dataset”, “Over-Sampling”, “Virtual Sample Generation” and “Interpolation Algorithm”, it was found from 4 sources of information: IEEE Xplore

(10 papers), ScienceDirect (5 papers), Springer Link (7 papers) and Google Scholar (4 papers).

From the systematic review conducted previously, were found 4 approaches to addressing the lack of data and their respective algorithms for synthetic data generation as shown in Table 4:

Table 4. Approaches and synthetic data algorithms

Approach	Algorithms	Description	Works
Over-Sampling	MDO, SMOTE, ADASYN, RWO SAMPLING, BORDERLINE SMOTE2 MSMOTE, BORDERLINE SMOTE1, BORDERLINE SMOTE, C-SMOTE, SMOTE-I, DSMOTE	Creates new synthetic instances for the minority class	[27–39]
Interpolation	Stair, S-Spline, Bicubic, Lanczos, Nearest Neighbor, Fractals, Linear	Determinates the values of a function at positions lying between its samples	[40–44]
Classifiers	AdaBoost, RAMOBoost, AdaBoost.M2, AdaBoost.M1, DataBoost.IM, SMOTEBoost, DataBoost, SMOTE Bagging, OverBagging	Classification or regression model that aims to predict the value of output variable from certain independent input variables	[30, 35, 36, 45–48]
Copy of Data	Bagging, Regression Trees, Random Forest, Statistical Distributions	Creates a copy from original data with a different representation that not reveal private information	[49–51]

To construct the guideline for increasing coffee rust samples, we used the approaches: oversampling, interpolation, and classifiers. Copy of data is not used because for this kind of problem is necessary use the original representation of data.

5 A Guideline for Building Large Coffee Rust Samples

In this section, we propose a guide to increase coffee rust samples. Figure 1 presents the process for generate discrete or numeric rust samples.

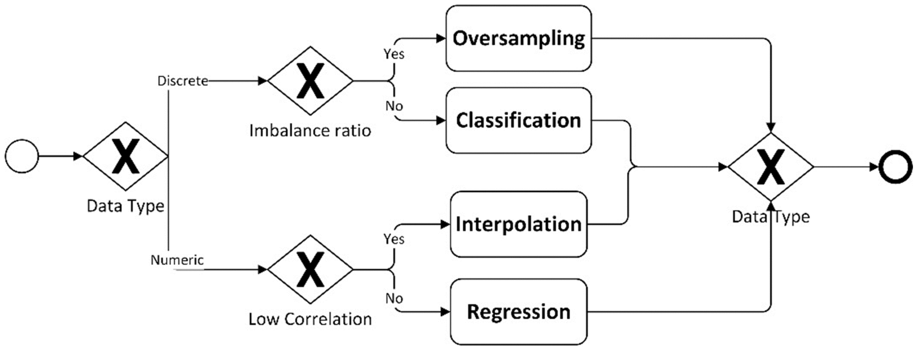


Fig. 1. Guide for increase coffee rust samples

When the coffee rust samples are discrete, the imbalance ratio must be evaluated. Commonly the Imbalance Ratio (IR) is used to measure the distribution of the classes:

$$IR = \frac{C^{+}}{C^{-}} \tag{2}$$

Where C+ represents the size of the majority class and C− the size of the minority class [52]. *Oversampling* techniques are used if IR > 1; otherwise *Classification* techniques are applied. The techniques for increase the discrete coffee rust samples are presented below.

5.1 Oversampling

Oversampling is used to increase examples from minority classes with aim have equitable distribution of the classes [28]. The algorithm most used are called SMOTE: Synthetic Minority Over-sampling Technique [27]; RUS Boost [55], Balance Cascade [56] and Easy Ensemble [56] also are considered Oversampling algorithms. Other algorithms used are MDO [35], SMOTE [31, 33, 39], ADASYN [29, 30], RWO SAMPLING [37], BORDERLINE SMOTE2 [38], MSMOTE [36], BORDERLINE SMOTE1 [38], BORDERLINE SMOTE [38], C-SMOTE [32], SMOTE-I [57], DSMOTE [34].

5.2 Classification

Classification algorithms are efficient methods to increase discrete samples where each new value is obtained from related cases in the whole set of records. Besides the capability to increase the coffee rust samples with plausible values that are as close as possible to the true value, classification algorithms should preserve the original data structure and avoid to distort the distribution of the original samples. The algorithms based in neighbours are the most used [58]. The algorithm most used are AdaBoost [30, 35, 36, 45], RAMOBoost [30], AdaBoost.M2 [36], AdaBoost.M1 [36], DataBoost.IM [59], SMOTEBoost [30, 36, 46], DataBoost [48], SMOTE Bagging [36], OverBagging [36].

In case of the coffee rust samples are numeric, a correlation analysis must be done. The range of values for the correlation coefficient is -1 to 1. A correlation of -1 indicates a perfect negative correlation, while a correlation of 1 indicates a perfect positive correlation. Values close to 0 indicates a low correlation. A regression approach is used when the dependent variables have a relationship strong with the predictor variables (correlation coefficient close to 1 or -1) [53]; otherwise the interpolation approach is used (correlation coefficient close 0) because these methods are focused in mathematic functions [54]. The techniques for increase the numeric coffee rust samples are presented below.

5.3 Interpolation

Interpolation is the process of determining the values of a function at positions lying between its samples. It achieves this process by fitting a continuous function through the numeric input samples. The interpolation can be addressed of two ways: univariate and multivariate [60].

Univariate Interpolation

The univariate interpolation is defined for values of a function f(x) = y, where only take part two variables (x, y), in which x is the series with full data and y is the incompleteness variable in the serie. Inside univariate interpolation, there are algorithms that interpolate the y values, of which three obtain good results: lineal interpolation, K-nearest neighborhood (KNN), and cubic spline interpolation [61–64].

Multivariate Interpolation

The univariate interpolation fitted of two-dimensional data points, while the multivariate interpolation make the points fit finding the surface that provides an exact fit to a series of multidimensional data points, considering a series of N distinct dimensional data points (x1, y1), (x2, y2), … (xN, yN), where $Xi = \left(X_i^1, X_i^2, \ \ldots \ X_i^d\right)$ is a vector for each i = 1, 2, … N. By this interpolation we find a function f: $R^d \rightarrow R$ such that: f(x1) = y1, f(x2) = y2, … f(xN) = yN [65]. The algorithms most used are Inverse Distance Weighting (IDW) and Kriging [66].

5.4 Regression

Similarly, to Classification algorithms, the coffee rust is treated as dependent variable and a regression is performed to increase coffee rust samples. Regression analysis is a machine learning approach that aims predict the value of continuous output variables (coffee rust samples) from certain independent input variables (e.g. temperature, humidity, etc.), via automatic estimation of their latent relationship from data [67]. Linear regression, logistic regression [68], regression trees [69], support vector regression [70] and multi-layer perceptron [71] are typical choices.

6 Conclusions

In this paper, a guideline for increasing the coffee rust samples was made. The first step to create the guideline was to make a systematic review about coffee rust in order to select appropriate algorithms to increase rust samples.

When the guideline for increasing the coffee rust samples is used, we consider important follow the next observations [72]:

- *Distance between meteorological station and coffee trees:* if a weather station is away from a coffee plot, the weather measurements are inaccurate, because coffee plot can have micro-climate influenced by: coffee plot orography and properties of crop sowing such as: plant spacing, shade on coffee trees, etc. Unfortunately, the weather stations are very expensive to have one per coffee plot.
- *Information about application of fungicides on coffee plots:* if fungicides are applied on coffee plots before germination of rust, the weather conditions can not be relevant factors to increase the rust incidence. We consider necessary this information to build a correct regression model based on meteorological variables.
- *Consider a margin of error in IRR measurements:* the insufficient data due to the expensive collection process that requires large expenditures of money and time [7]. The farmers must select 3 branches for each 60 coffee trees (minimum) per plot [23]. Usually one plot have 10000, 5000, or 2500 coffee trees [1], given that the maximum number of IRR measures that we can obtain for one plot are: 0.6%, 1.2%, or 2.4% respectively. Besides a coffee farm has over one coffee plot.
- Due to the coffee rust is an element of a agronomic pathosystem, it is necessary to study how the factors about crop administration (fungicides, production levels, and so on), the meteorological factors (humidity, temperature and so on) and the fungus development interact with each other, and how this interaction contribute to increase the coffee rust disease.

7 Acknowledgments

We thank to the Telematics Engineering Group (GIT) of the University of Cauca for the technical support. Finally, this work has been partially supported by AgroCloud project of the RICCLISA Program, and Colciencias for PhD scholarship granted to MsC. David Camilo Corrales and Cauca Innovacción Nucleus Group for the Young investigator program to Ing. Edwar Javier Girón.

References

1. Arneson, P.A.: Coffee rust. Plant Health Instr. (2000)
2. Avelino, J., et al.: The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. Food Secur. **7**(2), 303–321 (2015)
3. A Solution to the Coffee Rust Epidemic: How Spectrophotometry May Provide the Answers. HunterLab Horizons Blog, 12 January 2015

4. Corrales, D.C., Corrales, J.C., Figueroa-Casas, A.: Towards detecting crop diseases and pest by supervised learning. Ing. Univ. **19**(1), 207–228 (2015)
5. Corrales, D.C., Figueroa, A., Ledezma, A., Corrales, J.C.: An empirical multi-classifier for coffee rust detection in colombian crops. In: Gervasi, O., Murgante, B., Misra, S., Gavrilova, M.L., Rocha, A.M.A.C., Torre, C., Taniar, D., Apduhan, B.O. (eds.) Computational Science and Its Applications, ICCSA 2015, pp. 60–74. Springer, Heidelberg (2015)
6. Cintra, M.E., Meira, C.A.A., Monard, M.C., Camargo, H.A., Rodrigues, L.H.A.: The use of fuzzy decision trees for coffee rust warning in Brazilian crops. In: 2011 11th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 1347–1352 (2011)
7. Cesare di Girolamo, L.H.R.: Potencial de técnicas de mineração de dados para modelos de alerta da ferrugem do cafeeiro (2013)
8. Thamada, T.T., Rodrigues, L.H.A., Meira, C.A.A.: Predição da taxa de progresso da ferrugem do cafeeiro por meio de ensembles. Predicting infection rate of coffee rust by ensembles (2015)
9. Rivillas Osorio, C.A.: La roya del cafeto en Colombia, impacto, manejo y costos de control. Cenicafé: Chinchiná - Caldas - Colombia (2011)
10. Nutman, F.J., Roberts, F.M., Clarke, R.T.: Studies on the biology of Hemileia vastatrix Berk. & Br. Trans. Br. Mycol. Soc. **46**(1), 27–44 (1963)
11. Garcia, A.L.A.: RESUMO METODOLÓGICO DE AVALIAÇÃO DAS VARIÁVEIS FENOLÓGICAS E FITOSSÂNITÁRIAS DO SISTEMA DE AVISOS FITOSSÂNITÁRIOS DO MAPA/PROCAFÉ, Varginha, Brasil (2011)
12. Ng, A.: CS 229 machine learning course materials. In: Supervised learning. University of Stanford (2003)
13. Corrales, D.C., Ledezma, A., Andrés, J.P.Q., Hoyos, J., Figueroa, A., Corrales, J.C.: A new dataset for coffee rust detection in Colombian crops base on classifiers. Sist. Telemática **12**(29), 9–23 (2014)
14. Corrales, D.C., Casas, A.F., Ledezma, A., Corrales, J.C.: Two-level classifier ensembles for coffee rust estimation in colombian crops. Int. J. Agric. Environ. Inf. Syst. **7**, 41–59
15. Corrales, D.C., Peña, A.J.: Early warning system for coffee rust disease based on error correcting output codes: a proposal. Rev. Ing. Univ. Medellín **13**(25), 59–64 (2014)
16. Lasso, E., Thamada, T.T., Meira, C.A.A., Corrales, J.C.: Graph patterns as representation of rules extracted from decision trees for coffee rust detection. In: Garoufallou, E., Hartley, R.J., Gaitanou, P. (eds.) Metadata and Semantics Research, pp. 405–414. Springer, Heidelberg (2015)
17. Meira, C.A.A., Rodrigues, L.H.A., Moraes, S.A.: Análise da epidemia da ferrugem do cafeeiro com árvore de decisão. Trop. Plant Pathol. **33**(2), 114–124 (2008)
18. Pérez-Ariza, C.B., Nicholson, A.E., Flores, M.J.: Prediction of coffee rust disease using Bayesian networks. In: Andrés Cano, M.G.-O., Nielsen, T.D. (eds.) The Sixth European Workshop on Probabilistic Graphical Models. University of Granada, Granada, Spain (2012)
19. Cesare di Girolamo, L.H.R.: Desenvolvimento e seleção de modelos de alerta para a ferrugem do cafeeiro em anos de alta carga pendente de frutos (2013)
20. Meira, C.A.A., Rodrigues, L.H.A., de Moraes, S.A.: Warning models for coffee rust control in growing areas with large fruit load. Pesqui. Agropecuária Bras. **44**(3), 233–242 (2009)
21. di Girolamo Neto, C., Rodrigues, L.H.A., Meira, C.A.A.: Modelos de predição da ferrugem do cafeeiro (Hemileia vastatrix Berkeley & Broome) por técnicas de mineração de dados, 22 2014. http://www.alice.cnptia.embrapa.br/handle/doc/991078. Accessed 3 Feb 2016
22. Luaces, O., Rodrigues, L.H.A., Alves Meira, C.A., Bahamonde, A.: Using nondeterministic learners to alert on coffee rust disease. Expert Syst. Appl. **38**(11), 14276–14283 (2011)

23. Luaces, O., Rodrigues, L.H.A., Meira, C.A.A., Quevedo, J.R., Bahamonde, A.: Viability of an alarm predictor for coffee rust disease using interval regression. In: García-Pedrajas, N., Herrera, F., Fyfe, C., Benítez, J.M., Ali, M. (eds.) Trends in Applied Intelligent Systems, pp. 337–346. Springer, Heidelberg (2010)
24. Bhavsar, H., Ganatra, A.: A Comparative Study of Training Algorithms for Supervised Machine Learning
25. "Supervised Machine Learning: A Review of Classification …," 11:38:43 UTC
26. Segrera Francia, S., Moreno García, M.N.: Multiclasificadores: métodos y arquitecturas, March 2006. http://gredos.usal.es/jspui/handle/10366/21727. Accessed 29 Dec2015
27. Chawla, N.V.: Data mining for imbalanced datasets: an overview. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, pp. 853–867. Springer, New york (2005)
28. He, H., Ma, Y.: Foundations of imbalanced learning. In: Imbalanced Learning: Foundations, Algorithms, and Applications, p. 216. Wiley-IEEE Press (2013)
29. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. **21**(9), 1263–1284 (2009)
30. Thanathamathee, P., Lursinsap, C.: Handling imbalanced data sets with synthetic boundary data generation using bootstrap re-sampling and AdaBoost techniques. Pattern Recognit. Lett. **34**(12), 1339–1347 (2013)
31. Wong, G.Y., Leung, F.H.F., Ling, S.H.: A novel evolutionary preprocessing method based on over-sampling and under-sampling for imbalanced datasets. In: 39th Annual Conference of the IEEE Industrial Electronics Society, IECON 2013, pp. 2354–2359 (2013)
32. He, G., Han, H., Wang, W.: An over-sampling expert system for learning from imbalanced data sets. In: 2005 International Conference on Neural Networks and Brain, ICNN B 2005, vol. 1, pp. 537–541 (2005)
33. Pengfei, J., Chunkai, Z., Zhenyu, H.: A new sampling approach for classification of imbalanced data sets with high density. In: 2014 International Conference on Big Data and Smart Computing (BIGCOMP), pp. 217–222 (2014)
34. Mahmoudi, S., Moradi, P., Akhlaghian, F., Moradi, R.: Diversity and separable metrics in over-sampling technique for imbalanced data classification. In: 2014 4th International eConference on Computer and Knowledge Engineering (ICCKE), pp. 152–158 (2014)
35. Abdi, L., Hashemi, S.: To combat multi-class imbalanced problems by means of over-sampling techniques. IEEE Trans. Knowl. Data Eng. **28**(1), 238–251 (2016)
36. Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **42**(4), 463–484 (2012)
37. Zhang, H., Li, M.: RWO-Sampling: a random walk over-sampling approach to imbalanced data classification. Inf. Fusion **20**, 99–116 (2014)
38. Han, H., Wang, W.-Y., Mao, B.-H.: Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) Advances in Intelligent Computing, pp. 878–887. Springer, Heidelberg (2005)
39. Kerdprasop, N., Kerdprasop, K.: Predicting rare classes of primary tumors with over-sampling techniques. In: Kim, T., Adeli, H., Cuzzocrea, A., Arslan, T., Zhang, Y., Ma, J., Chung, K., Mariyam, S., Canción, X. (eds.) Database Theory and Application, Bio-science and Bio-technology, pp. 151–160. Springer, Heidelberg (2011)
40. Malpica, J.A.: Splines interpolation in high resolution satellite imagery. In: Bebis, G., Boyle, R., Koracin, D., Parvin, B. (eds.) Advances in Visual Computing, pp. 562–570. Springer, Heidelberg (2005)

41. Hung, K.-W., Siu, W.-C.: Learning-based image interpolation via robust k-NN searching for coherent AR parameters estimation. J. Vis. Commun. Image Represent. **31**, 305–311 (2015)
42. Rui, L., Qiong, L.: Image sharpening algorithm based on a variety of interpolation methods. In: 2012 International Conference on Image Analysis and Signal Processing (IASP), pp. 1–4 (2012)
43. Bentbib, A.H., El Guide, M., Jbilou, K., Reichel, L.: A global Lanczos method for image restoration. J. Comput. Appl. Math.
44. Shi, Z., Yao, S., Li, B., Cao, Q.: A novel image interpolation technique based on fractal theory. In: 2008 International Conference on Computer Science and Information Technology, ICCSIT 2008, pp. 472–475 (2008)
45. Sun, Y., Kamel, M.S., Wang, Y.: Boosting for learning multiple classes with imbalanced class distribution. In: 2006 Sixth International Conference on Data Mining, ICDM 2006, pp. 592–602 (2006)
46. Chawla, N.V., Lazarevic, A., Hall, L.O., Bowyer, K.W.: SMOTEBoost: improving prediction of the minority class in boosting. In: Lavrač, N., Gamberger, D., Todorovski, L., Blockeel, H. (eds.) Knowledge Discovery in Databases. PKDD 2003, pp. 107–119. Springer, Heidelberg (2003)
47. Viktor, H.L., Guo, H.: Multiple classifier prediction improvements against imbalanced datasets through added synthetic examples. In: Fred, A., Caelli, T.M., Duin, R.P.W., Campilho, A.C., de Ridder, D. (eds.) Structural, Syntactic, and Statistical Pattern Recognition, pp. 974–982. Springer, Heidelberg (2004)
48. Guo, H., Viktor, H.L.: Boosting with data generation: improving the classification of hard to learn examples. In: Orchard, B., Yang, C., Ali, M. (eds.) Innovations in Applied Artificial Intelligence, pp. 1082–1091. Springer, Heidelberg (2004)
49. Drechsler, J., Reiter, J.P.: An empirical evaluation of easily implemented, nonparametric methods for generating synthetic datasets. Comput. Stat. Data Anal. **55**(12), 3232–3243 (2011)
50. Anderson, J.W., Kennedy, K.E., Ngo, L.B., Luckow, A., Apon, A.W.: Synthetic data generation for the internet of things. In: 2014 IEEE International Conference on Big Data (Big Data), pp. 171–176 (2014)
51. Albuquerque, G., Lowe, T., Magnor, M.: Synthetic generation of high-dimensional datasets. IEEE Trans. Vis. Comput. Graph. **17**(12), 2317–2324 (2011)
52. Verbiest, N., Ramentol, E., Cornelis, C., Herrera, F.: Improving SMOTE with fuzzy rough prototype selection to detect noise in imbalanced classification data. In: Advances in Artificial Intelligence, IBERAMIA 2012, pp. 169–178 (2012)
53. Törn, A.A.: Correlation coefficients of linear regression models of human decision making. Omega **8**(3), 393–394 (1980)
54. Field, A., Miles, J., Field, Z.: Discovering Statistics Using R (2012)
55. Seiffert, C., Khoshgoftaar, T.M., Hulse, J.V., Napolitano, A.: RUSBoost: a hybrid approach to alleviating class imbalance. IEEE Trans. Syst. Man Cybern. Part Syst. Hum. **40**(1), 185–197 (2010)
56. Albayrak, A.S.S.: Alleviating the Class Imbalance problem in Data Mining (2013)
57. SMOTE: Synthetic Minority Over-sampling Technique. https://www.cs.cmu.edu/afs/cs/project/jair/pub/volume16/chawla02a-html/chawla2002.html. Accessed 19 June 2017
58. Beretta, L., Santaniello, A.: Nearest neighbor imputation algorithms: a critical evaluation. BMC Med. Inform. Decis. Mak. **16**(Suppl), 3 (2016)
59. Guo, H., Viktor, H.L.: Learning from imbalanced data sets with boosting and data generation: the DataBoost-IM approach. SIGKDD Explor. Newsl. **6**(1), 30–39 (2004)

60. Mohanty, P.K., Reza, M., Kumar, P., Kumar, P.: Implementation of cubic spline interpolation on parallel skeleton using pipeline model on CPU-GPU cluster. In: 2016 IEEE 6th International Conference on Advanced Computing (IACC), pp. 747–751 (2016)
61. Phillips, G.M.: Univariate interpolation. In: Interpolation and Approximation by Polynomials, pp. 1–48. Springer, New York (2003)
62. Keogh, E., Chu, S., Hart, D., Pazzani, M.: An online algorithm for segmenting time series. In: Proceedings 2001 IEEE International Conference on Data Mining, pp. 289–296 (2001)
63. Hamed, Y., Shafie, A., Mustaffa, Z.B., Idris, N.R.B.: An application of K-Nearest Neighbor interpolation on calibrating corrosion measurements collected by two non-destructive techniques. In: 2015 IEEE 3rd International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), pp. 1–5 (2015)
64. Li, H., Wan, X., Liang, Y., Gao, S.: Dynamic time warping based on cubic spline interpolation for time series data mining. In: 2014 IEEE International Conference on Data Mining Workshop, pp. 19–26 (2014)
65. Multivariate - Interpolation - Approximation - Maths Reference with Worked Examples. http://www.codecogs.com/library/maths/approximation/interpolation/multivariate.php. Accessed 20 Feb 2017
66. Influence of DEM interpolation methods in Drainage Analysis. https://www.researchgate.net/publication/237116945_Influence_of_DEM_interpolation_methods_in_Drainage_Analysis. Accessed 20 Feb 2017
67. Yang, L., Liu, S., Tsoka, S., Papageorgiou, L.G.: A regression tree approach using mathematical programming. Expert Syst. Appl. **78**, 347–357 (2017)
68. Magnani, M.: Techniques for Dealing with Missing Data in Knowledge Discovery Tasks (2004)
69. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and Regression Trees. Taylor & Francis (1984)
70. Vapnik, V., Golowich, S.E., Smola, A.J.: Support vector method for function approximation, regression estimation and signal processing. In: Mozer, M.C., Jordan, M.I., Petsche, T. (eds.) Advances in Neural Information Processing Systems 9, pp. 281–287. MIT Press (1997)
71. Neural Networks: A Comprehensive Foundation (2nd edn.) Neural Networks: A Comprehensive Foundation. ResearchGate. https://www.researchgate.net/publication/233784957_Neural_Networks_A_Comprehensive_Foundation_2nd_Edition_Neural_Networks_A_Comprehensive_Foundation. Accessed 16 June 2017
72. Corrales, D.C., Gutierrez, G., Rodriguez, J.P., Ledezma, A., Corrales, J.C.: Lack of data: is it enough estimating the coffee rust with meteorological time series? In: Computational Science and Its Applications, ICCSA 2017, pp. 3–16 (2017)

Towards an Alert System for Coffee Diseases and Pests in a Smart Farming Approach Based on Semi-supervised Learning and Graph Similarity

Emmanuel Lasso(✉) and Juan Carlos Corrales

Grupo de Ingeniería Telemática, Universidad del Cauca,
Campus Tulcán, Popayán, Cauca, Colombia
{eglasso,jcorral}@unicauca.edu.co

Abstract. Smart Farming represents a new approach based on management of observation, measurement and response to internal and external variations in crops. This approach is closely related to a current trend area in Information and Communication Technologies such as Big Data. The application of machine learning techniques to agriculture data allows to assist in decision making and predict what will happen in the future (predictive analysis). From predictive models, the inexact graph matching would allow to establish the probability of occurrence of one or another disease or in such case the presence of a pest, based on the analysis of the crop conditions. This paper presents a review of some areas involved in the definition of an alert system for diseases and pests in a Smart Farming approach, based on machine learning and graph similarity. Finally, the integration of the mentioned areas for their application in coffee crops is proposed.

Keywords: Semi-supervised learning · Graph similarity · Pattern matching · Coffee disease · Pest

1 Introduction

In agricultural production systems, Smart Farming represents a new approach based on management of observation, measurement and response to internal and external variety in crops [1, 2]. In this sense, the tasks of administration, decision making and management of sudden events (diseases and pests) are improved by analyzing a large amount of data characterizing the environment around crops (climate, physical properties, agronomic management). For coffee crops, there are several diseases such as leaf rust, South America leaf spot, brown leaf spot; and also pests such as borer beetle, which greatly influence the quality and costs of their production. Some researchers [3–9] have focused their efforts on determining the relationships between climatic conditions and agronomic properties of crops, with the episodes of phenomena mentioned. In Colombia, there are some organizations that have developed *agroclimatic* platforms to monitor climatic

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_9

variability in different areas of the country, such as AgroCloud[1] platform belonging to RICCLISA[2]. However, there are only a few initiatives focused on analyzing intensively the monitored data in order to generate a timely response and contingency measures against diseases and pests that affect coffee trees, which in turn generate large losses and decrease quality in crops.

Smart Farming is closely related to a current trend area in Information and Communication Technologies such as Big Data. Moreover, the application of Big Data in Agriculture would have no real value without the data analysis (Big Data Analytics) [10], where the value and variability of the data are exploited through machine learning techniques and data mining. This task represents a tool for the extraction of knowledge, generation of models to determine factors that limit yield and quality in crops, and improve the response to sudden phenomena such as epidemics of diseases and pests [11]. Such models can be represented as rule sets and decision trees, which provide an easy-to-understand structure for the user and generate a prediction from the definition of ranges in the values of each variable involved. For example, the conditions that favor a disease or pest in the crop can be represented as graph patterns. Graph-based information structures consist of a set of nodes (for the specific case of coffee rust, a node can represent the data monitoring instances) that are related through edges (again, in the case of coffee rust, an edge can represent the semantic context between measured variables in the crops), which have a generally dynamic nature with capacities for the manipulation of large amounts of interconnected data, coming from heterogeneous sources [12]. A graph-based representation can be taken as information structure for the storage and analysis of the different types of variables present in a crop environment.

After defining the diseases or pests in terms of graph patterns, they are stored in a repository. These patterns can be searched in a data graph that represents weather conditions obtained from monitoring services and agronomic properties obtained from users, and as a result it will indicate whether the conditions for the disease or pest are present. In computer science, this task is called *Graph Pattern Matching* [13], and is based on the search for subgraphs within a data graph that meet the characteristics described in a pattern, which in this case are favorable conditions for one of the problems mentioned. The matching can be exact or inexact. In the case of exact matching, the conditions in the patterns must be fulfilled in their totality, without giving margin to near approximations to a certain pattern. The authors of [14, 15] have developed a prototype that allows the detection of favorable conditions for three rates of coffee rust infection in Colombian coffee crops through the application of exact graph pattern matching, leaving aside the possibility of finding conditions conducive to other diseases or pests that may have similar characteristics in the patterns that characterize them. On the other hand, in the case of inexact matching, the closeness of a pattern to a subgraph is quantified through measures such as: graph edit distance [16], maximum common subgraph [17] and similarity score [18]. In this sense, from the multiple data obtained in an environment of intelligent agriculture, the inexact graph matching would allow to establish the probability of occurrence of one or another disease or in such case the presence of a pest,

[1] https://agrocloudcolombia.com.

[2] http://ricclisa.org/.

based on the analysis of the crop conditions (the similarity with the graph patterns extracted for each disease or pest, can determine how similar are the conditions that lead to the occurrence of each of them).

Thus, in a scenario of Smart Farming, where coffee production organizations in Colombia lack a technological system to identify favorable conditions for the occurrence of diseases and pests in coffee, we propose an approach to provide elements necessary for an early response to events, such as those mentioned above, that regularly generate negative impacts on the quality and costs of coffee production. The remainder of this paper is organized as follows: Sect. 2 describes the state of the art; Sect. 3 presents the contributions and gaps of related works in order to present our proposal and Sect. 4 address the conclusions.

2 State of the Art

There are several knowledge areas involved in the development of tools that contribute to the improvement of agricultural production systems. In addition, some research aims to develop intelligent systems to counteract the effects of diseases that attack crops. Various technological advances offer algorithms and techniques to achieve this. In order to understand the context in which this work is developed, the topics and related works of different areas of knowledge, such as Big Data Analytics, Graphs and Graphs Similarity, are discussed below.

2.1 General Context

2.1.1 Big Data Analytics

Big Data concept is related to the generation, manipulation and analysis of large amounts of data from heterogeneous sources [19]. The Big Data value chain is composed of: generation, acquisition, storage and analysis of data. Precisely the latter, known as Big Data Analytics, aims to interpret the data to assist in decision making and predict what will happen in the future (predictive analysis) [20]. This can be achieved through machine learning techniques, which are characterized by learning from a dataset [21], used to analyze the relationships between the variables present in it and finally generate a prediction model. In a Big Data environment, there are large datasets from different sources that represent the variables that affect a particular problem, and generally the largest portion of their instances are not labeled (they do not contain a relationship of the data with a target variable). Semi-Supervised Learning (SSL) is based on the analysis of both labeled (related to a target variable) and unlabeled data, in order to generate a classification function with the maximum possible generalization [22, 23]. This function can predict the label of a new input data (inductive learning) or infer the label for the unlabeled instances in the dataset (transductive learning) [24]. This approach assumes that the information available in the unlabeled instances can influence the classification function and generate more accurate classifiers, since these contain more examples of the analyzed domain. In this way, the greatest possible amount of information collected in a specific environment is used.

The SSL can consider a graph-based representation of the dataset instances and label propagation mechanisms between similar nodes for the label inference. Graphs are used since they are a uniform representation for homogeneous data, easily scalable for handling large amounts of information and the Graph-based SSL has proven to be more effective in practice compared to other techniques of automatic learning [24].

2.1.2 Graphs

A graph is a set of nodes that represent the entities within an application domain and edges that constitute links or relations between these entities [25], allowing to characterize the distribution of large sources of information, the strategic positions of its elements, and the dynamics within a knowledge base [12]. The nodes and edges can be characterized from labels containing information of the modeled domain and indicators of their position in the formed network. The new generation of database systems, which generally work with structured documents, often model information through trees and graphs. Semantics that can be implicit in nodes and edges labels allow you to enrich and generate additional information from the stored data [26]. A graph that considers the semantics in its structure is conformed by classes of an ontology (nodes), and the relations between them (connections) [27].

2.1.3 Graph Similarity

A Graph Pattern Matching is defined as: "*Given a Data Graph G, and a graph pattern Q, find all the matches in G for Q*" [28]. This is usually aimed at finding entities that have specific characteristics in their attributes and relations with other nodes of the graph and, therefore, a characterization of the similarity between two graphs. In this sense, the searched pattern can be seen as a series of conditions within the attributes of the graph, similar to the evaluation made by decision trees. In addition, systems based on this technique are divided into exact pattern matching and error tolerant pattern matching systems. While exact match offers a rigorous way to get a result in mathematical terms, it is usually only applied to a limited set of real-world problems. Instead, the error-tolerant system is able to cope with the distortion in the data, generating a measure of similarity or closeness, which is a frequent element in real-world problems. For many applications of these techniques, it is of great interest to know the degree of similarity that exists between two graphs, instead of knowing if there is a maximum coincidence [29]. In this approach, it is possible to quantify the closeness of a pattern to a subgraph through measures such as: graph edit distance [16], maximum common subgraph [17] and similarity score [18]. The measures of similarity between graphs take into account the differences between node types, label values and structure of the compared graphs [30]. In this way, there are two types of similarities: structural and based on the elements properties. In the case that the graphs contain semantic information, the different processes and similarity metrics represent a degree of semantic approximation [30].

2.2 Related Work

Next, we present the research related to the proposal under development, considering the automatic detection of diseases in coffee, the application of concepts and techniques of Big Data to Smart Farming, in addition to the graph pattern matching, such as areas of interest for the development of this project.

2.2.1 Automatic Disease Detection in Coffee Crops

In computer science, there are several approaches that address the use of machine learning techniques to solve problems in agricultural domains. Next, the most relevant works, specifically applied to coffee production, are presented.

First, the research paper presented in [31] proposes the use of fuzzy decision trees in order to generate alerts for the appearance of coffee rust. The models obtained represent thresholds of different variables, both for situations of disease prevention and treatment. In addition, the process is carried out by analyzing a dataset of approximately 8 years of disease records. For its evaluation, the tool is compared with the traditional decision trees, obtaining better values of performance. In the same approach, the authors of [32, 33] make use of 364 samples containing information on temperature, precipitation and relative humidity, in order to training a decision tree induction algorithm, proposed by Han and Kamber [34]. The model provides support for understanding how the interaction between the variables analyzed leads to rust epidemics. After its execution, the model correctly classifies 78% of the training dataset, and its accuracy is estimated at 73% for the classification of new samples.

In the research carried out by Corrales et al. [35, 36] this problem is approached through the use of multiclassifiers and assembly methods, in order to reduce errors in classification models. The method is based on two-level classifiers, and these classifiers are chosen from the performance measures comparison of algorithms such as Support Vector Machines, Artificial Neural Networks, Bayesian Networks, Decision Trees, among others. The tests performed show that this approach presents better values of correlation coefficient, mean absolute error and quadratic error.

Finally, Lasso et al. [14] propose the generation of a graph-based representation of coffee rust growth patterns, modeled according to the variables related to this disease and based on rules extracted from the induction of decision trees and expert knowledge. The patterns obtained provide a greater expressiveness and interpretation of the climatic phenomena that favor the development of the disease. The same author proposes the construction of an expert system that makes use of graph pattern matching to validate the rules and the knowledge produced by experts, in order to find the crops that present favorable conditions for a rust epidemic [15].

A common problem in the previous related work is the number of examples (instances) used to train the algorithms, since the disease records were not sufficient to obtain an ideal training set for supervised learning tasks and discarded the information contained in the unlabeled data.

2.2.2 Smart Farming and Big Data

The emergence of new technologies for the monitoring of a wide variety of crop conditions and properties has allowed a transition from precision agriculture to intelligent agriculture, where the great amount of information obtained is harnessed from its analysis and interpretation. Following are some research that addresses the integration of Big Data concepts and Smart Farming.

Wolfert et al. [37] present a review of the research carried out around the application of Big Data in Smart Farming, mentioning that it is in an early stage of development. In addition, similar documents are compiled that present revisions of these two topics and analyzed how they could be articulated. In this way, some opportunities for the integration of the two concepts are identified, such as: predictive modeling, improvement of risk management models in crops, improvements in decision-making processes, among others [37, 38]. To achieve these purposes, the variety and value of the data can be exploited from the Big Data Analysis. This approach is followed by Shah et al. [39], which proposes an architecture for Big Data analysis in agricultural advice systems. This architecture considers the technologies suggested for each step in the Big Data value chain, emphasizing the analysis stage, implemented for crop yield prediction. To achieve this, a machine learning technique is applied. This prediction is used to decision-making support for farmers. The authors of [40] make use of Big Data Analytics to find relationships between environmental factors and crop growth cycles. These relationships are used to improve decision-making, agricultural production and research processes. In addition, in [2] an exploration of two supervised learning techniques is carried out: Support Vectors Machines and Artificial Neural Networks; in order to be applied in a Big Data Analysis task. The research is focused on mixed systems (agriculture and livestock). As a result, a model for bovine intake prediction is obtained, in order to control its impact on nearby crops.

In [41, 42] a conceptual framework based on Big Data Analytics for the identification of diseases in crops is presented, taking as case study the rice blight. This framework does not make use of machine learning techniques, but the analysis is based on the similarity between dataset instances. The presented tool makes recommendations for the solution to the disease, based on the similarity between the symptoms of a plant at a given time and the records of symptoms presented in past episodes of the disease.

Finally, an approach in the use of semi-supervised learning is presented in [43], for detecting beetles pests in crops. The dataset used contains labeled and unlabeled instances and corresponds to historical climate records, crop growth characteristics, pest growth, among others. In this dataset two key elements of Big Data can be identified: the data variety and value. An algorithm to process the labeled segment based in association rules is used and the unlabeled segment is analyzed through ISODATA (Iterative Self-Organizing Data Analysis Technique Algorithm) [44]. In this way, it is assumed that in a densely distributed region of the data, the models should obtain similar outputs. The predictive results of the semi-supervised approach are more accurate than supervised learning approaches such as Support Vector Machines, Neural Networks and K-means Clustering.

2.2.3 Graph Similarity

The evaluation of graphs similarity is known as a matching process between graphs. Below, different research papers that propose techniques and optimizations for the similarity calculation are presented.

Fan research [45, 46] make use of graph pattern matching to evaluate the similarity between a pattern and a data graph. Their proposals are oriented towards the elimination of the number of excessive results before a query and the generation of a score for the matches found. In this way, an ordered list of candidate subgraphs is generated. In addition, other characteristics of the proposed algorithms are: efficiency in the searches evaluations in large graphs; Support to graph data and structure update; and graph compression. The support for graph updates presents a better performance to the methods proposed by similar approximations. In turn, the graph compression generates a reduction of the original size by 57%, which also reduces the matching runtime by 70%. In a similar way, in [47] a software called GMT (Graph Matching Toolkit) is presented, which graphically constructs graph patterns for its search. GMT uses a pattern matching algorithm [48] that generates a binary search tree to analyze the graph nodes and the graphs distance concept proposed by Riesen [49]. The results show that the similarity calculation process is optimized in computational resource use.

A common point in several investigations is the A-star algorithm [50], which has been the basis for the construction of new techniques, such as the proposals presented in [51–53]. This algorithm has a great precision and aims to find the edit distance between two graphs, performing a classic tree search based on exploring the space of all possible mapping of nodes and edges of graphs. The above mentioned proposals are focused on improving the execution time of A-star, which is exponential with respect to the graph size, and its high consumption of computational resources. In [54, 55] some techniques for the correct relationship of the structural similarities with the elements properties (labels or attributes) similarities between two graphs are proposed, which allows to obtain a resulting similarity of greater relevance. Consequently, the similarity calculation process between graphs must be adapted to the application domain [30].

The most recent approaches in this area propose improvements aimed at parallel processing and easily scalable operations for graph similarity calculation [56, 57], as this is a common requirement given the large amount of data currently being produced In different application domains.

3 Discussion

From the previous review of the related work in each identified area, it is possible to define their most important contributions and gaps for our proposal:

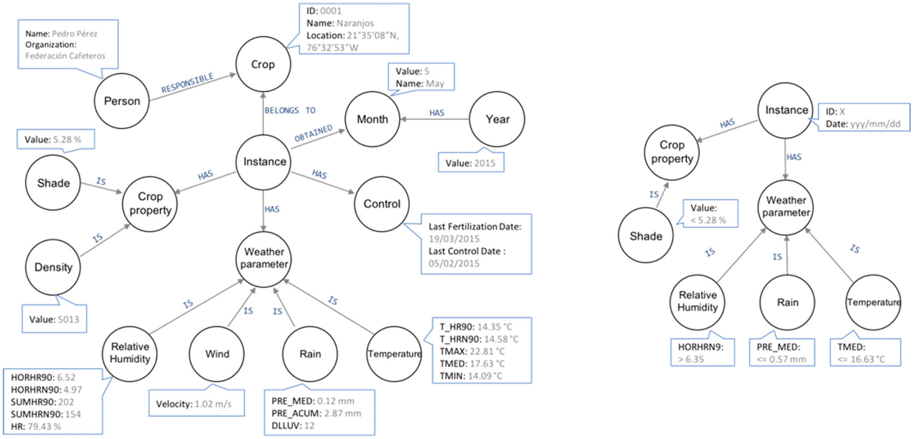
Automatic disease detection in coffee crops [14, 15, 31–36, 58]: These studies demonstrate that the use of CRISP-DM (Cross Industry Standard Process for Data Mining) [59] as a methodology for the discovery of knowledge from machine learning techniques, applied to crop monitoring data, generate valuable solutions around diseases in crops. In this scenario, the additional use of unlabeled data in the application domain, which are in greater proportion than label data, would theoretically have a positive impact on the classification task. This advantage would be possible from semi-supervised learning (SSL).

Smart Farming and Big Data [2, 37–43]: The use of machine learning techniques to carry out the Big Data Analysis stage is presented. In Smart Farming a large amount of information is produced, being necessary to take advantage of both labeled and unlabeled data.

Graph Similarity [30, 45–57]: Related work is focused on improving well-known techniques for graph pattern matching. The most flexible method for the similarity calculation is the error-tolerant pattern matching based on the graph-editing distance. The similarity functions considered and proposed only consider graphs labeled with numerical values or text strings. However, the labels in nodes and edges that express numerical ranges are not considered, such as the coffee rust graph patterns proposed in [14]. Similarly, other coffee diseases and pests may have similar patterns that determine the conditions for their occurrence.

In computer sciences, several techniques allow to take advantage of the data value and variety obtained in crops monitoring around a Big Data environment, in order to obtain predictive models to solve several problems of agricultural production. However, these models are often "black box" elements, where their response is known, but the process to get to it is unknown. Precisely, there are models for which their structure is observable and understandable to anyone, known as "white box", such as rule-based and decision trees algorithms. In some application domains, these models are often less precise but more understandable [60]. Given the existence of experts in coffee diseases and pests, they can contribute with their knowledge in the patterns optimization, obtained from "white box" models, since its structure expresses the way in which a prediction is made. In addition, the above presents an advantage to obtain a representation of the favorable environment conditions of crops for diseases and pests, as proposed in [14].

From the above considerations, our proposal aims to identify favorable conditions for coffee diseases and pest based on Smart Farming, Semi-Supervised Learning and Similarity between Graphs. As an example, in Fig. 1 the structure that could have the data graph and patterns is shown. The nodes represent entities of the application domain, such as Crop, Instance, some weather parameters, among others. The edges of the graph represent dependencies and relationships between entities. On the other hand, nodes and edges contain labels, which characterize their properties from the value of significant variables for each. The data graph represents the characteristics of a crop, the Instance entity being in charge of representing the conditions (weather and crop properties) that the crop has at a given time. Figure 1(b) shows a graph pattern which may contain the ranges of the crop conditions associated with a disease or pest. The matching of the pattern and the data graph may result in the degree of similarity between the present conditions and the patterns that determine the occurrence of one of the problems mentioned. The structural difference allows to verify the coherence in the relationships that exist between the entities of the application domain whereas the comparison between tags of the elements define the closeness between the values of the data graph and the ranges expressed in each pattern. In this sense, each task of the pairing (structural difference, semantic difference, difference between labels) must be assigned a weight, in order to adapt the results to the application domain.



a) Data graph example

b) Graph Pattern example

Fig. 1. Data graph and pattern examples

The architecture proposed is shown in Fig. 2, and the components are described below:

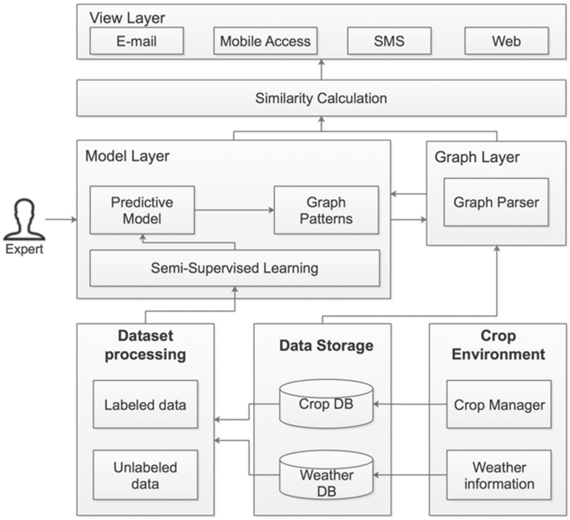


Fig. 2. Architecture proposed for coffee disease and pest detection

- **Crop environment:** This component represents the two information sources: crop properties and weather information, obtained in a Smart Farming environment.

- **Storage data:** Persistence and access to the information obtained from coffee crops environment, making use of Database Manager System.
- **Dataset processing:** Identification of labeled and unlabeled elements in the crops and weather Databases, in order to generate a training dataset.
- **Model layer:** This component contains the elements related to the model generation for the identification that favorable conditions for diseases and pests. The training dataset is used by a Semi-Supervised Learning algorithm to produce a predictive model. From this model, a set of graph patterns are generated through the Graph Parser. Additionally, the patterns can be reviewed for any expert to validate and/or improve it.
- **Graph layer:** A graph parser that generates a graph-based representation of predictive models and data from crops monitoring and management.
- **Similarity Calculation:** Composed by an Inexact Graph Pattern Matching algorithm, in order to find the similarity scores between graph patterns and data graph. The similarity obtained is used to characterize the conditions of any crop related to a disease or pest.
- **View layer:** This layer shows the results of the similarity calculation through several interfaces (e-mail, mobile access, SMS, Web).

4 Conclusions

We presented the most relevant research around automatic coffee diseases detection, smart farming and graph similarity. Our proposal is based on an integration between the areas of knowledge addressed. Smart Farming supported in Big Data approach is still in an early-stage development, but with great potential for development. The information that can be obtained in this approach (crops monitoring) can be exploited through the application of a semi-supervised learning process following a methodology for the discovery of knowledge in databases, such as CRISP-DM, in order to obtain predictive models. Thus, from these models a set of rules that characterize diseases and pests of coffee can be extracted. These rules can be represented as graph patterns and graph similarity techniques would allow to determine the probability of a disease or pest occurrence, based on the analysis of the conditions (climatic and agronomic) present in the area where the crop is located (the similarity between the graph patterns extracted for each disease or pest, can determine how similar are the conditions that lead to the occurrence of each one of them). As future work we consider the implementation of the different components of the proposed system and its validation within the Colombian coffee production.

Acknowledgements. The authors are grateful to the University of Cauca and its Telematics Engineering Group (GIT), the Colombian Administrative Department of Science, Technology and Innovation (Colciencias), AgroCloud project of The Interinstitutional Network of Climate Change and Food Security of Colombia (RICCLISA) for supporting this research and the InnovAccion Cauca project of the Colombian Science, Technology and Innovation Fund (SGR-CTeI) for PhD scholarship granted to MSc. Emmanuel Lasso.

References

1. Kempenaar, C., Kocks, C.G.: Van precisielandbouw naar smart farming technology. Kenniscentrum Agrofood en Ondernemen (2013)
2. Kempenaar, C., et al.: Big Data Analysis for Smart Farming. Wageningen University & Research (2016)
3. Rivillas, C., Serna, C., Cristancho, M., Gaitán, A.: Roya del Cafeto en Colombia: Impacto, Manejo y Costos del Control. Chinchiná Bol. Téc., **36** (2011)
4. Waller, J.M., Bigger, M., Hillocks, R.J.: Coffee Pests, Diseases and their Management. CABI (2007)
5. Muller, R.A., Berry, D., Avelino, J., Bieysse, D.: Coffee diseases. Coffee Grow. Process. Sustain. Prod. Guideb. Grow. Process. Traders Res., 491–545 (2004)
6. Machado, P., Gil, Z., Constantino, L.M., Villegas, C., Giraldo, M.: Plagas del café. Man. Cafe. Colomb. Investig. Tecnol. Para Sostenibilidad Caficultura, 215–306 (2013)
7. Avelino, J., et al.: The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. Food Secur. **7**(2), 303–321 (2015)
8. Kushalappa, A.C., Eskes, A.B.: Advances in coffee rust research. Annu. Rev. Phytopathol. **27**(1), 503–531 (1989)
9. Gaitán, Á., Rivillas, C., Castro, B., Cristancho, M.: Manejo integrado de enfermedades. Man. Cafe. Colomb. Investig. Tecnol. Para Sostenibilidad Caficultura, 143–178 (2013)
10. Sun, Z.F., Du, K.M., Zheng, F.X., Yin, S.Y., et al.: Perspectives of research and application of big data on smart agriculture. J. Agric. Sci. Technol. Beijing **15**(6), 63–71 (2013)
11. Kshetri, N.: Big Data's Big Potential in Developing Economies: Impact on Agriculture, Health and Environmental Security. CABI (2016)
12. Erétéo, G., Buffa, M., Gandon, F., Corby, O.: Analysis of a real online social network using semantic web frameworks. Springer (2009)
13. Fan, W., Li, J., Ma, S., Tang, N., Wu, Y., Wu, Y.: Graph pattern matching: from intractable to polynomial time. Proc. VLDB Endow. **3**(1–2), 264–275 (2010)
14. Lasso, E., Thamada, T.T., Meira, C.A.A., Corrales, J.C.: Graph patterns as representation of rules extracted from decision trees for coffee rust detection. In: Garoufallou, E., Hartley, R.J., Gaitanou, P. (eds.) Metadata and Semantics Research, pp. 405–414. Springer International Publishing (2015)
15. Lasso, E., Corrales, J.C.: Expert system for crop disease based on graph pattern matching: a proposal. Rev. Ing. Univ. Medellín **15**(29), 81–98 (2016)
16. Gao, X., Xiao, B., Tao, D., Li, X.: A survey of graph edit distance. Pattern Anal. Appl. **13**(1), 113–129 (2010)
17. Bunke, H.: On a relation between graph edit distance and maximum common subgraph. Pattern Recognit. Lett. **18**(8), 689–694 (1997)
18. Zager, L.A., Verghese, G.C.: Graph similarity scoring and matching. Appl. Math. Lett. **21**(1), 86–94 (2008)
19. Gantz, J., Reinsel, D.: Extracting value from chaos. IDC Iview **1142**(2011), 1–12 (2011)
20. Hu, H., Wen, Y., Chua, T.-S., Li, X.: Toward scalable systems for big data analytics: a technology tutorial. IEEE Access **2**, 652–687 (2014)
21. Ayodele, T.O.: Types of machine learning algorithms. In: Zhang, Y. (ed.) New Advances in Machine Learning, pp. 20–48. In-Tech, India (2010). intechweb.org
22. Zhu, X.: Semi-supervised Learning. In: Seel, N.M. (ed.) Encyclopedia of Machine Learning, pp. 892–897. Springer (2011)
23. Chapelle, O., Schlkopf, B., Zien, A.: Semi-supervised Learning, 1st edn. The MIT Press, Cambridge (2010)

24. Subramanya, A., Talukdar, P.P.: Graph-based semi-supervised learning. Synth. Lect. Artif. Intell. Mach. Learn. **8**(4), 1–125 (2014)
25. Hanneman, R.A., Riddle, M.: Introduction to social network methods. University of California Riverside (2005)
26. Lasso-Sambony, E.G., Ortega-Ponce, S.M., Corrales, J.C.: Semantic enrichment and inference of relationships in an online social network. Ing. Univ. **17**(2), 355–373 (2013)
27. Rowe, M.: Applying semantic social graphs to disambiguate identity references. In: The Semantic Web: Research and Applications, pp. 461–475. Springer (2009)
28. Wang, X.: Graph Pattern Matching on Social Network Analysis. University of Edinburgh (2013)
29. Bunke, H., Neuhaus, M.: Graph matching. exact and error-tolerant methods and the automatic learning of edit costs. Min. Graph Data, 17–32 (2007)
30. Gallagher, B.: Matching structure and semantics: a survey on graph-based pattern matching. AAAI FS **6**, 45–53 (2006)
31. Cintra, M.E., Meira, C.A.A., Monard, M.C., Camargo, H.A., Rodrigues, L.H.A.: The use of fuzzy decision trees for coffee rust warning in Brazilian crops. In: 2011 11th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 1347–1352 (2011)
32. Meira, C.A., Rodrigues, L.H., Moraes, S.A.: Análise da epidemia da ferrugem do cafeeiro com árvore de decisão. Trop. Plant Pathol. **33**(2), 114–124 (2008)
33. Meira, C.A.A., Rodrigues, L.H.A.: ÁRVORE DE DECISÃO NA ANÁLISE DE EPIDEMIAS DA FERRUGEM DO CAFEEIRO (2009)
34. Jiawei, H., Kamber, M.: Data mining: concepts and techniques, vol. 5. Morgan Kaufmann, San Franc. CA (2001)
35. Corrales, D.C., Figueroa, A., Ledezma, A., Corrales, J.C.: An empirical multi-classifier for coffee rust detection in colombian crops. In: 15th International Conference on Computational Science and Its Applications – ICCSA 2015, Banff, AB, Canada, 22–25, June 2015, Proceedings, Part I, vol. 9155, pp. 60–74 (2015)
36. Corrales, D.C., Casas, A.F., Ledezma, A., Corrales, J.C.: Two-level classifier ensembles for coffee rust estimation in Colombian Crops. Int. J. Agric. Environ. Inf. Syst. IJAEIS **7**(3), 41–59 (2016)
37. Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M.-J.: Big data in smart farming–a review. Agric. Syst. **153**, 69–80 (2017)
38. Lesser, A.: Big data and big agriculture, 08 October 2014. [En línea]. Disponible en: https://gigaom.com/report/big-data-and-big-agriculture/. [Consultado: 19 May 2017]
39. Shah, P., Hiremath, D., Chaudhary, S.: Big data analytics architecture for agro advisory system. In: High Performance Computing Workshops (HiPCW), pp. 43–49 (2016)
40. Hao, F., Luo, X., Mu, C.: Research on key technologies of intelligent agriculture based on agricultural big data. In: International Conference on Smart City and Systems Engineering (ICSCSE), pp. 598–601 (2016)
41. Kaur, R., Garg, R., Aggarwal, H.: Big data analytics framework to identify crop disease and recommendation a solution. In: International Conference on Inventive Computation Technologies (ICICT), vol. 2, pp. 1–5 (2016)
42. Garg, R., Aggarwal, H.: Big data analytics recommendation solutions for crop disease using Hive and Hadoop Platform. Indian J. Sci. Technol., **9**(32) (2016)
43. Li, T., Yang, J., Peng, X., Chen, Z., Luo, C.: Prediction and early warning method for flea beetle based on semi-supervised learning algorithm. In: Proceedings of the 2008 Fourth International Conference on Natural Computation, Washington, DC, USA, vol. 04, pp. 217–221 (2008)

44. Ball, G.H., Hall, D.J.: ISODATA, a novel method of data analysis and pattern classification. Stanford Research Inst Menlo Park CA (1965)
45. Fan, W., Wang, X., Wu, Y.: Diversified top-k graph pattern matching. Proc. VLDB Endow. **6**(13), 1510–1521 (2013)
46. Ma, S., Cao, Y., Fan, W., Huai, J., Wo, T.: Capturing topology in graph pattern matching. Proc. VLDB Endow. **5**(4), 310–321 (2011)
47. Ogaard, K., Roy, H., Kase, S., Nagi, R., Sambhoos, K., Sudit, M.: Discovering patterns in social networks with graph matching algorithms. In: Social Computing, Behavioral-Cultural Modeling and Prediction, pp. 341–349. Springer, (2013)
48. Sambhoos, K.P.: Graph matching applications in high level information fusion. In: ProQuest (2007)
49. Riesen, K., Emmenegger, S., Bunke, H.: A novel software toolkit for graph edit distance computation. In: International Workshop on Graph-Based Representations in Pattern Recognition, pp. 142–151 (2013)
50. Messmer, B.T., Bunke, H.: A decision tree approach to graph and subgraph isomorphism detection. Pattern Recognit. **32**(12), 1979–1998 (1999)
51. Riesen, K., Neuhaus, M., Bunke, H.: Bipartite graph matching for computing the edit distance of graphs. In: International Workshop on Graph-Based Representations in Pattern Recognition, pp. 1–12 (2007)
52. Fankhauser, S., Riesen, K., Bunke, H.: Speeding up graph edit distance computation through fast bipartite matching. In: International Workshop on Graph-Based Representations in Pattern Recognition, pp. 102–111 (2011)
53. Zaslavskiy, M., Bach, F., Vert, J.-P.: Many-to-many graph matching: a continuous relaxation approach. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 515–530 (2010)
54. Wolverton, M., et al.: LAW: a workbench for approximate pattern matching in relational data. In: IAAI, 2003, vol. 3, pp. 143–150 (2003)
55. Berry, P.M., Harrison, I., Lowrance, J.D., Rodriguez, A.C., Ruspini, E.H.: Link analysis workbench. DTIC Document (2004)
56. Kollias, G., Sathe, M., Schenk, O., Grama, A.: Fast parallel algorithms for graph similarity and matching. J. Parallel Distrib. Comput. **74**(5), 2400–2410 (2014)
57. Zhao, X., Xiao, C., Zhang, W., Lin, X., Tang, J.: Improving performance of graph similarity joins using selected substructures. In: International Conference on Database Systems for Advanced Applications, pp. 156–172 (2014)
58. Meira, C.A.A., Rodrigues, L.H.A., Moraes, S.A.: Analysis of coffee leaf rust epidemics with decision tree. Trop. Plant Pathol. **33**(2), 114–124 (2008)
59. Chapman, P., et al.: CRISP-DM 1.0 Step-by-step data mining guide (2000)
60. Romero, C., Olmo, J.L., Ventura, S.: A meta-learning approach for recommending a subset of white-box classification algorithms for Moodle datasets. In: Educational Data Mining 2013 (2013)

Deploying Timely Alerts Through Converged Services: An Application for Colombian Agriculture

Julián Eduardo Plazas(✉) and Juan Carlos Corrales

Grupo de Ingeniería Telemática (GIT), Universidad del Cauca, Popayán, Colombia
{jeplazas,jcorral}@unicauca.edu.co

Abstract. Crop diseases and pests are one of the major concerns of Colombian farmers, since the affected crops could highly reduce their yields, negatively impacting the livelihood of their families. Though recent researches have focused on opportunely detecting possible infections, most of the alerts are deployed through ineffective legacy systems wasting the opportunities brought by these researches. Thus, we propose an ICT-based easily-scalable dissemination system for coffee-rust early warning systems, capable of deploying alerts through both telecommunications and Web 2.0 services, and seamlessly connecting to current rust-detection systems through a standard Web interface. This system will allow Colombian agriculture-oriented alert systems to increase their efficiency and effectiveness, reducing the impact of crops pests and diseases on the Colombian economy, mainly for smallholders.

Keywords: Converged services · Early warning systems · Dissemination & Communication · Coffee rust

1 Introduction

Agriculture is a major economic activity for almost every country in the world, including Colombia, where the agricultural sector represents more than 10% of the National Domestic Product and the livelihood of almost 4 million people [1]. Furthermore, coffee production is the main agricultural activity in Colombia; more than 350,000 families depend on harvesting coffee as their sole incomes. Therefore, economic factors alongside diseases, pests and weather changes could significantly reduce these Colombian families' quality of life [2].

The most important and severe disease currently affecting coffee production in Colombia is the coffee rust. This disease has caused production losses up to 30% in susceptible varieties of *coffea arabica* in Colombia, drastically reducing the livelihood of thousands of smallholders and harvesters [3].

Due to the severity of the coffee rust, both public and private entities have adopted different strategies for overcoming this disease and avoiding new epidemics. Half of the rust-susceptible Colombian coffee crops have been replaced with new resistant varieties. Moreover, surveillance systems have been implemented in order to monitor the coffee

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_10

rust incidence in commercial fields, allowing for the early detection of coffee rust outbreaks [3].

Notwithstanding these measures, about 40% of the Colombian coffee crops are still susceptible to the coffee rust [3] since variety is crucial for satisfying all the markets' requests. This scenario highlights the importance of crop-monitoring systems in their task of preventing future coffee rust epidemics.

Therefore, many researchers have focused their efforts on improving the performance of rust-detection systems: proposing new methods and techniques for automatically detecting diseases in the crops [4, 5]; collecting and preparing data for developing future behavior models [2]; defining novel sets of rules and patterns for detecting the coffee rust [6]; developing multi-classifier systems capable of estimating and detecting the coffee rust incidence in selected crops [7, 8]; and even validating the estimations against abnormalities in order to avoid false positives [9]. These researches are focused on improving current monitoring and alert systems by increasing their reliability and reducing their detection time. Nevertheless, their actual impact on preventing epidemics might be inhibited by such systems, which base their alerts on quarterly bulletins [3], wasting the opportunity these new detection methods bring.

Aiming to overcome the limitations of current rust-alert systems, some researchers have focused on leveraging Information and Communication Technologies (ICT) in order to deliver timely alerts, exploiting the opportunity brought by the new detection methods. Some works have completely designed new early warning systems [10], while others have developed dissemination components intended to enhance the timeliness of current systems, considering the lack of advanced telecommunications networks in Colombian rural areas [11]. However, though being interesting approaches, the implementation of these proposals as actual dissemination systems is too difficult to achieve in the necessary scale or is merely considered as a shallow idea [10, 11].

Hence, we propose an ICT-based easily-scalable dissemination component for agriculture-oriented early warning systems, capable of deploying alerts through both telecommunications and Web 2.0 services (converged services), and seamlessly connecting to current rust-detection systems through a standard Web interface. This component will allow Colombian agriculture-oriented alert systems to increase their efficiency and effectiveness, reducing the risks for thousands of Colombian families.

The remaining of this paper is arranged as follows. The next section presents the dissemination system design. The Sect. 3 describes the experimental prototype and defines the experiment conditions. The Sect. 4 presents the obtained results and their consequent discussion. Finally, the Sect. 5 exposes the obtained conclusions and proposes future works for further research.

2 System Design

This section presents the main features of the design of the proposed dissemination system. Firstly, we establish the system's objective, goals and requirements. Secondly, we define the different roles and agents in the proposed system. Finally, we illustrate the relationships between the system agents through a sequence diagram.

2.1 System Requirements

The idea of conceiving, designing, implementing and operating a system of automatic dissemination of alerts for diseases in crops (coffee rust) through convergent services (telecommunication services and Web 2.0) reporting the severity of the situation to the farmers, is born from the relevance of control of coffee rust for Colombian agriculture and the opportunity offered by solutions such as early warning systems [3].

These early warning systems are composed of four key components: the Knowledge component is based on previous models to estimate the risk that each situation may represent; the Monitoring component comprises all the infrastructure that monitors the occurrence of relevant situations and detects any imminent risk (based on the Knowledge component); the Dissemination component is responsible for ensuring that alert messages reach end users (or Crop Managers in this case) through multiple channels or services; and the Response module includes all people involved with the risky situation (all managers and potentially affected farmers) and indicates how they should be organized and proceed to mitigate the risk [9].

Rust-oriented early warning systems based on the aforementioned researches have functional Knowledge, Monitoring and Response components. However, the dissemination of alerts is done by low-effective slow channels (bulletins), thus, extending it with convergent services could greatly improve its performance.

As part of a larger system, the Dissemination component must receive the request for disseminating the alert from the Monitoring component (based on the Knowledge component), and then proceed to display the alert messages through the different service providers, which must deliver the alerts to the respective Crop Managers.

Nevertheless, considering that the Dissemination component -and not the Monitoring component- is responsible for selecting the services to be activated according to the Knowledge component, and that the Knowledge component defines at least three types of alerts considering their risk level (Green or low, Yellow or intermediate, and Red or high alerts), the alert dissemination system based on convergent services should be responsible for sending the alert messages by different means to different Managers depending on the alerts' level and location.

In order to build a system for the automatic dissemination of early warnings for diseases in crops, ICT will be used as a means of sending alerts to the Crop Managers. Specifically, mobile telephony has been chosen because of its potential to timely deliver alert messages to every interested person regardless of their location (assuming they are in the mobile coverage area); Web 2.0 services have also been chosen for their capability to transmit information of different types to a huge number of people in a very short time [12].

Among the services offered by mobile telephony, short message service (SMS) was chosen since it is very fast to send for the system and light to read for the Crop Managers, so they can check the website for recommendations [13], understand the implications of the alert, and proceed to effectively and efficiently mitigate the risk; Twilio® platform was selected as telecommunications service provider. The voice call service was discarded for the time it takes to establish a connection with the users and the added computing resources it consumes.

Moreover, two Web 2.0 services were chosen: Twitter® (status and direct message) and Gmail® (email). The first service allows to inform all the subscribed users ("followers") what is happening with the crops, in addition to inform the interested managers directly; and the second one allows to deliver detailed information on the alert conditions to the managers in a more private way.

2.2 System Agents

Some relevant agents and their roles in the deployment of coffee rust early warnings (regardless the alert level) can be identified from the system design. The main agents are:

- Crop Manager: the person who is in charge of the control of the crop affected by coffee rust, this is the end user of the system and is the one who, alongside the affected farmers, must proceed to mitigate the incidence of rust in their crops.
- SMS provider: (Twilio) is the provider of telecommunications services over the Internet, it is capable of sending short text messages to almost any mobile phone in the world -for a small price ($ 0.06 USD)-, as long as it is connected to a cellular network.
- Email provider: (Gmail) is the email service provider, it has the ability to send large messages to several people simultaneously in a timely fashion, assuming they have an electronic address registered in any email provider, most of which have no cost.
- Social network provider: (Twitter) is the social network service provider, it is capable of sending free short messages in two ways: as a broadcast message that instantly reaches a virtually unlimited number of people within the social network (status or tweet), or as a private message addressed to a single person (direct message). Most of the services of this provider have no cost.
- Dissemination system for convergent services: this is the service to be developed, it must perform all necessary processing of the alerts in order to identify and send them quickly and effectively through Twilio, Gmail and Twitter.
- Managers database: this database is part of the Knowledge component in the monitoring system; it contains all the relevant contact information for sending the alerts to the managers, mainly their mobile telephone of contact, email address and Twitter ID.
- Alert generation system: this is a combination of the Knowledge and Monitoring components, this system is responsible for monitoring crop areas, identifying risky situations, identifying the interested managers, and generating the necessary alerts (green, yellow or red).

2.3 Relationship Diagram

Once the roles and functions of each agent involved in the dissemination of coffee rust early warnings are defined, the relationships among them are important, since they should relate differently depending on the alert level.

On Fig. 1 a sequence diagram illustrating the relationships between the seven system agents is presented. These relationships work on a standardized way, the Alert-detection system, the Managers database, and the Crop managers are always present on the alert deployment; however, not all the service providers or their services are invoked on each alert level.

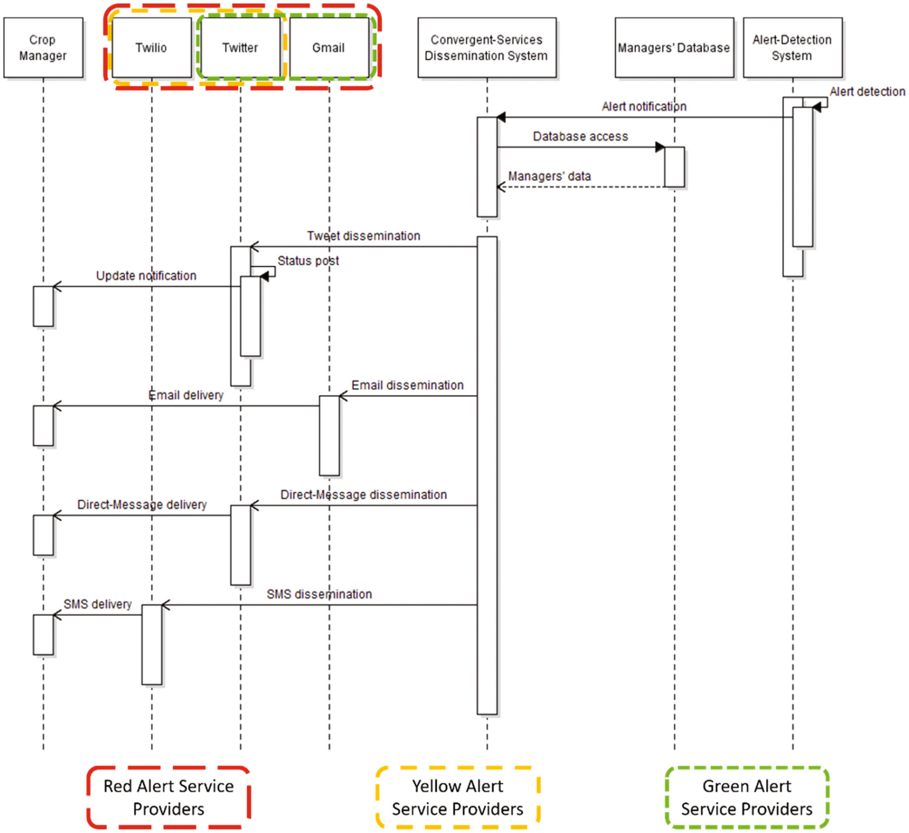


Fig. 1. Agents' relationships diagram

Green Alerts

Green alerts are the least severe, they indicate that the identified conditions could allow for some coffee trees to become infected with rust. For this alert level, the Alert-detection system identifies the described situation, sending a notification indicating the alert severity and the threatened crops to the Dissemination system, which looks for the interested managers on the Database and starts deploying the alerts through two services: the tweet (a broadcast message via Twitter), and the email (a multicast message via Gmail). Finally, the Crop Manager receives the warning and should start working with the farmers in order to prevent the coffee rust from infecting their crops.

Yellow Alerts

Yellow alerts have a medium level of severity, they indicate that the identified conditions are allowing for some coffee trees to become infected with rust. For this alert level, the Alert-detection system identifies the described situation, sending a notification indicating the alert severity and the threatened crops to the Dissemination system, which looks for the interested managers on the Database and starts deploying the alerts through three services: the tweet (a broadcast message via Twitter), the email (a multicast message via Gmail), and the direct message (a private message via Twitter). Finally, the Crop Manager receives the warning and should start working with the farmers in order to prevent the coffee rust from spreading over their crops.

Red Alerts

Red alerts are the most severe, they indicate that the identified conditions have allowed for the coffee rust to infect the crops and to spread. For this alert level the Alert-detection system identifies the described situation, sending a notification indicating the alert severity and the threatened crops to the Dissemination system, which looks for the interested managers on the Database and starts deploying the alerts through all the services: the tweet (a broadcast message via Twitter), the email (a multicast message via Gmail), the direct message (a private message via Twitter), and the SMS (a private message sent to the managers' phones via Twilio). Finally, the Crop Manager receives the warning and should start working with the farmers in order to avoid the coffee rust from killing their crops.

3 Experimental Prototype

This section provides a complete description of the experimental prototype, built for evaluating the feasibility of a large scale commercial system. In the first place, we illustrate the deployment model of the system. Subsequently, we explain the services components, illustrating their internal operation. Thereafter, we describe the implementation environment, which acts as a testbed for the experiment. Finally, we provide the experimental design, describing the main methodologies and features of the experiment.

3.1 Deployment Model

The deployment model of the experimental prototype (Fig. 2) allows to identify four main categories where the system agents (described in Subsect. 2.2) can be classified.

Main System

Though this is the dissemination component, *i.e.*, the developed system, it must communicate with the other components of the early warning system in order to succeed when mitigating the coffee rust risks. The agent classified on this category is the Converged-services dissemination system, which is implemented in the Mobicents JSLEE Server.

JSLEE is a short name for the JAIN SLEE Specification (Java APIs for Integrated Networks - Service Logic Execution Environment), which is a low-latency high-performance standard for building, managing and executing robust network applications [11]. This

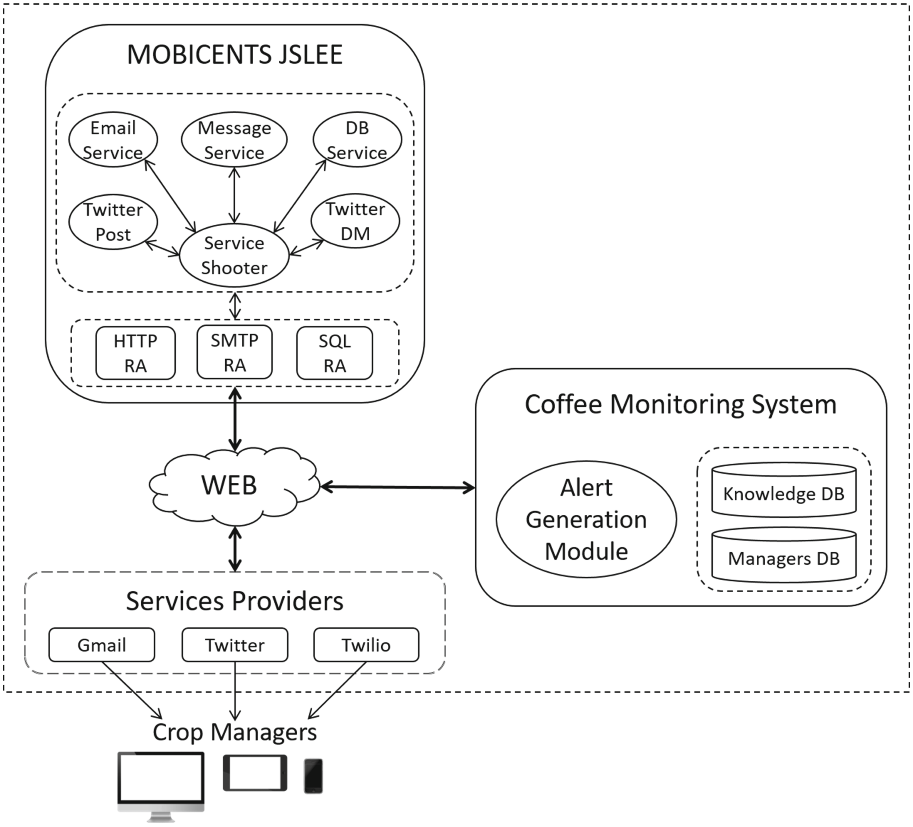


Fig. 2. Deployment model of the experimental prototype.

specification allows us to develop and deploy the converged services via Mobicents, a JBoss-based open-source server for JSLEE applications.

This system is divided on two main parts: Service Logic Container and Service Adaptation Container. The first one contains all the defined services, which are explained in Subsect. 3.2; while the second one contains the Resource Adaptors (RA), which allow the server and the services to connect with external components through different protocols, *i.e.*, HTTP, SMTP and SQL.

Partner System

This system includes the Knowledge and the Monitoring components of the early warning system. Though it was developed separately [14], it is essential for mitigating the coffee rust risks. The agents classified on this category are the Alert-detection system (the Monitoring component supported on the Knowledge component) and the Managers' database (part of the Knowledge component), which are implemented in a Coffee Monitoring System.

The two early warning system components can be identified on this system: the Knowledge component is represented by the database part, storing the risk knowledge and the managers' information; while the Monitoring component is represented by the Alert Generation Module. Further information about this module can be found in [14].

Third Parties
These are the service providers we selected for delivering the alerts. The agents classified on this category are Twilio, Twitter and Gmail. These service providers can be accessed through the Web, though they can deliver the alerts via different networks.

Final Users
The agent classified in this category is the Crop Manager, who will receive the alerts through different channels in different means (*e.g.*, computer, tablet or phone), and should proceed to mitigate the incidence of coffee rust in their crops, alongside the affected farmers.

3.2 Services Components

The deployment model of the experimental prototype (Fig. 2) allows to identify six services inside the Service Logic Container. These services have various components that enables their operation, thus, the services' behavior can be defined from analyzing the functions and relationships of their internal components (Figs. 3 and 4).

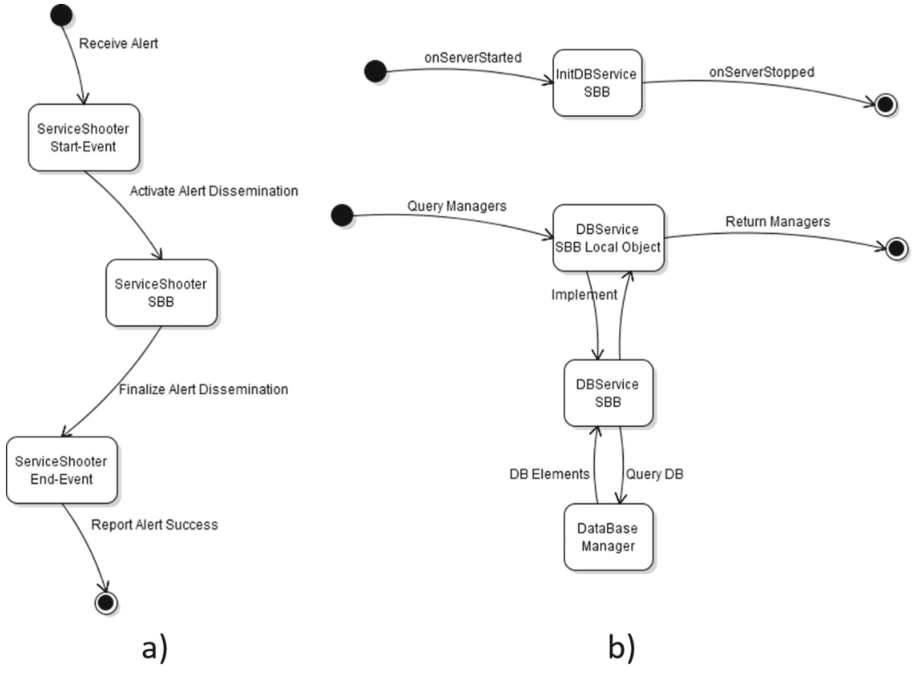


Fig. 3. Control and data services

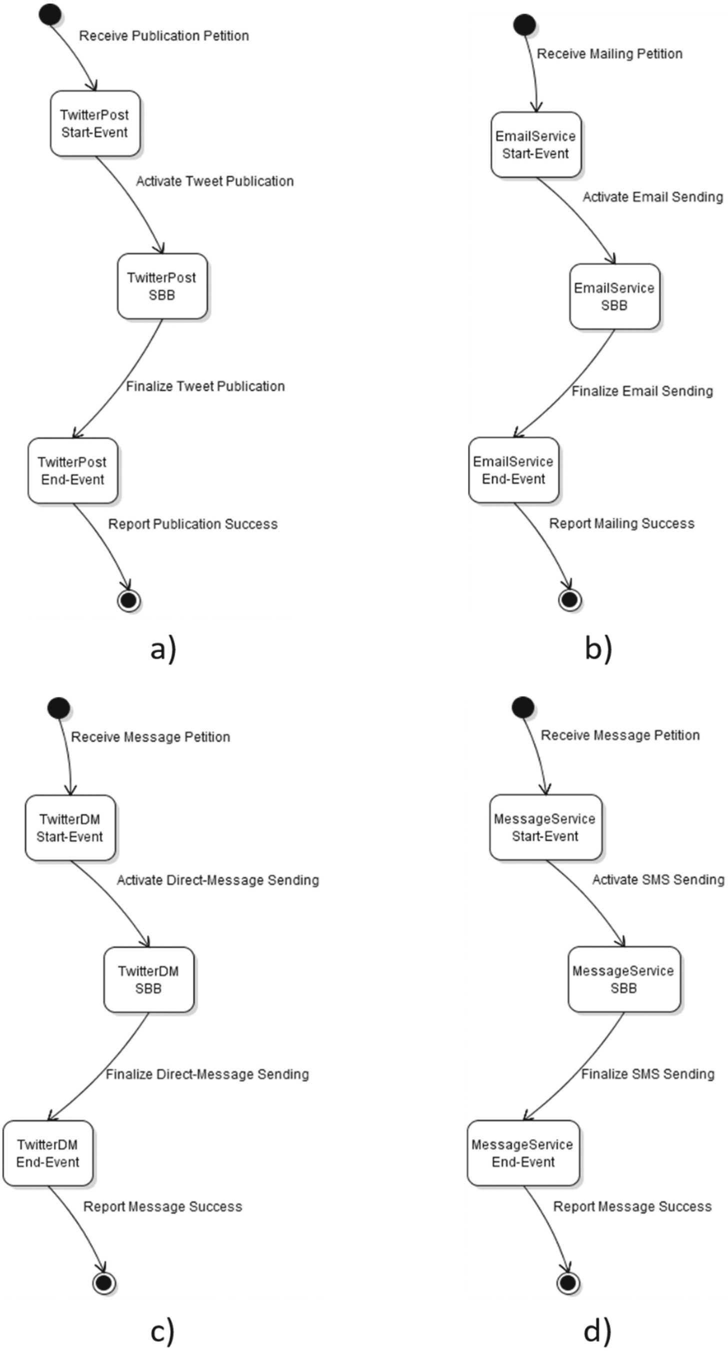


Fig. 4. Dissemination services

The services can be catalogued in three categories: Control (Service Shooter), Data (DB Service), and Dissemination (Twitter Post, Email Service, Twitter DM, and Message Service).

Control Service - Service Shooter

This service is responsible for receiving dissemination requests from the Coffee Monitoring System, and activating the different dissemination services. For this purpose, this service must rely on three distinct components (Fig. 3-a).

ServiceShooter Start-Event. This component allow the Service Shooter to be activated by the monitoring system, so the alerts can be deployed. This component has two functions: Receive the Request (allows to connect with any partner system through HTTP), and Activate Dissemination (activates the next component: ServiceShooter SBB).

ServiceShooter SBB. This component allows the Service Shooter to execute all the service control logic for accessing the database and activating the different services combinations for each alert level (Green, Yellow, and Red). This component has 13 functions: Format the Alert (integrates the received data to operate it in the JSLEE server), Select the Services (presets the services accordingly with the alerts' levels, as stated in Subsect. 2.3), Query Managers (connects with the DB Service in order to obtain the interested managers), Format the Managers (integrates the managers' data to operate it in the JSLEE server), Configure Twitter Post (prepares the message to be posted as a tweet), Connect with Twitter Post (sends the prepared message to the Twitter Post service), Configure Email (prepares the message to be sent as an email), Connect with Email Service (sends the prepared message to the Email Service), Configure Twitter DM (prepares the message to be sent as a Twitter direct message), Connect with Twitter DM (sends the prepared message to the Twitter DM service), Configure SMS (prepares the message to be sent as a SMS), Connect with Message Service (sends the prepared message to the Message Service), and Finalize Dissemination (activates the next component: ServiceShooter End-Event).

ServiceShooter End-Event. This component allows the Service Shooter to finalize its operation, sending a confirmation to the partner system. This component has two functions: Inform Finalization (sends a finalization message to the partner system), and Finalize Execution (releases the reserved resources).

Data Service - DB Service

This service is responsible for accessing the Managers' Database in order obtain the interested managers, considering the alert level and threatened crops. For this purpose, this service must rely on four distinct components (Fig. 3-b).

InitDBService SBB. This component allows the DB Service to remain always-available. It is based on two functions: Initialize (starts the service indefinitely), and Finalize (stops the service releasing the reserved resources).

DBService SBB-Local-Object. This component allows the DB service to be invoked by the Service Shooter at any time. It has the same functions that the DBService SBB, however, since it is an interface, it does not execute any logic.

DBService SBB. This component executes all the logic from the SBB-Local-Object for accessing the database. This component has six functions: Start Query (receives the conditions of the managers), Establish Credentials (sets the database URI, user, and password), Format the Query (prepares the query in SQL), Execute Query (sends the query to the DataBase Manager), Format Data (prepares the extracted data for the Service Shooter), and Return Data (sends the managers' data to the Service Shooter).

DataBase Manager. This component allows the DB Service to communicate with the database, sending queries and returning the managers' data. It is based on two functions: Query Managers (connects with the database management system for executing the query), and Return Managers (sends the received data to the DBService SBB).

Dissemination Services

The dissemination services are responsible for connecting with the services providers in order to deliver the alerts to the managers. For this purpose, four services have been defined: Twitter Post (Fig. 4-a), Email Service (Fig. 4-b), Twitter DM (Fig. 4-c) and Message Service (Fig. 4-d). Both of the Twitter services connect with the Twitter provider, while the Email connects with Gmail, and Message connects with Twilio.

As observed on Fig. 4, all the dissemination services have the same components structure, therefore, it is explained as follows:

Start-Event. This component allows its service to be initialized by the Service Shooter in order to deliver an alert. This component has two functions: Receive the Request (allows to connect with the Service Shooter), and Activate Dissemination (activates the next component: the SBB, for initializing the dissemination).

SBB. This component allows its service to execute all the service logic for connecting with the provider and delivering the alert. This component has four functions: Format the Data (prepares the necessary data for sending the message), Establish Provider's Credentials (sets the account configuration required by the provider), Request Message Delivery (connects with the service provider to deliver the message), and Finalize Dissemination (closes the connection and activates the next component: End-Event).

End-Event. This component allows its service to finalize the operation, sending a confirmation to the Service Shooter. This component has two functions: Inform Finalization (sends a finalization message to the Service Shooter), and Finalize Execution (releases the resources of the service).

3.3 Implementation Environment

All the tests were realized on an Asus® K555L laptop, with a 1-TB Hard Disk Drive, an Intel Core i7-5500U Central Processing Unit, 12 GB of RAM memory, and Ubuntu 14.04 LTS. The selected server was the Mobicents JSLEE 2.7.0. FINAL - RAY, with the default configuration. Finally, the Internet connection for communicating the services with the providers was a standard 5-Mb/s connection, activating some bandwidth-demanding applications for emulating a busy network.

3.4 Experimental Design

In the context of early warning systems, there are two main aspects to consider when evaluating the dissemination services: the execution time (Time, in seconds), and the success rate (Success, in percentage), since these parameters allow to determine if the alert message achieves at reaching the necessary percentage of crop managers (users) opportunely, *i.e.*, with enough time for taking actions to mitigate the risks of the detected situation [11].

Moreover, considering that such convergent services represent the Dissemination component of an early warning system, the rate of informed users (Informed Users, in percentage) should be considered too, since this parameter indicates the percentage of users that have received the alert through any of the channels defined by the component. In the context of rust alerts, the user is considered as informed or alerted when he/she has received a personal alert message (to his/her social network, email or mobile phone) and the alert has also been broadcasted (in a social network).

In order to evaluate the performance of the rust-warning dissemination services through convergent services in terms of execution time, success rate, and informed user rate, a black box model was used; observing only the input data (alert request), and the obtained results (public messages and messages received by the managers). Following this model, three tests were performed for a total of five managers, considering the three alert scenarios defined in Subsect. 2.3:

- Green Alert: low risk alert, only two managers are interested in this type of alerts (2 users).
- Yellow Alert: intermediate risk alert, in addition to the two previous managers, two more are interested in this type of alerts (4 users).
- Red Alert: high risk alert, in addition to the four previous managers, one more is interested in this type of alerts (5 users).

The tests were repeated five times for a total of 15 runs, averaging the results at the end of each test.

4 Results and Discussion

After executing the 15 test runs, we present the obtained results (Fig. 5) considering the System Design and Experimental Design in order to establish the requirements for the number of services per user that each type of alert should send, and thus how the success rate is evaluated on each test.

- Green Alert: This alert type should alert users through a tweet on Twitter and an email message. The test performed for this alert type showed a success rate of 100% in all iterations, therefore, an informed user rate of 100%, and an average time of 2.75 s per user (Fig. 5).
- Yellow Alert: This alert type should alert users through a tweet on Twitter, a direct Twitter message, and an SMS to the personal phone. The test performed for this alert

type showed an average success rate of 99%, despite the informed users were the 100%, with an average time of 1.5 s per user for displaying all the alerts (Fig. 5).

- Red Alert: This alert type should alert users through a tweet on Twitter, an email message, a direct Twitter message, and an SMS to the personal phone. The test performed for this alert type showed an average success rate of 98%, despite the informed users were the 100%, with an average time of 2.1 s per user for displaying all the alerts (Fig. 5).

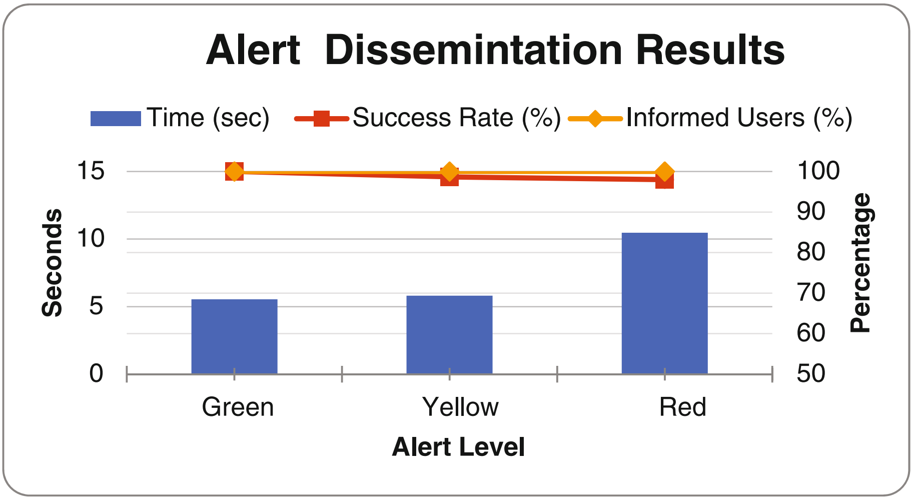


Fig. 5. Experimental results

In addition to the of informed user rate (Informed Users), success rate (Success) and execution time (Time), the error reports were collected in the events of message failure, obtaining the error 30008 of Twilio [15] in all the events, since the SMS was the only service that failed.

It is also important to highlight that the alerts sent by the dissemination service did not confuse those responsible for each alert, that is, the red-alert users only received red alerts, the yellow-alert users received red and yellow alerts, and the green-alert users received all three types of alerts, in all cases and all iterations.

Although the results presented in Fig. 5 show an excellent behavior in the services delivery, the success rate decreased as the alert relevance increased, an undesired scenario that must be analyzed. For the most severe alerts a success rate of 98% and a time of 2.1 s per user could be not enough, since the system could handle many individual users (crop managers) who would be important to alert in a very short time.

Thus, regarding the execution time, crops diseases such as the coffee rust do not usually generate massive alerts and have very large windows of opportunity, even longer than one hour, therefore, the converged services developed are optimal for the dissemination of rust alerts on coffee crops.

On the other hand, regarding the success rate, it is clear for early warning systems that dissemination services and channels will not always work ideally, so they may fail (especially in emergency situations) [12]. However, it is desirable that the services achieve a 100% success rate under normal (non-emergency) conditions.

We have sought for a solution to the identified problem (Error 30008 - Message not delivered by Twilio), but it is an unknown error for the service provider [15], so we can assume that the problem may be related only to the destination phones, or to the telecommunications operators of such phones (since all the errors occurred with Virgin Mobile® users).

Finally, regarding the informed users rate, a percentage of 100% was reached thanks to the multiplicity of dissemination channels [11], combining different telecommunications services and Web services. Nevertheless, more telecommunications services are required, as the SMS was the only failing service, and users without an Internet connection could remain unwarned.

5 Conclusions and Future Works

The obtained results, presented in Sect. 4, allow us to conclude this dissemination system could be seamlessly deployed as a large scale component for the coffee-rust early warning system. This dissemination component will enhance the current early warning systems, allowing for a better exploitation of the opportunities the new researches on knowledge and monitoring bring. These opportunities could highly reduce the impact of coffee pests and diseases on the Colombian economy, mainly for the vulnerable smallholders.

Furthermore, the standard web interface of the dissemination system allows it to easily connect with any other crop-oriented early warning system, creating a better protection for any crop and shielding the Colombian agriculture sector. This interface could also allow other early warning systems from any area accepting opportunity windows of minutes to leverage our dissemination services. Moreover, this proposal could be turned into a public dissemination service for early warnings.

The proposed system could also support precision agriculture systems increasing the profitability of the Colombian crops by leveraging different agriculture-oriented models as the presented on [16], which will enhance the agricultural shielding against crises [3].

As future works we propose to complete the coffee-rust early warning system. As a first step, we would need to implement more Web and telecommunications services, like different social networks (*e.g.*, Facebook® or WhatsApp®), USSD messages, and automatic calls. Secondly, we should integrate a validation component as the proposed in [9]. In the third place, we have to implement a large-scale system for the alert delivery. Finally, aiming to complete the early warning system, the Response component must be enhanced, *i.e.*, give the crops managers and farmers training for correctly mitigating the rust risks, accordingly with each different alert.

Acknowledgements. The authors would like to thank Universidad del Cauca and AgroCloud project of the RICCLISA program for supporting this research, and Colciencias (Colombia) for the PhD scholarship granted to MSc(C) Julián Eduardo Plazas.

References

1. Ramirez-Villegas, J., Salazar, M., Jarvis, A., Navarro-Racines, C.E.: A way forward on adaptation to climate change in Colombian agriculture: perspectives towards 2050. Clim. Change **115**(3–4), 611–628 (2012)
2. Corrales, D.C., Ledezma, A., Peña, A.J.Q., Hoyos, J., Figueroa, A., Corrales, J.C.: A new dataset for coffee rust detection in Colombian crops base on classifiers. Sist. Telemática, vol. 12, no. 29, pp. 9–23, June 2014
3. Avelino, J., et al.: The coffee rust crises in Colombia and Central America (2008–2013) impacts, plausible causes and proposed solutions. Food Secur. **7**(2), 303–321 (2015)
4. Corrales, D.C., Corrales, J.C., Figueroa-Casas, A.: Towards detecting crop diseases and pest by supervised learning. Ing. Univ. **19**(1), 207–228 (2015)
5. Sambony, E.L., Corrales, J.C.: Sistema experto para enfermedades en cultivos basado en emparejamiento de patrones en grafos: una propuesta. Rev. Ing. Univ. Medellín, vol. 15, no. 29, pp. 81–98, May 2016
6. Lasso, E., Thamada, T.T., Meira, C.A.A., Corrales, J.C.: Graph patterns as representation of rules extracted from decision trees for coffee rust detection. In: Metadata and Semantics Research, pp. 405–414 (2015)
7. Corrales, D.C., Figueroa, A., Ledezma, A., Corrales, J.C.: An empirical multi-classifier for coffee rust detection in Colombian crops. In: Computational Science and Its Applications – ICCSA 2015, pp. 60–74 (2015)
8. Corrales, D.C., Casas, A.F., Ledezma, A., Corrales, J.C.: Two-level classifier ensembles for coffee rust estimation in Colombian crops. Int. J. Agric. Environ. Inf. Syst. IJAEIS **7**(3), 41–59 (2016)
9. Plazas, J.E., Rojas, J.S., Corrales, D.C., Corrales, J.C.: Validation of coffee rust warnings based on complex event processing. In: Gervasi, O., Murgante, B., Misra, S., Rocha, A.M.A.C., Torre, C.M., Taniar, D., Apduhan, B.O., Stankova, E., Wang, S. (eds.) Computational Science and Its Applications – ICCSA 2016, pp. 684–699. Springer International Publishing (2016)
10. Corrales, D.C., Peña, A.J.Q., León, C., Figueroa, A., Corrales, J.C.: Sistema de alerta temprana para la roya en el café basado en códigos de salida de corrección de error: una propuesta. Rev. Ing. Univ. Medellín, vol. 13, no. 25 (2014)
11. Plazas, J.E., Rojas, J.S., Corrales, J.C.: Mejorando los sistemas rurales de alertas tempranas a través de la integración de OpenBTS y jain slee, Rev. Ing. Univ. Medellín, vol. 16, no. 30, pp. 195–207, May 2017
12. Plazas, J.E., Rojas, J.S., Corrales, J.C.: An architecture based on JAIN SLEE and OpenBTS for rural early warning systems. Presented at the I Workshop de Tecnologías de la Información y la Comunicación aplicadas a la Agricultura - TIC-@gro, Popayán (Cauca), Colombia, pp. 51–56 (2015)
13. ALERTAS - AgroCloud. http://190.90.112.7:8081/web/guest/alertas. Accessed 08 Dec 2016
14. Lasso, E., Valencia, Ó., Corrales, J.C.: Decision support system for coffee rust control based on expert knowledge and value-added services. In: Computational Science and Its Applications – ICCSA 2017, pp. 70–83 (2017)
15. Error - 30008. https://www.twilio.com/docs/api/errors/30008?utm_source=docs&utm_medium=social&utm_campaign=guides_tags. Accessed 13 Dec 2016
16. Plazas, J.E., López, I.D., Corrales, J.C.: A tool for classification of cacao production in Colombia based on multiple classifier systems. In: Computational Science and Its Applications – ICCSA 2017, pp. 60–69 (2017)

Improving Early Warning Systems for Agriculture Based on Web Service Adaptation

Oscar Ricardo Valencia(✉), Emmanuel Lasso, and Juan Carlos Corrales

Grupo de Ingeniería Telemática, Universisdad del Cauca,
Campus Tulcán, Popayán Cauca, Colombia
{ovalencia,eglasso,jcorral}@unicauca.edu.co

Abstract. Early warning systems are designed to inform the largest number of users, such as a country or a region, about a risky situation. However, in specific domains such as agriculture, it is commonly required that these alerts be more specific according to the crops location and their properties, consequently the web services of these systems must be adapted. On the other hand, the Enterprise Services Bus with its mediation capabilities (such as message transformation and routing) and Complex Event Processing with their monitoring characteristics can be integrated to meet the adaptation requirements of web services at runtime. This paper presents an improvement for Early Warning System for coffee production that, according to the area in which a crop is located and its phenology, manages the adaptation of alerts for coffee rust, based on the integration of an Enterprise Services Bus and a Complex Events Processing.

Keywords: Web service adaptation · Early warning systems · Enterprise service bus · Complex event processing

1 Introduction

In agriculture, farmers are faced with several problems during the production of different crops [1, 2]. The diseases and pests can affect crops which decrease the quality and quantity of production. Generally, their origin is due to the attack of fungi, bacteria, plagues, and other agents. The generation of diseases mainly depends on the physiological status of the crop, meteorological conditions, production zones of the country, among others.

Over time, diseases have devastated crops throughout the world, causing epidemics and losses [3–5]. This implies that important measures must be taken to mitigate the effects and improve productivity through practices such as crop protection, the use of clean technologies, among others. For this reason, different tools that allow the detection of different diseases for subsequent treatment are used. An Early Warning System (EWS) can perform this detection, through monitoring and analysis of crops conditions, which allow to provide timely information to the farmers about possible natural, biotic and physical threats in such a way that they can prepare and take the necessary measures according to the level of crop risk [6].

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_11

The EWSs are usually designed to inform a large number of users, e.g. a large population such as a country or a region, of any risk situation. In the agriculture, these alerts should be more specific, so that they include characteristics as the location of the crop, as well as some of its properties. Taking into account the above considerations, an improved EWS for diseases in crops is proposed for the generation of alerts and their adaptation through the integration of an Enterprise Services Bus (ESB) and Complex Events Processing (CEP), according to the area in which the crop is located and its phenology.

This paper is organized as follows: Sect. 2 presents the motivation; Sect. 3 contains the state of the art; Sect. 4 describes the solution proposed from the integration of the ESB and CEP to the EWS; Sect. 5 presents the evaluation of the solution through a case study; Sect. 6 describes the results obtained. Finally, Sect. 7 relates the conclusions and future work.

2 Motivation

Currently, EWSs are an important tool to prevent or reduce risk in the face of disasters, illnesses or any predictable events [7]. These systems allow all users to have timely, clear and effective information, in order to take appropriate measures to avoid, reduce or respond adequately to a risk. An effective and efficient EWS contains four key elements: **Knowledge of Risk**, that provides essential information for assessing current risk and prioritizing its mitigation, prevention and response strategies; **Monitoring and Warning Service** component, which allows an analysis of captured information to provide timely estimates of potential risk in order to generate accurate and early warnings; **Dissemination and Communication**, responsible for managing all necessary procedures to send reliable, simplified and understandable warning messages to the involved users through various channels or services; **Responsiveness** to handle the knowledge of risk and response mechanisms and communicate them to the population, e.g. appropriate action plans to react to a danger situation, in addition to the tools that are available to deal with a phenomenon [7].

Typically, an EWS produces notifications in a general way, e.g. covering the greatest number of populations, such as alerts for an entire country or region. Due to this, many alerts are sent to all users within the territory, but not necessarily the alert affects all notified users, in other words, only some of the users could be affected by the risk announced in the alert. In agriculture, an EWS needs to be much more specific, since the crops are highly related to the particular conditions of the place where they are located. Therefore, in order to get early warnings relevant for the farmers, The EWSs should include crops information as the location, agronomic information, among others, otherwise the alert generated would not achieve the purpose of avoiding or reducing risk for a specific problem and area.

This paper proposes the implementation of two of the four main EWS components oriented to coffee rust detection: *knowledge of the risk*, acquired through techniques for the identification of favorable weather conditions for the appearance of Coffee Rust, and the *monitoring and warning services*, in order to personalize early alerts based on user information, through the adaptation of the alert-generation services.

3 State of the Art

This section provides relevant information about studies carried out in the development of EWSs in agriculture and personalization and adaptation of web services.

The EWSs are widely used in different domains as the detection of diseases in crops. These include key elements such as: knowledge and threat mapping, monitoring and forecasting of impending events, and diffusion of alerts [8]. The correct combination of these elements results in a complete and reliable system that provides timely and understandable information to the population. In addition, an EWS should be people-centered, allowing communities threatened by a phenomenon to be able to act in a timely and appropriate manner so as to reduce potential losses.

In [9], weather data is included within an EWS for *late blight* in potato crops. Eight experimental plots with weather monitoring were used. After the data analysis and processing, a fixed calendar was obtained containing the alarm dates for the disease occurrence and treatments to be carried out in the crops. In [10] an EWS to mitigate the impact of cockcrowing on coffee crops is presented. This system is based on mathematical models that use meteorological data to predict the favorable conditions for the development of the disease and recommend the fungicides application at the right time, thus preventing its development. On the other hand, weather data monitored in crops is used in [11] as support for decision making, helping producers to determine the right moment for disease control.

Some EWSs make use of web services to provide data to the user. Precisely, the web services personalization collects user information during its interaction with the system, so that it is used to provide adequate contents and services for each user, in order to improve the user experience [12]. The purpose of using web service personalization is to improve an EWS with relevant information of each user, allowing that its response can be delivered only to those potentially affected.

Furthermore, web services can be modified in such a way that it can be adjusted to different circumstances or conditions, either by the specific user needs or by the environment in which they are running [13]. This feature is called *Adaptation.* In [14] an approach using the mediation characteristics of the ESB to provide a dynamic adaptation of the services is presented. In the same way, the research presented in [15] describes an ESB implementation using JBossESB, focused on the description of a services adaptation framework based on the ESB capabilities previously mentioned. In [16] an architecture for the adaptation of context-based services is proposed, making use of the user information captured and analyzed by a CEP. Similarly, the research described in [17] proposes a solution that uses the ESB mediation patterns to adapt the services to a context, making use of the CEP to analyze the events received from external resources and detect the relevant situations that require an adaptation.

The studies presented above address the mediation features provided by the ESB to perform the adaptation of services, and some use CEP to identify context events (e.g. climate) that are relevant to the user. Although they consider an integration with CEP, this is conceived as an external and additional component of the ESB. In fact, some studies propose the use of the ESB to integrate event-driven architectures using CEP [18].

Therefore, we propose to carry out the adaptation of web services based on an ESB-CEP integration, making use of the mediation capabilities of the ESB to combine Services-Oriented Architectures (SOA) for the generation of intercommunicated services [19] and Event-Driven Architectures (EDA) to design and implement applications based on events [20]. The combination of these architectures allows the integration of different applications in heterogeneous environments and, through ESB, these can be unified to operate in specific areas. For the same purpose, the CEP monitors relevant situations (events) according to the system requirements. Our approach proposes a system with the ability to react to environmental events (expected or unexpected), in order to ensure the proper operation of web services.

4 ESB – CEP Integration for Web Services Adaptation

Although in [14, 15] an adaptation of the services using the mediation characteristics of the ESB is implemented, these proposals do not use CEP to detect relevant events that require adaptation. Our proposal is aimed at the web services adaptation around the *monitoring* component of an EWS, since these systems do not provide solutions for a dynamic and automatic adaptation at runtime based on users needs [21]. This is addressed through the integration between ESB and CEP, acting as a mediator element through which all messages are transported and can be transformed, routed, invoked, among others. Also, it provides the analysis, detection and correlation of the messages that are exchanged by the system actors, in order to identify relevant events of interest in an application scenario.

Therefore, following the approach proposed in [15], where the adaptation is performed within the ESB using its mediation capabilities, we introduce the model presented in Fig. 1.

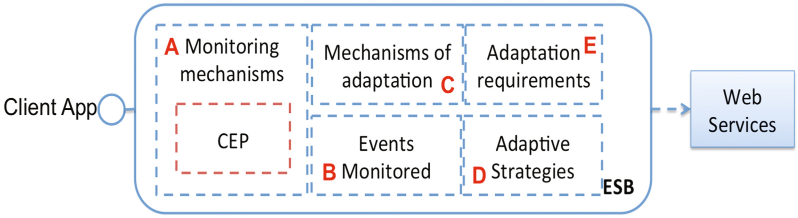


Fig. 1. Model of web services adaptation

Figure 1 shows the ESB-CEP integration to perform the adaptation process, which contains the elements presented in the adaptation and monitoring framework proposed by the S-Cube project [22]. This framework was adjusted to include the CEP within the ESB. The components are presented below.

Monitoring mechanisms: check if the current situation corresponds to the expected one. For example, a mechanism that monitors the parameters related to the invocation of the service, addition of new operations, among others. This component can be addressed through the complex event processing (CEP) in charge of monitoring the events of interest.

Monitored events: provide relevant information on the execution, evolution or context of the services, detected through the monitoring mechanisms. These events represent the difference in the status, functionality or expected environment. For example, the number of parameters entered when invoking an operation exceeds the expected amount.

Adaptation requirements: represents the need to modify a service to eliminate the difference between the current and expected state. For example, the inclusion of parameters related to the user location in the service operations. These may include requirements for reliability or functional accuracy, optimization, interoperability and usability.

Adaptation mechanisms: contain the techniques and facilities provided by the ESB to carry out strategies according to the adaptation requirements.

Adaptation strategies: alternatives to meet adaptation requirements according to the current situation. For example, to invoke an equivalent operation that takes into account new parameters. Different strategies can be used for the same situation, which requires decision mechanisms to select the most appropriate.

The components described above are involved in the adaptation process as follows: the monitoring mechanisms are responsible for detecting the situations of interest. These include the elements that contribute to the identification of monitored events and adaptation requirements. The adaptation mechanisms provide the possible elements that will be used to perform the adaptation based on the basic characteristics of the ESB (routing, transformation, mediation, among others). The strategies and mechanisms of adaptation depend entirely on the messages handling property of the ESB. Finally, the adaptation is performed from the requirements and monitored events.

In summary, the general process proposed in Fig. 1 is executed as follows: **A.** Relevant information is collected through monitoring mechanisms; **B.** Critical events are detected; **C.** The need for adaptation is identified; **D.** A strategy for adaptation is identified; and **E.** Adaptation is executed using the available adaptation mechanisms [22].

Algorithm 1, realize adaptation process describe previously, which has as inputs data the user location (posx, posy) and flowering date of crop (day, month, year). The algorithm will response a recommendation adapted to the inputs. For this the algorithm start getting coffee rust rate from climatic conditions using the following variables: crop shade (sombra), average precipitation (prdmed), hours of relative humidity higher than 90% on the day (horhr90), hours of relative humidity higher than 90% in night (horhrm90), average temperature when relative humidity higher than 90% (thrn90), maximum temperature (tmax), average temperature (tmed) and minimum temperature (tmin). These variables are calculated for that reason are not entered by the user. After the algorithm verifies the user location if is a valid location invoked function getLocation this function find if location are in the risk zone of coffee rust and based on the response generate a recommendation. In case the algorithm verifies the valid flowering date invoke function getRecommendationFlowering this function generate a recommendation based on the date and coffee rust rate.

Algorithm recommendation is based on user provided information that is to say when algorithm know user localization the recommendation include information about location related with coffee rust rate. e.g., if algorithm knows that user location are Risaralda department and besides that coffee rust rate are medium, algorithm recommendation in this case are fungicides, dosage and spray equipment used for this rate and available in Risaralda department. When algorithm know crop flowering date the recommendation are based on this date and coffee rust rate. e.g., if algorithm know that coffee rust rate are high and flowering date suggest days for high risk of coffee rust, the recommendation are measures for disease control.

Adaptation of Recommendation Function

```
INPUTS: posx, posy, month, year
OUTPUS: Coffe rust recommendation

Start
$rustRate = getRustRate(sobra, predmed, horhr90, horhrn90, thrn90, tmax, tmed, tmin)$
if exist an event with $name \neq null$ and $posx, posy$ are within the range of coordinates from Colombia
    $location = getLocation(posx, posy)$
    if $location$ contains $warning$
        $recommendation = location + rustRate$
    else
        $recommendation = rustRate$
    End if
else
    if exist an event with $day > 0$ and $month > 0$ and $year > 0$
        $flowering = getRecommendationFlowering(day, month, year)$
        $recommendation = rustRate + flowering$
    else
        $recommendation = rustRate$
    End if
Return recommendation
End
```

Algorithm 1. Adaptation of Recommendation Function

The model of adaptation described above can be applied to any environment. We applied it for the generation of early warnings for crop diseases, taking as a particular case the Coffee Rust, which will be explained in the next section.

5 Case Study: Coffee Rust Detection in Colombian Crops

In this section we present an application of the proposal mentioned above, describing the problem domain and the cases of study to apply our solution.

5.1 Application Domain

Coffee Rust is the main high-impact phytosanitary problem for coffee cultivation, caused by the pathogen *Hemileia vastratix* [23], a fungus that develops only in the living tissue of its host plant, for this case the coffee leaves. In Colombia, this disease affects crops located between 600 to 1000 m above sea level, with a negative impact on more than one million families that depend on the coffee production. The consequences related to the disease mainly are relate with economic and quality losses for coffee farmers. There is in fact an urgent need to take preventive measures for reducing the losses generated by this disease. The infection of coffee rust and subsequent progress depend on the simultaneous occurrence of four factors, described below:

- Susceptible plant or host: the variety of coffee sown in a crop can make it susceptible to rust infection or not. In addition, the physiological status of the plant, soil, excess water, nutritional deficiencies, among others factors, can affect the disease development.
- Pathogenic organism or causal agent: the causal agent of coffee rust is the fungus *Hemileia vastatris* belonging to the uredinales family.
- Environment: refers to weather states conducive to the generation of the disease. The fungus needs very special conditions to start the process of dispersion between leaves and plants such as splashing rain, presence of a layer of water on the underside of the leaves to germinate, all accompanied by temperatures between 16 and 28 °C and low solar brightness conditions.
- Human action: everything related to agronomic management depends on human action, such as: crop management, fertilization, shade, chemical control, among others.

Taking into account the factors mentioned above, some proposals have been made from the computer sciences related to the automatic identification of conditions conducive to this disease, making use of machine learning techniques. In these studies [25–31], the authors make use of the different variables involved in the development of coffee rust in crops, which have been monitored in them. As a result, a training dataset that relates the conditions of a crop with episodes of the disease is obtained. Several techniques such as induction of decision trees, support vector machines, assembly methods, are applied to the dataset generated. In the particular case of [28], an induction of decision trees to generate a model that predicts an infection rate of coffee rust from monitored data in crops is presented. This rate indicates whether the infection level of the disease tends to remain stable or reduced, to increase moderately or accelerate. Consequently, the results of this research can be used as the main element to alert coffee producers about favorable conditions (weather and agronomic properties) present in crops for the disease.

This last research is implemented within the AgroCloud project, which belongs to the Interinstitutional Network of Climate Change and Food Security of Colombia[1] (RICCLISA). In this project, a web platform that aims to provide monitoring services to warn users about threat situations for coffee crops located in the upper Cauca River basin has been developed. For this purpose, a set of services are deployed: Web services of climatic data (*WS_DClima*), in charge of obtaining all the weather data that are monitored in different coffee production zones; Web services for the calculation of the variables that describe the crop condition against the disease, generated from weather data (*WS_VClimaticas*); Web services for alert management (*WS_Alertas*), which perform the database management of generated alerts and allow their search, update and deletion.

However, these approximations do not consider the alert personalization, taking into account factors that determine how timely its diffusion, given the different coffee production zones in the country, which have distinct production cycles, flowering and harvest dates.

Thus, two case studies are presented for the personalization of early warnings by adapting web services. The first case takes the coffee crop location as an additional factor for the personalization and the second takes the flowering date for the same purpose.

5.2 Case 1: Coffee Crop Location

The first case is based on the generation of early warnings for coffee rust making use of weather data monitored in the area where a crop is located. From this location, a crop can be classified within a coffee zone.

In Colombia, there are three coffee zones divided by semesters where the main harvest is presented [24]; these are:

- Zone 1: main harvest in the second half of the year. It includes the departments of Guajira, Magdalena, Cesar, Norte de Santander, Santander, Antioquia, Caldas, Risaralda and Meta (Fig. 2).
- Zone 2: main harvest in the first half of the year. It includes the departments of Cauca and Nariño (Fig. 2).
- Zone 3: main harvest in the two semesters. It includes the departments of Valle del Cauca, Huila, Tolima, Quindío, Cundinamarca and Boyacá (Fig. 2).

5.3 Case 2: Coffee Flowering Date

The second case aims to generate early warnings for coffee rust based on weather data in the area where a crop is located an its flowering dates provided by the farmer. Flowering can be understood as a phenological state, directly associated with crop production [32]. Therefore, flowering periods in coffee are directly related to outbreaks of diseases. In the case of coffee rust, if the date of flowering is known, it is possible to identify the moment in which the crop is more prone to the disease, and from this, to establish the ideal dates to start the control process.

[1] www.ricclisa.org.

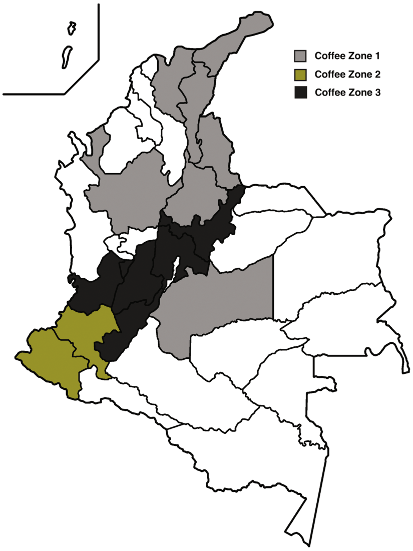


Fig. 2. Colombian coffee zones. Based on [24]

According to [24], knowing the quantity and concentration of flowering can establish the period where the application of fungicides is more effective. Figures 3, 4 and 5 show the map of the Colombian coffee zones and the phenological behavior of crops.

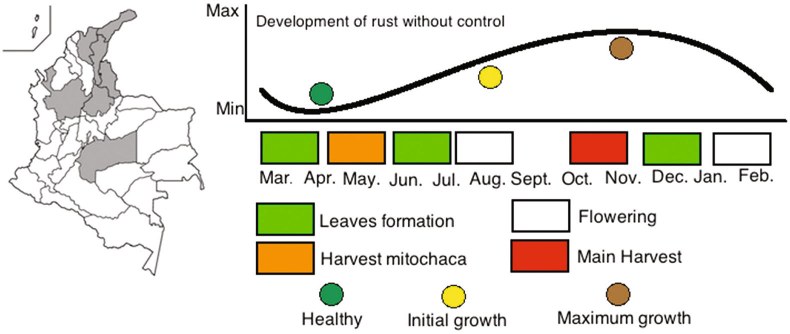


Fig. 3. Coffee zone 1 and phenology. Based on [24]

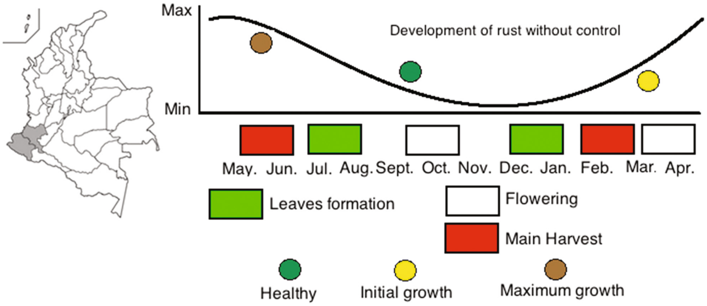


Fig. 4. Coffee zone 2 and phenology. Based on [24]

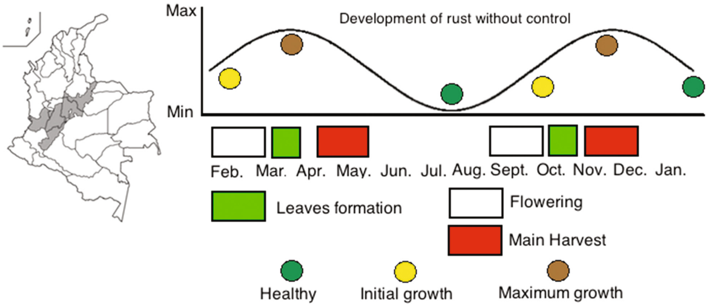


Fig. 5. Coffee zone 3 and phenology. Based on [24]

Figure 3 shows the development of coffee rust without chemical control in coffee zone 1, which affects the main harvest during the second half of the year. The phenological behavior of the plant is determined by: formation of leaves (green squares), *mitaca* harvest (orange squares), flowering (white squares) and main crop (red squares). In addition, a curve representing the growth of the disease over time can be observed as follows: healthy (green circles), initial growth (yellow circles) and maximum growth (brown circles). The estimated time of flowering before the main harvest for this case is between the months of July, August and September, which indicates that from these dates the process of disease development onset. This suggests that, after flowering, coffee rust management should be initiated in the crops, which, depending on the disease amount and concentration, should be started at 60 days, continuing with applications at 90, 120 and 180 days or following the fixed sprays schedule.

Similarly, Fig. 4 shows the development of coffee rust without chemical control in coffee zone 2 (The conventions of elements and colors are similar to those of Fig. 3). In this case, the disease affects the main harvest present during the first half of the year. The estimated time of flowering is between the months of September, October, November, March and April. The disease management should be made with the same considerations reported for zone 1.

Finally, Fig. 5 shows the development of coffee rust without chemical control in coffee zone 3. In this case, there are two main harvests per year and coffee rust affects each one and the estimated time of flowering before them is between the months of February, March, September and October. The disease management should be made with the same considerations reported for zone 1.

In summary, the adaptation proposal for each case is presented in Table 1:

Table 1. Adaptation summary per case

Case	Adaptation
Case 1	Adding of the user's location to identify the coffee zone in which he is located
Case 2	Adding of the flowering date to identify the critical periods of coffee rust onset

6 AgroCloud: WSA

This section addresses the implementation of the proposed solution for the study cases mentioned above within AgroCloud project, specifying the technology and the architecture used.

6.1 Technology

Our proposal was implemented using the following technologies:

- Enterprise Service Bus (ESB): Mule ESB [33].
- Complex Event Processing Engine (CEP): Esper [34].
- Application server: Glassfish [35].
- Database management server: PostgreSQL [36].

6.2 Implementation of the Proposed Solution

The generation of early warnings for coffee rust is addressed from the case studies described above. Consequently, we personalize the services by adapting them using the scheme proposed in Fig. 6, which contains the interaction of the elements related in Sect. 4.

The components of the scheme are described below.

- Web services: facilities implemented by AgroCloud to detect the favorable weather conditions for the appearance of coffee rust and generate the disease alerts (WS_DClima, WS_VClimaticas, WS_Alertas).
- GlassFish server: contains all the services developed (weather information, alert management, database access, among others).
- PostgreSQL server: persistence of the data handled by services.
- Mule ESB: through this bus, all messages between services, servers and applications are transported. Additionally, everything related to adaptation is developed in this element.

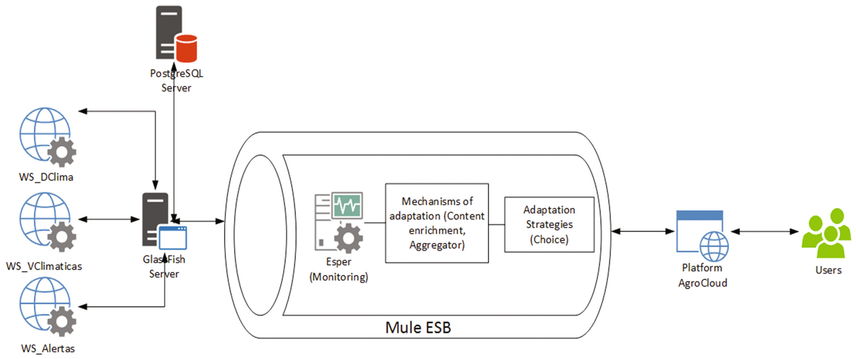


Fig. 6. Adaptation of ESB-CEP proposed

- Esper: CEP Engine in charge of monitoring the relevant situations presented in the case studies, in order to find all events that require an adaptation, according to the information provided by the users and their needs.
- Mechanisms of adaptation: tasks available within the ESB that allow the adaptation.
- Adaptation strategies: procedures to satisfy the requirements of adaptation according to the current situation.

Case Adapted 1

In order to implement the personalization of services for Case 1 (described in Sect. 5.2), the proposed execution flow is the following:

a. The user makes a query through the platform AgroCloud and the location of the crop is requested or detected.
b. Monitoring mechanisms (located on Esper CEP) process the location and recognize it as an event of interest.
c. The need for adaptation is identified and the mechanisms available in the ESB (content enrichment, aggregator) are determined.
d. The most appropriate adaptation strategy to meet the requirement is identified.

Case Adapted 2

In order to implement the personalization of services for Case 2 (described in Sect. 5.3), the proposed execution flow is as follows:

a. The user makes a query through the platform AgroCloud and introduces the date where the crop flowering was presented.
b. Monitoring mechanisms (located on Esper CEP) process the date of flowering and recognize it as an event of interest.
c. The need for adaptation is identified and the mechanisms available in the ESB (content enrichment, aggregator) are determined.
d. The most appropriate adaptation strategy to meet the requirement is identified.

7 Results

This section presents the tests performed on the proposed solution. We took into account the response time of web services, use of computational resources and efficiency in the events of interest detection. Each of the tests was performed considering situations with and without the proposed adaptation. It's important clarifying that the proposal is in the first version for that reason has not yet been tested by final users.

7.1 Use of Computational Resources and Response Time

In order to measure the use of computational resources (CPU and RAM) in the web services execution with and without performing the proposed adaptation, the VisualVM v1.3.9 tool was used. These services were deployed on a server with the following features: 16 GB of RAM, processor Intel Xeon 1.8 Ghz and Ubuntu 12.04 as operating system. The results are shown in Table 2. The use of both CPU and RAM is low and very similar in all case, indicating that the proposed solution does not generate an overload in computational resources. On the other side, in order to measure the response time of web services with and without performing the proposed adaptation process, the SoapUI tool v5.3.0 was used to invoke each of them. The results obtained based on 1200 invocations for each of the different web services are presented in Table 2.

Table 2. Use of computational resources

Web service	Average percentage of CPU usage	Average RAM usage (MB)	Average time (ms)
WS_Ubication	1.58	53.20	9.97
WS_Flowering	1.72	45.28	33.07
WS_TasaRoya	1.55	50.43	8.77
WS_Adaptation Case 1 (adapted for Case 1)	1.39	51.02	39.17
WS_Adaptation Case 2 (adapted for Case 1)	1.52	44.10	14.66

According to [37], the values obtained are considered as acceptable, since the average response times are below 200 ms.

7.2 Decision Effectiveness

In the test performed to the CEP to verify the effectiveness in the detection of interest events, ten cases were considered. First, 100 examples were generated divided into true instances (events requiring adaptation) and negative instances (events that do not require adaptation). Subsequently, the number of instances correctly classified by CEP was observed, determining the true positives, false negatives, true negatives and false positives. For the following cases the number of examples was increased in 100 successively until obtaining 1000 with a random proportion of true/negative instances in each of them. As a result, the number of false negatives and false positives was zero in all cases, which means that the CEP was able to identify the need for adaptation in the instances for each of the cases tested. This indicates that the process of detecting events of interest carried out by the CEP is effective and adequate for the proposed solution.

8 Conclusions

This paper presents an adaptation proposal for web services used in early warning systems for crops diseases, through a CEP-ESB integration, based on the adaptation and monitoring framework proposed by the S-Cube project.

The main contribution of our proposal is the use of the mediation capabilities provided by the ESB to carry out the adaptation process, in addition to the use of the CEP as a component in charge of monitoring events of interest that may require such adaptation. The proposal was developed using MuleESB and Esper, which allowed the implementation of all the components required by the solution. The use of computational resources (CPU and RAM) produced by the solution in its execution is among the acceptable ranges. Regarding the effectiveness of the CEP, acceptable results were obtained, however, it was not taken into account possible drawbacks with the input data (e.g. the introduction of a text string instead of an integer value) which would generate an error in the execution. The EWS proposed is aimed at the generation of for coffee rust that considers local properties of crops. Precisely, as related in [38], to combat this disease, the systematization and dissemination of these type of computational tools should be considered from regional research platforms to producer farm levels.

It is proposed as future work, to address other adaptation requirements, as well as the implementation and testing of other ESB and CEP products. On the other hand, incorrect input data could be solved by handling exceptions in web services.

Acknowledgements. The authors are grateful to the University of Cauca and its Telematics Engineering Group (GIT), the Colombian Administrative Department of Science, Technology and Innovation (Colciencias), AgroCloud project of The Interinstitutional Network of Climate Change and Food Security of Colombia (RICCLISA) for supporting this research and the *InnovAccion Cauca* project of the Colombian Science, Technology and Innovation Fund (SGR-CTeI) for PhD scholarship granted to MsC. Emmanuel Lasso.

References

1. Gómez, G.C.: Desarrollos científicos de Cenicafé en la última década. Rev. Acad. Colombaina Cienc. Exactas Físicas Nat. vol. 1, no. 30, pp. 89–100 (2005)
2. de Camargo, A., Pereira, A.R.: Agrometeorology of the coffee crop. Geneva World Meteorological Organization (1994)
3. De León, C.: Enfermedades del maíz: Guía para su identificación en el campo (1974)
4. Gauhl, F., et al.: Multilocational evaluation of black Sigatoka resistance in banana and plantain. Evaluación, en varios lugares, de bananos y plátanos resistentes a la Sigatoka negra. IITA Res. Guide 47 (1993)
5. Agrios, G.N.: Fitopatología, Segunda Edición. México: UTEHA-Noriega

6. García, A., Obín, D.: Sistemas de Alerta Temprana para Prevención de Enfermedades y Plagas, December 2013
7. Wiltshire, A.: Developing early warning systems: a checklist. In: Proceedings of 3rd International Conference on Early Warning (EWC) (2006)
8. Ocharan, J.: Sistemas de Alerta Temprana. Fotografía actual y retos. Cuad. Int. Tecnol. Para El Desarro. Hum. no. 6, p. 2 (2007)
9. Flores, N.P., Lerdon, F.J., Bravo, H.R., Acuña, I.: Factibilidad De Implementar Pronosticadores Automatizados Para Controlar El Tizón Tardio De La Papa En El Sur De Chile. Agro Sur, vol. 36, no. 1, pp. 37–42, May 2008
10. Barquero Miranda, M.: Sistema de alerta temprana para el ojo de gallo. Rev. Inf. (2012)
11. Gleason, M.L., et al.: Obtaining weather data for input to crop disease-warning systems: leaf wetness duration as a case study. Sci. Agric. **65**(SPE), 76–87 (2008)
12. Bonett, M.: Personalization of web services: opportunities and challenges. Ariadne, no. 28 (2001)
13. García Gutiérrez, V.: Sistema para la adaptación de servicios a nivel de presentación y de navegación en portales web (2013)
14. González, L., Ruggia, R.: Towards dynamic adaptation within an ESB-based service infrastructure layer. In: Proceedings of the 3rd International workshop on Monitoring, Adaptation and Beyond, pp. 40–47 (2010)
15. González, L., Laborde, J.L., Galnares, M., Fenoglio, M., Ruggia, R.: An adaptive enterprise service bus infrastructure for service based systems. In: Service-Oriented Computing–ICSOC 2013 Workshops, pp. 480–491 (2013)
16. Ortiz, G., Boubeta-Puig, J., de Prado, A.G., Medina-Bulo, I.: Towards event-driven context-aware web services. In: Adaptive Web Services for Modular and Reusable Software Development: Tactics and Solutions, pp. 148–159 (2012)
17. González, L., Ortiz, G.: An event-driven integration platform for context-aware web services. J. UCS **20**(8), 1071–1088 (2014)
18. González, L., Ortiz, G.: An ESB-based infrastructure for event-driven context-aware web services. In: Advances in Service-Oriented and Cloud Computing, pp. 360–369. Springer (2013)
19. Service-Oriented Architecture (SOA) Definition. Service Architecture. http://www.service-architecture.com/articles/web-services/service-oriented_architecture_soa_definition.html. Accessed 03 Feb 2017
20. Yuan, S.-T., Lu, M.-R.: An value-centric event driven model and architecture: a case study of adaptive complement of SOA for distributed care service delivery. Expert Syst. Appl. **36**(2), 3671–3694 (2009). Part 2
21. González, L.: Plataforma ESB Adaptativa para Sistemas Basados en Servicios. Universidad de la República, Montevideo, Uruguay (2011)
22. Kazhamiakin, R.: Adaptation and monitoring in S-Cube: global vision and roadmap. In: Workshop on Service Monitoring, Adaptation and Beyond, p. 67 (2009)
23. Gaitan, A., et al.: Evento de La Niña en Colombia: Recomendaciones para la caficultura (2016)
24. Osorio, C.A.R., Giraldo, C.A.S., Ardila, M.A.C., Bustamante, A.L.G.: La Roya del Cafeto en Colombia Impacto, manejo y costos del control, February 2011
25. Corrales, D.C., Ledezma, A., Peña, A.J., Hoyos, J., Figueroa, A., Corrales, J.C.: Un nuevo conjunto de datos para la detección de roya en cultivos de café Colombianos basado en clasificadores. Sist. Telemática, vol. 12, no. 29, pp. 9–23 (2014)

26. Corrales, D.C., Figueroa, A., Ledezma, A., Corrales, J.C.: An empirical multi-classifier for coffee rust detection in Colombian crops. In: Proceedings of 15th International Conference Computational Science and Its Applications – ICCSA 2015, Banff, AB, Canada, 22–25 June 2015, Part I, vol. 9155, pp. 60–74 (2015)
27. Corrales, D.C., Casas, A.F., Ledezma, A., Corrales, J.C.: Two-level classifier ensembles for coffee rust estimation in Colombian crops. Int. J. Agric. Environ. Inf. Syst. IJAEIS **7**(3), 41–59 (2016)
28. Lasso, E., Thamada, T.T., Meira, C.A.A., Corrales, J.C.: Graph patterns as representation of rules extracted from decision trees for coffee rust detection. In: Garoufallou, E., Hartley, R.J., Gaitanou, P. (eds.) Metadata and Semantics Research, pp. 405–414. Springer International Publishing (2015)
29. Meira, C.A., Rodrigues, L.H., Moraes, S.A.: Análise da epidemia da ferrugem do cafeeiro com árvore de decisão. Trop. Plant Pathol. vol. 33, no. 2, pp. 114–124 (2008)
30. Meira, C.A.A., Rodrigues, L.H.A., de Moraes, S.A.: Modelos de alerta para o controle da ferrugem-do-cafeeiro em lavouras com alta carga pendente. Pesqui. Agropecuária Bras. vol. 44, pp. 233–242 (2009)
31. Cintra, M.E., Meira, C.A.A., Monard, M.C., Camargo, H.A., Rodrigues, L.H.A.: The use of fuzzy decision trees for coffee rust warning in Brazilian crops. In: 2011 11th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 1347–1352 (2011)
32. Ramírez, V.H., et al.: Variabilidad climática y la floración del café en Colombia (2013)
33. Dossot, D., D'Emic, J., Romero, V.: Mule in action. Manning, Greenwich (2014)
34. EsperTech - Products. http://www.espertech.com/products/. Accessed 02 Mar 2016
35. Goncalves, A.: Beginning Java EE 6 with GlassFish 3. Apress, Berkely (2010)
36. Momjian, B.: PostgreSQL: Introduction and Concepts, vol. 192. Addison-Wesley, New York (2001)
37. Liang-Jie, Z.: Web Services Research and Practices. Idea Group Inc (IGI), Hershey (2008)
38. PROCAGICA by Jacques Avelino - Research Project on ResearchGate. ResearchGate. https://www.researchgate.net/project/PROCAGICA

Crop Monitoring in High Andean Ecosystems of the Upper Basin of the Palacé River Using Planet Images

Julián Muñoz[1](✉), Leonairo Pencue[1], Apolinar Figueroa[2], and Carlos Guzmán[1]

[1] Grupo de Óptica y Láser, Universidad del Cauca, Popayán, Colombia
{julianfer, leonairo, cjguzman}@unicauca.edu.co
[2] Grupo de Estudios Ambientales, Universidad del Cauca, Popayán, Colombia
apolinar@unicauca.edu.co

Abstract. The Palacé River basin, in its upper section, comprises a lagoon complex characterized by the presence of multiple natural water reservoirs between 2900 and 3600 m above sea level. It provides water for human consumption to the municipalities of Totoró, Cajibío and the north of Popayán, an expanding urban sector supplying water to more than 200,000 people. Intensive agricultural processes, associated with the cultivation of potatoes, canola and onion, feature in the basin, and the expansion of this agricultural frontier is a fundamental aspect regarding the processes of conservation and sustainability of the paramo ecosystems therein. This paper explores the conditions and conflicts that make possible and enhance this crop expansion process, presenting an updated analysis of the coverages in the upper basin using Planet images. Classification made it possible to identify regions that affect its vegetation cover, particularly in relation to the conservation of forests and their wetlands. Classification achieved an accuracy of 96.2% and a kappa coefficient of 0.96. Establishing itself as the most up-to-date classification available for the upper part of the Palacé River basin.

Keywords: Imagery classification · Landcover · Monitoring · Planet satellites · Remote sensing · Machine learning · Combined production systems

1 Introduction

Understanding and studying the dynamics of change that occur in natural systems and especially in terrestrial cover is an important task in the monitoring and conservation of ecosystems [1]. These studies allow the identification of changes and processes in natural resources, allowing the creation of tools for environmental management schemes and spatial planning in order to ensure the proper management of the ecosystems and the production systems of settled societies. This understanding facilitates and enhances remote sensing through satellite images for monitoring, identification of conflicts, and degradation processes that require decision making in favor of conservation of the different areas under study.

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_12

Currently, multiple sources of free-access satellite imagery (Landsat, Modis, Sentinel-1 and -2) and private information [2] are collected to monitor the evolution of coverage. Planet satellites make available a database of images of the earth with daily revisit rates and high spatial resolutions. Its high temporal resolution allows the capture of images of regions where cloud cover is present for most of the year. This is the case for the upper Palacé River basin, which represents a region of immense importance for the whole of Colombia, but especially for municipalities such as Totoró, Cajibío and Popayán. Its high agricultural and cattle production cater for the Department of Cauca [3]. Geographically, it is close to the Nevado del Huila and to the volcanic chain of the Kokonukos (Pan de Azúcar and Puracé). It is a high mountain region with paramo characteristics, featuring a vegetation structure with two canopies, one of frailejón (*Espeletia hartwegiana*) and pajonales or grasslands (*Calamagrostis recta*), and other vegetative associations such as mortiño (*Vaccinium Meridionale sp*), chupaya de páramo (*Puya sp*), fern (*Blechnum sp*), and moss (*Sphagnum sp*), which together with such elements as vegetation, soil and subsoil, have developed a great potential to intercept, store and regulate surface and underground water flows. Other dominant coverages are natural forest, gallery forest and scrublands. Species making up the **natural forest** include copec (*Clusia multiflora*), encenillo blanco (*Weinmannia rollotti killip*), encenillo negro (*Weinmannia tomentosa* L.f), ahumado (*Ocotea infrafoveolata van der Weff*), siete cueros (*Tibouchina lepidota*), mortiño (*Vaccinium cf. Meridionale Sw*), cántaro (*Miconia sp*.), and ferns (*Blechum sp*). In the **gallery forest,** arborean species between two and five meters predominate, including mate cuy (*Ageratina tinifolia*), chilco (*Escallonia myrtilloides L.f.*), granizo (*Hedyosmum bonplandianum*), bodoquero (*Viburnum triphyllum*), castaño (*Tibouchina sp*), pulem (*Otholobium mexicanum*), pepo (*Solanum sp*), and salvia (*Lepechinia conferta*). In the **scrublands**, species such as chilico (*Escallonia myrtilloides*), mate cuy (*Ageratina tinifolia*), salvia (*Lepechinia conferta*), manzano (*Esperomeles sp*), angucho (*Bejaria mathewsii*), castaño (*Tibouchina sp*), cántaro (*Miconia sp*), tachuelo (*Zanthoxylum quinduense*), and coroso (*Symplocos quitensis*) predominate.

The average annual multiyear temperature in the study area ranges from 10 °C in the Gabriel López area to 8 °C in the upper part of the basin, with an annual average multiyear rainfall ranging from 1050 mm located in the district of Gabriel López at 2900 m.a.n.s.l., up to 2350 mm in the highest areas of the paramo that reach 3600 m.a. n.s.l. These conditions lead to the appearance of dense cloud cover making it difficult to monitor the area. The information provided by Planet satellites has managed to capture a clean scene of the Malvazá Valley, with which this analysis is made.

The structure of this document is organized as follows: Sect. 2 presents related work. Section 3 presents the description of the study zone, source of data, coverage analyzed and the characteristics extracted for the classification of the images. Section 4 shows the experimental results on the Malvazá Valley constructed dataset. In addition, it shows the zones vulnerable to loss of coverage and contamination of forests and wetlands present in the region. Finally, some remarks and suggestions for future work are presented.

2 Related Work

Classification of satellite images encompasses different processes in the creation and arrangement of the dataset: pre-processing of the image, radiometric correction of scenes [4, 5], extraction of spectral radiometric bands [6], selection of attributes to increase precision and reduce computational processing cost [7], and finally, a choice of method of supervised or unsupervised machine learning [8, 9] is required for the classification of the different records or samples that describe the region of interest.

Several works on image classification emerged in 2010 with an approach based on the combination of a decision tree and the theory of D-S evidence [10]. Thanks to these, an algorithm is proposed that classifies three types of coverage: vegetation, urban areas and water, the result of this research determined the land use for a Yantai zone. Other work focuses on the employment of a diversity of machine learning algorithms [11] with the purpose of comparing a decision tree, maximum likelihood classifier and support vector machine. These were tested over a dataset of Landsat images dating from 1986 to 2001. The result of this exercise yielded evidence to support the superiority of the decision tree for the classification of the dataset. In 2014 in Trentino Italy, work was carried out with Very High Resolution (VHR) multi temporal images for the detection of changes in the area of study [12].

Remote sensing provides tools for decision-making in the conservation of natural resources, ecosystem management, spatial planning, political and strategic decision-making for development, linking it to environmental and restoration processes: 2015, [13], for example, saw a review toward the application of remote sensing to support the parameterization and validation of models in ecosystem processes, analyzing ecological attributes of the forest such as composition, structure, productivity, and evapotranspiration.

In 2016, a multi-temporal and multi-sensor classification of the Palacé River basin [6] was achieved using Landsat images 4, 5 and 7 that can classify with a high precision 11 types of plant cover. Its greatest limitation lies in the images of the Landsat family of satellites which have a resolution of 30 m, avoiding the detection of crops with small areas respective to the size of the pixel; The spectral mixture between cover is greater, and its temporal resolution of 15 days does not allow recent images to be obtained.

In 2017, a multi-temporal statistical study with Landsat images in the riparian forests of the stream corridor of the upper Cauca River basin [14] allowed classification of different plant covers, generating a map of susceptibility for riparian areas in these rivers.

3 Materials and Methods

To carry out the plant cover classification process, training and test data were obtained from the Malvazá Valley (or Valle de las Papas) study area. This valley is in the upper part of the Palacé River sub-basin, in the municipality of Totoró, comprising the towns of Agua Bonita, Chuscales, Malvazá, San Pedro, Tabaco, Portachuelo, Gabriel López, and Calvache (see Fig. 1). The total analyzed area was 10910.4 ha. Plant covers taken into account were natural forest, planted forest, potato crop, secondary crops (strawberry, ullucus and broccoli), bare soil, paramo, cloud cover, urbanized areas, wetland, and grassland, defined in the CORINE project Land cover (CLC) [15].

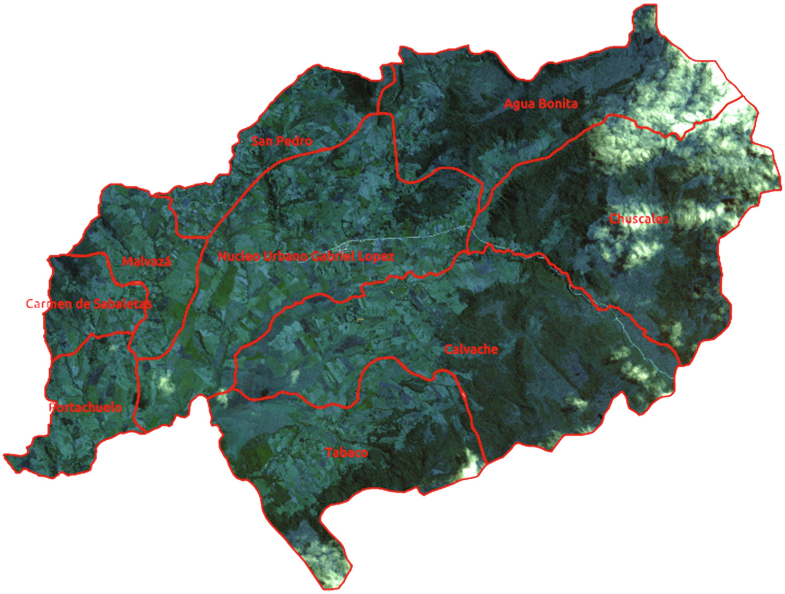


Fig. 1. Study area: The Malvazá valley, true color Planet image.

Data obtained from participatory workshops and informal interviews with villagers from Tabaco, Agua Bonita, San Pedro, Portachuelo, and Gabriel López allow the assertion that the rates of deforestation that occurred during the 1980 s and 1990 s have declined in the last years, since this practice has lost its validity, thereby avoiding the disappearance of larger areas of natural forest. The open natural forest type plant cover, comes from dense natural forest, which presents distinct levels of intervention with characteristics of a successional state of advanced secondary forest, in which some of the original natural forest structure remains dense, but with a greater predominance of herbaceous and shrub strata. The bushes correspond to a very early state of secondary forest. Nonetheless, as with natural forest, this presents similar qualities for providing wood, fuel, and food, also operating as a refuge for birds and lending protection and regulation of water flow by adding its role as a CO_2 capture zone.

In the gallery forest [14] tree vegetation has been maintained on river banks, in strips of two to ten meters wide, although in some rivers such as Cazadores and Casas Viejas there are relicts of up to 80 meters wide, which operate as stabilizers and protectors of the riverbanks, serving as a protective zone against rising water levels in periods of heavy rains, attenuating the speed and force to the currents of the rivers and streams, therefore becoming very important in flood prevention. The wetlands of Lagunas and Ciénagas occupy 21.91% of the Malvazá valley, that in Ciénaga being the bigger of the two and known commonly by locals as Laguna de Calvache, with 406.38 ha, equivalent in area to 4.26% of the valley. Laguna de Calvache is located in the center of the Malvazá valley, in an area permanently covered by a water film ranging from a few centimeters to 1.5 m in depth towards the center of the valley. It has patches of vegetation called water cushion *(Distichia sp*) and paramo ferns (*Blechnum sp*) at 3000 meters above sea level. These patches particularly appear where the water film is shallow. The Laguna has a permanent layer of water toward the center of its depression and a temporary one nearer its banks, operating as a regulator of the water flow of the Palacé River. In the rainy seasons, its borders present typical characteristics of wetlands, but these shores are seasonal, a situation exacerbated by anthropic pressure as the wetland is being severely affected through innumerable drainage channels with the purpose of rapid drainage to incorporate its land into livestock activities through the planting of grasses.

The majority of the crops comprise a number of different varieties of potato, which cover an extension of 1216 ha, corresponding to 12.77% of current occupation of the Malvazá valley. In this article, the source of the information is Planet satellite imagery. These have a spatial resolution of 3 meters and 4 spectral bands are available for each scene in radiance values (blue, green, red and nir). A radiometric correction process was carried out, taking into account the available metadata for each image [2] and converting them into reflectance values. The high temporal resolution of the Planet satellites allowed images to be obtained with a percentage of low cloudiness for the studied area. Two “Planet scenes” were needed to make up the entire Malvazá Valley. The scene IDs, their acquisition date, and cloudiness percentage are shown in Table 1.

Table 1. Identifiers and acquisition dates of Planet scenes for the Malvazá Valley

Planet image ID	Acquisition date	Percent cloudiness (%)
20170716_144315_1041	July 16, 2017	10
20170716_144316_1041	July 16, 2017	16

The number of samples extracted for each class, using photo-interpreted points [6] and field visits to the study region are presented in Table 2.

The total of collected samples was of 116,890 registries, for the training of the classification process a Random Forest (RF) algorithm was used, available in the WEKA library [16]. The default parameters of this RF-related data mining library were used in this research. The training and test sets were constructed considering the 70/30 division, that is, 81823 training samples and 35067 testing samples.

Each sample or record of the dataset consists of 13 characteristics: spectral bands in reflectance values of Planet satellites: Blue, Green, Red and NIR; Indices of vegetation [6] built from the spectral bands: ARVI, CRI, EVI, NDVI, NDWI, SAVI, SIPI, SR and TVI. The final constructed model enabled classification of the Malvazá valley; the classified images were exported in.tif format with their respective georeference, using the OpenCV and GDAL libraries for this task.

Table 2. Characterization of classified covers.

Cover	Number of samples	Tag assigned by class
Clouds	11724	0
Water Bodies	11723	1
Urban areas	11724	2
Paramos	11724	3
Bare soil	11724	4
Potato crops	9750	5
Grasslands	11724	6
Natural forests	11725	7
Planted forests	11724	8
Secondary crops	1625	9
Wetlands	11723	10

4 Results

The final classification process is based on the Random Forest algorithm. Its implementation was done through the use of WEKA, a library specialized in data mining. The algorithm parameters were configured using the default values tuned to WEKA. A first experiment was carried out that consisted in performing a classification taking into account all the characteristics. The general accuracy of the classifier was 96.1% with a kappa coeficient of 0.95. The correctly classified instances were 33729 and the mean square error 0.0734. The confusion matrix is shown in Table 3.

To improve the classification capacity, reduce the computational cost of training a dataset with 13 characteristics and 81823 records, and later classify large areas with high resolution images a process of selection of characteristics is performed. Such process is available in the options of WEKA: select attributes. It consists of a wrapper based on a random forest and cross validation to choose the best training precision based on a set of characteristics. The search method is based on the greedy type algorithm called GreedyStepwise. The selection of attributes returned a list of characteristics with a percentage equivalent to their level of importance in the classification process, see Table 4.

Selecting the attributes with a percentage greater than 70% of importance, the original dataset is shortened to 7 characteristics: blue band, green band, red band, ARVI, CRI, NDWI and TVI. A new classification was performed with this dataset resulting in a Kappa coeficient of 0.96 and an overall precision of 96.20%, correctly classified instances increased to 33736 and the mean squared error was reduced to 0.0733. Therefore, better results are obtained with a lower computational cost. The confusion matrix of the classification process is shown in Table 5.

In general terms, the classification process is satisfactory, as evidenced by the kappa coefficient and general accuracy. It is worth noting that the classifier achieves its highest precision for the labels of water bodies (class 1) and planted forest (class 8); its accuracy decreases for secondary crops (class 9). These crops have a small area so it was not possible to obtain more samples which led to a decrease in accuracy. In

Table 3. Confusion matrix for all features initially built.

True classes	Classified as											Total	Precision (%)
	0	1	2	3	4	5	6	7	8	9	10		
0	3550	0	4	0	34	0	1	0	0	0	0	3589	97.1
1	0	3482	0	0	5	0	0	0	0	0	0	3487	99.9
2	43	0	3377	5	79	10	9	0	0	0	2	3525	98.4
3	0	0	1	3529	18	7	0	7	0	5	3	3570	93.4
4	63	2	38	116	3137	15	53	8	1	4	61	3498	93.8
5	0	0	10	81	18	2600	26	35	0	66	111	2947	92.0
6	1	0	1	0	35	5	3366	0	0	0	0	3408	97.3
7	0	0	0	7	1	7	0	3525	0	5	5	3550	96.7
8	0	0	0	0	0	0	0	0	3504	0	0	3504	100
9	0	0	1	30	5	124	4	64	0	267	9	504	76.5
10	0	0	0	10	14	59	0	8	0	2	3392	3485	94.7
Total	3657	3484	3428	3778	3346	2827	3459	3647	3505	344	3583	35067	96.1
Recall (%)	98.9	99.9	95.8	98.9	89.7	88.2	98.8	99.3	100	53.0	97.3	96.2	96.1

Kappa 0.95

Table 4. Importance of variables during the classification process

Variables	blue	green	red	nir	arvi	cri	evi	ndvi	ndwi	savi	sipi	sr	tvi
Importance (%)	100	100	100	50	100	100	40	20	60	40	50	20	70

addition, due to these areas, small footprint spectral mixtures are produced, generating ambiguities in the spectral response of such plant covers forming areas difficult to classify by the algorithm.

Regarding recall, it can be observed that for the potato crop and secondary crop (class 5 and class 9 respectively) labels a percentage lower than 90% is observable, this is explained by the fact that in crops the dynamics of change is quite strong, which means that the cultivated area is sometimes reflected as bare soils (planting season) and grasslands (growing crops). The confusion matrix shows that for the secondary crops the highest instances incorrectly classified are in potato crops and to a lesser extent in bare soils, paramo and wetlands.

Once the model based on the Random Forest algorithm was trained, it was proceeded to classify the trails that make up the Malvazá valley (see Fig. 2). The objective of this procedure was to verify the location and area of the plant cover, mainly studying potato crops because they represent the main economic activity in the region and the main source of employment. In addition, natural forests and wetlands were verified. These are the water regulators of the region, with a substantial impact on the quality of life of the municipalities of Totoró, Cajibío and the north of Popayan.

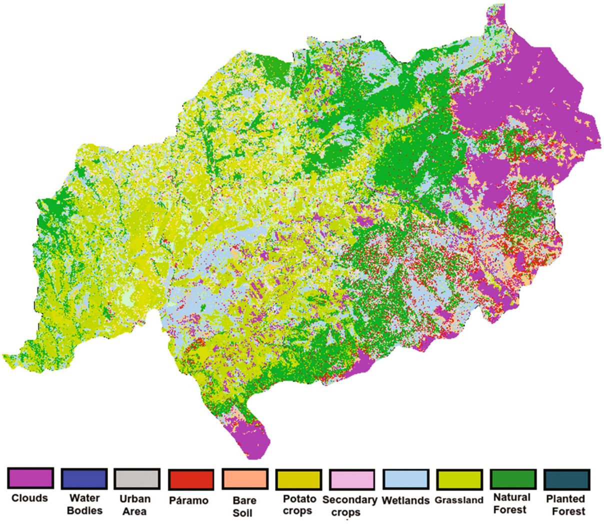


Fig. 2. Classification of plant cover in the Malvazá valley.

A supervised field test was carried out to corroborate the classified images in the study area. For this purpose, photographs of the areas of interest were taken. These were traced, the polygons and geo-referenced points of the corroborated plant covers were drawn for comparison, and verification done with the image of Fig. 2. In Fig. 3, georeferenced points X (2.5122, −76.2870), and Y (2.5065, −76.2849) were located to indicate the place of capture of the photograph. The results are shown below:

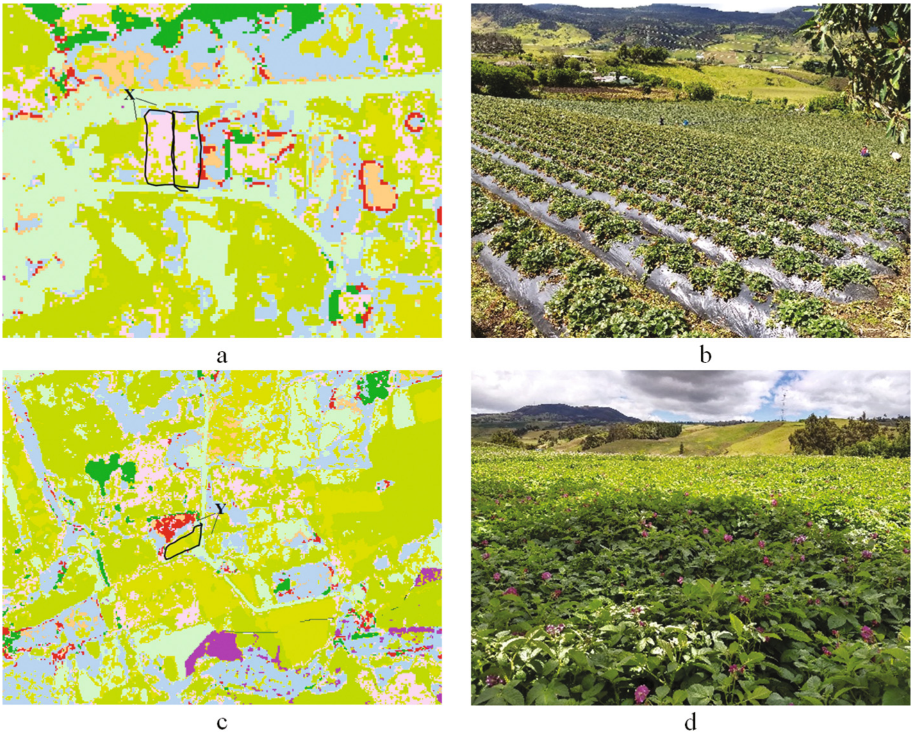


Fig. 3. Verification of the classification with the supervised check, **a** and **c** regions visited with a closure on strawberry and potato crops; **b** and **d** pictures of the area.

The visual inspection corroborates that the classification had a high precision. It is possible to verify how the crops shown in the photographs are detected and also how the plant covers that are observed at the bottom of the landscape also coincide.

In this research, the result obtained in the classification enabled statistical analyzes to be carried out related to the area of occupancy of the cover. Table 6 summarizes the above.

The results show that the three covers that predominate in the Malvazá valley are forests, wetlands and grasslands. These percentages match because in the studied municipalities, livestock is extensive and dedicated mainly to beef and dairy cattle. Additionally, the zone has paramo conditions at heights above 2900 m.a.n.s.l., becoming a very humid region with 213 to 298 days of precipitation per year.

It is important to note that the accumulated percentage of forests and wetlands (47.55% of the actual area analyzed) is comparable to the sum of cover crops such as potato crops, secondary crops, bare soil and pastures (45.13% of the actual area analyzed). This shows that the upper part of the basin presents a high degree of environmental impact caused by the combined effect of intensive production systems (beef

Table 5. Confusion matrix for the dataset with feature selection.

True Classes	Classified as											Total	Precision (%)
	0	1	2	3	4	5	6	7	8	9	10		
0	3551	0	6	0	31	0	1	0	0	0	0	3589	97.2
1	0	3483	0	0	4	0	0	0	0	0	0	3487	99.9
2	42	0	3386	5	74	8	8	0	0	0	2	3525	98.6
3	0	0	0	3527	18	9	0	7	0	3	6	3570	93.4
4	58	3	31	115	3150	13	56	4	2	5	61	3498	93.9
5	0	0	8	80	18	2611	24	38	0	64	104	2947	91.8
6	1	0	1	0	41	5	3360	0	0	0	0	3408	97.3
7	0	0	0	7	0	8	0	3518	0	9	8	3550	96.6
8	0	0	0	0	1	0	0	0	3503	0	0	3504	99.9
9	0	0	1	32	7	122	3	64	0	267	8	504	76.3
10	0	0	1	10	12	69	0	11	0	2	3380	3485	94.7
Total	3652	3486	3434	3776	3356	2845	3452	3642	3505	350	3569	35067	96.1
Recall(%)	98.9	99.9	96.1	98.8	90.1	88.6	98.6	99.1	100.0	53.0	97.0	96.2	96.2

Kappa 0.96

Table 6. Area of occupation of the cover over the Malvazá valley.

Cover	Area (ha)	Percentage of current occupancy (%)
Clouds	1389.05	NA
Water bodies	0.00	0.00
Urban areas	289.37	3.04
Paramos	407.60	4.28
Bare soil	578.36	6.07
Potato crops	1216.00	12.77
Grasslands	1928.61	20.25
Natural forest	2442.79	25.64
Planted forest	0.00	0.00
Secondary crops	575.40	6.04
Wetlands	2087.33	21.91
Total	10914.50	100.00
Effective area	9525.48	

and dairy cattle) that are handled in the region, affecting soils with erosive processes, increasing turbidity in the water and favoring the opening of the natural forest for agricultural activities.

The Malvazá valley is a territory recognized for its high agricultural yield, potato crops being the main source of income in the region; large swathes of land are cultivated either by their owners or by people who rent the land. This has led to the spread of agriculture throughout the region. The attained results of the classification allow the identification of areas where covers such as wetlands and natural forests, the main contributors to the conservation of ecosystems and water resources, are vulnerable to deterioration (See Fig. 4).

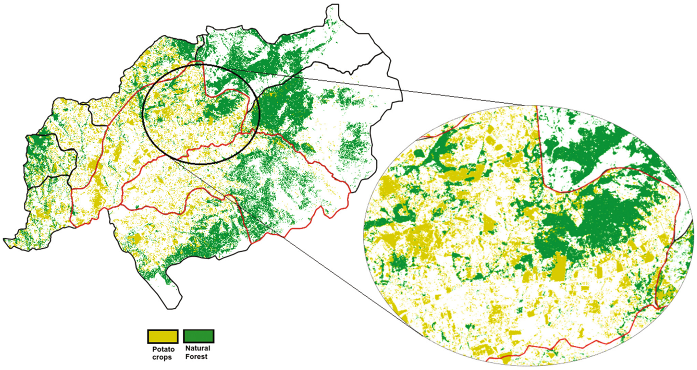


Fig. 4. Potato crops and natural forests. Zones of vulnerability in the region.

In Fig. 4, it is observed how the potato crops are located on the limits of the natural forest, which extends the agricultural frontier by the pressure exerted in the expansion of the sowing area. This circumstance is due to slash and burn activities carried out as conditioning for agricultural processes, leading to the reduction of the forest area. Natural forest cover, to a large degree, is private property, a condition that limits protection activities, since land renting is considered a profitable activity above the conservation of the forest.

In Fig. 5, a point Z (2.5145, −76.2773) is located, in which a photograph is taken showing how a potato crop zone joins the edge of the beginning of the natural forest. The classification clearly details this situation.

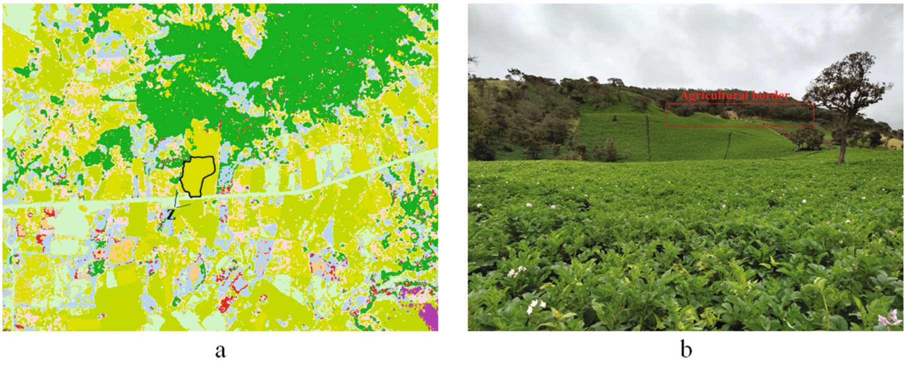


Fig. 5. (a) Classification of a region of the Malvazá valley, (b) agricultural border near a natural forest.

The same process described above is present in the wetland (see Fig. 6). The growers drain the wetland through canals, with the purpose of gaining arable lands and incorporating them into the agricultural-livestock activity. Because of the intensive use

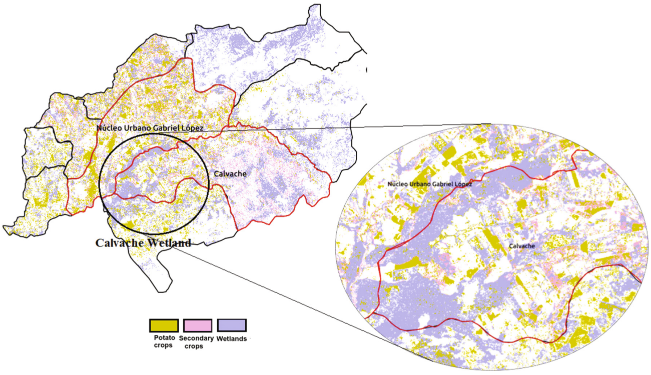


Fig. 6. Interaction of wetlands in the Malvazá valley against potato and secondary crops.

of agrochemicals in agricultural activity, this condition favors the deterioration of water and soils. These activities, which have been carried out throughout the last 30 years, have deteriorated the Calvache wetland, located between the boundaries of the Calvache and Gabriel Lopez municipalities, affecting the availability and regulation of the water for the Palacé River. Potato crops here yield approximately 40 to 50 tons per hectare and information acquired in the workshops with the community in this investigation determined that each hectare of cultivated potato requires 10 bags of $CaCO_3$, 20 bags of NPK (10:30:10), 20 bags of NPK (10:20:20), and additionally for each production cycle a cocktail of up to 12 agrochemicals with different levels of toxicity. This is applied, making use of up to 200 cm^3 per product, diluted in 55,000 L of water, to be applied on one hectare. Figure 6 shows the current state of the wetland.

Comparing the wetland detected by the classification algorithm and a study carried out on the Malvazá valley in 2006, a modification of the form or limits corresponding

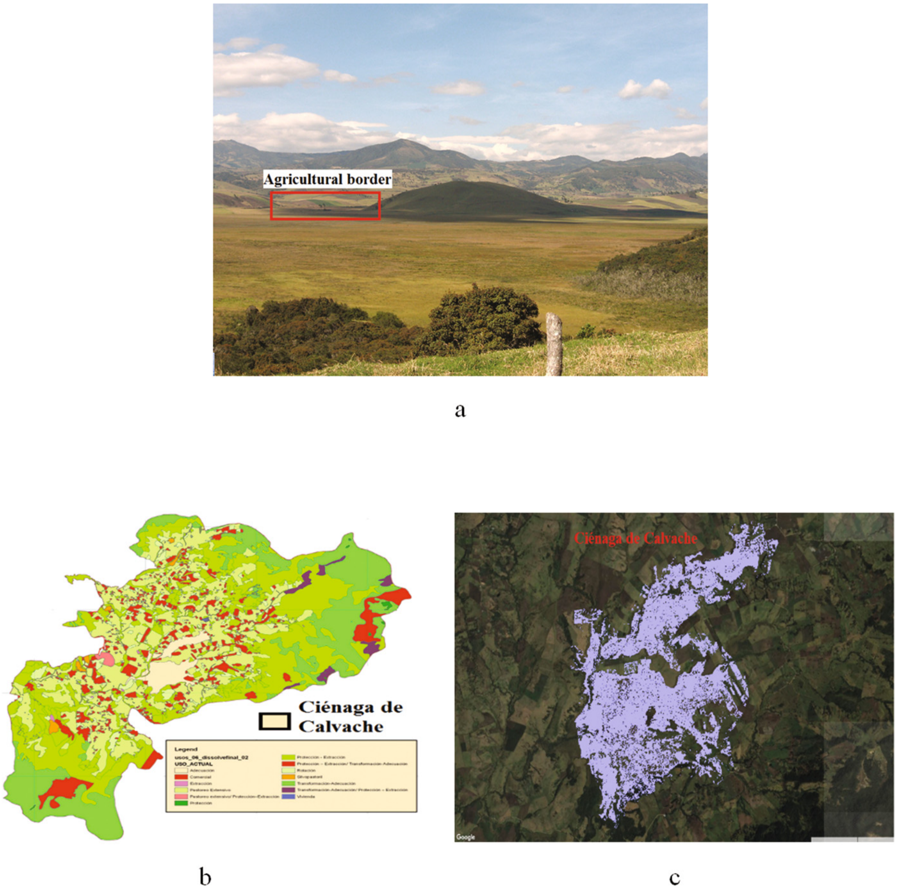


Fig. 7. Modification of limits of Calvache Ciénaga between years 2006 and 2017. (a) Photograph of Calvache Ciénaga. (b) Supervised classification of land use year 2006. (c) Detection of Calvache Ciénaga through the classification made in this research

to the Calvache wetland (see Fig. 7) is noticeable. In addition, it is observed that around the wetland, potato and strawberry crops appear, generating pollution hazards. The wetland is vulnerable to threats of pollution generated by the proximity of crops and agrochemicals in the region. In general terms, in the upper part of the Palacé River basin (Malvazá valley) the wetlands have some closeness to crops, except in the Agua Bonita municipality, in which case the classification of coverings did not report the occurrence of crops.

5 Conclusions and Future Work

A classification process based on the Random Forest algorithm and using Planet satellite imagery was performed. The overall accuracy was 96.1%, with a recall rate of 96.2%, indicating that confusions between classified coverages were very low. The value of the kappa coefficient was 0.96, obtaining precision rates for each coverage above 91% in 10 of the 11 classes evaluated. In this investigation, a technical visit to the study area was undertaken. The results showed the coincidence between the coverages observed in the region and the classified areas. The work carried out in this research in the Malvazá valley represents the latest classification to date. This region is characterized by parametric conditions, having a favorable climate for the sowing of crops (potato, ullucus, broccoli and strawberry) and the practice of livestock (beef and dairy cattle) activities; additionally, the capacity to supply water to the Palacé River, responsible for providing water for human consumption in the municipalities of Totoró, Cajibío and the north of Popayán. An environmental impact assessment in change of covers such as forest and wetlands was therefore carried out. The analysis showed that crop sowing is being carried out near these covers, which leads to loss of area, affected ecosystem services, and contamination of their ecosystems due to agrochemicals used in agricultural activities.

As a future work, the authors plan to enrich the training data set, especially the secondary crop class, with the aim of balancing the dataset and improving the power of the classification. The classification will furthermore be extended over the whole of the Palacé river basin and in general in the upper Cauca basin, which is greatly significant for Colombia in terms of agriculture and water resources, and will also detect the regions in the basin with greatest impact in plant cover, fundamental to the conservation of water resources and ecosystems.

Acknowledgments. This work was developed within the "Programa para el fortalecimiento de la Red Interinstitucional de Cambio Climático y Seguridad Alimentaria" project – RICCLISA, funded by Colciencias and the University of Cauca. The authors acknowledge the support of the University of Cauca to carry out this search. We are especially grateful to Colin McLachlan for suggestions relating to the English text.

References

1. Onur, A.C., Tezer, A.: Ecosystem services based spatial planning decision making for adaptation to climate changes. Habitat Int. **47**, 267–278 (2015)
2. Planet—Monitoring. https://www.planet.com/products/monitoring/. Accessed 24 July 2017
3. Fernandez, J., Guevara, J., Cáceres, C., Mazabuel, C., Lopera, D., Becerra, C., Chaux, C., Guevara, R., Mora, L., Borrero, A., Figueroa, E.: Plan de ordenacion y manejo de la parte alta de la subcuenca hidrografica del río palacé, Popayán (2008)
4. Richards, J.A.: Sources and characteristics of remote sensing image data. In: Remote Sensing Digital Image Analysis, pp. 1–26. Springer, Heidelberg (2013)
5. Richards, J.A.: Correcting and Registering Images, pp. 27–77. In: Remote Sensing Digital Image Analysis. Springer, Heidelberg (2013)
6. Pencue-Fierro, E.L., Solano-Correa, Y.T., Corrales-Munoz, J.C., Figueroa-Casas, A.: A semi-supervised hybrid approach for multitemporal multi-region multisensor landsat data classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **9**(12), 5424–5435 (2016)
7. Richards, J.A.: Feature reduction. In: Remote Sensing Digital Image Analysis, pp. 343–380. Springer, Heidelberg (2013)
8. Richards, J.A.: Supervised classification techniques. In: Remote Sensing Digital Image Analysis, pp. 247–318. Springer, Heidelberg (2013)
9. Richards, J.A.: Clustering and unsupervised classification. In: Remote Sensing Digital Image Analysis, pp. 319–341. Springer, Heidelberg (2013)
10. Li, X., Xing, Q., Kang, L.: Remote sensing image classification method based on evidence theory and decision tree, p. 78570Y (2010)
11. Otukei, J.R., Blaschke, T.: Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. Int. J. Appl. Earth Obs. Geoinf. **12**, S27–S31 (2010)
12. Correa, Y.T.S., Bovolo, F., Bruzzone, L.: Change detection in very high resolution multisensor optical images, p. 924410 (2014)
13. Masek, J.G., Hayes, D.J., Hughes, M.J., Healey, S.P., Turner, D.P., Masek, J.G., Hayes, D.J., Hughes, M.J., Healey, S.P., Turner, D.P.: The role of remote sensing in process-scaling studies of managed forest ecosystems, vol. 355, pp. 109–123
14. Pencue-Fierro, E.L., Solano-Correa, Y.T., Corrales-Muñoz, J.C., Figueroa-Casas, A.: Analysis of riparian forest buffers dynamics in Colombian basins by landsat time series. In: 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp), pp. 1–4 (2017)
15. Feranec, J., Hazeu, G., Christensen, S., Jaffrain, G.: Corine land cover change detection in Europe (case studies of the Netherlands and Slovakia). Land Use Pol. **24**(1), 234–247 (2007)
16. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software. ACM SIGKDD Explor. Newsl. **11**(1), 10 (2009)

WSN Prototype for African Oil Palm Bud Rot Monitoring

Miguel Piamonte[1], Monica Huerta[2(✉)], Roger Clotet[3], John Padilla[4], Tito Vargas[1], and David Rivas[5]

[1] Universidad Santo Tomás, Bucaramanga, Colombia
miguelpiamonte@gmail.com, tivarher@ustabuca.edu.co
[2] Universidad Politécnica Salesiana, Cuenca, Ecuador
mhuerta@ups.edu.ec
[3] Universidad Simón Bolívar, Caracas, Venezuela
clotet@usb.ve
[4] Universidad Pontificia Bolivariana, Bucaramanga, Colombia
jhon.padilla@upb.edu.co
[5] Universidad de las Fuerzas Armadas - ESPE, Sangolquí, Ecuador
drrivas@espe.edu.ec

Abstract. The oil palm (Elaeis guineensis) is the most productive oleaginous on the planet. The world's largest producers of oil palm are located in Asia, Colombia is the fourth largest producer in the world and the first in America. In recent years, the sowing of oil palm has taken a great importance in food industry and biofuel production. Bud rot is among the factors that are most affecting this type of crop, generating to palm farmers large economic losses and the country's social problems due to unemployment. Early detection of abiotic factors that may trigger bud rot is one of the strategies that would allow palm farmers to minimize the impact on the crops. In this research, a WSN was developed to acquire, process and transmit in real time to a server acquired data as: pH, humidity, temperature and luminosity.

Keywords: WSN · Oil palm · Bud rot

1 Introduction

The oil palm (Elaeis guineensis) is the most productive oleaginous on the planet, it has a productive life that exceeds 25 years. The world's largest producers of oil palm are located in Asia, the first producer is Indonesia followed by Malaysia, Colombia is the fourth largest producer of oil palm in the world and the first in America [1].

In recent years, the sowing of oil palm has taken a great importance in food industry and biofuel production. The African oil palm represents an important crop in the Colombian economy. This sector has been hit hard by the bud rot disease, that reaches the crops in a biotic or abiotic way.

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_13

At present, in Colombia, African oil palm plantations represent an important part of the country's agricultural production, with a total of 450,131 hectares planted. However, bud rot is among the factors that are most affecting this type of crop, generating to palm farmers large economic losses and country's social problems due to unemployment.

There are two hypotheses about bud rot: the first one has a biotic origin, as a classical disease caused by a specific species of bacteria or fungus, sometimes with complications by several species of insects [7]. The second is of abiotic origin generally related to climatic and edaphic factors as causes of a very complex nutritional disorder but without precise specifications in causes [7].

Palm farmers have the risk that their main source of subsistence will be extinguished because of African oil palm bud rot. For this reason, it is necessary to deploy crop monitoring systems based on wireless sensor network that allows an early warning of diseases.

Different management strategies of bud rot disease have been designed: chemical control, elimination of affected tissue, planting of tolerant materials and edafoclimatics analysis, among others. Early detection of abiotic factors that may trigger bud rot is one of the strategies that would allow palm farmers to minimize the impact on the crop. This is possible if with the design and development of a system to monitoring crops using wireless sensor networks [9].

Wireless sensor networks (WSN) are composed of nodes, which are autonomous devices capable of capturing and transmitting different variables that are around the target to be monitored, such as temperature, humidity, pH and luminosity. Different researchers have used WSN to monitor the climatic variables of different crops such as: Aji Habanero [3], tomato [6], Orchids [5] and roses [4], among others. Their results indicate that they have been able to optimize the production of these crops by applying ICT (Information and Communication Technologies). When WSN have to be deployed outdoor or remote places, in many cases, they do not have access to the electrical system and data networks. In addition, it have to handle a large volume of information, therefore, the design of reliable nodes it is necessary, so Palm farmers can monitor crops in real time.

In Colombia, most crops are handled in an artisan way. Precision agriculture has begun to have importance in large business groups, this type of agriculture is going to generate a large amount of information, which is why it is necessary to use tools to process them with Big Data analysis techniques.

In this research, a nodes and the WSN have been designed and developed, thus allowing to transmit the acquired values of different sensors that are connected to the nodes. Each node of the WSN it is a unit capable of acquire, process and transmit, in real time, to a server data related to important variables to determined if there is a bud rot triggered event. Variables such as: pH, humidity, temperature and luminosity. Additionally, nodes of WSN system is equipped with power supply and geo-referencing system. The acquired data make possible to create a DataSet to be processed with Big Data analysis tools.

2 System Design

2.1 Edafoclimatic Requirements of African Palm

The African palm is a tropical region plant, therefore, it is located in those areas with average monthly temperatures that oscillate between 26 C and 28 C. Temperatures below 17 C for several days cause a reduction in the development of adult plants and in the seed nursery it could stops the growth of the seedlings. As for rainfall, the favorable conditions for this species are determined by the amount and distribution of rainfall, which ranges from 1800 mm to 2300 mm per year.

In relation to the sunlight, the African palm is identified as heliophile plant, because its high requirements of sunlight. To achieve high productions, 1500 h of sunlight per year are required. Therefore, areas with average monthly values greater than 125 h of sunlight are considered suitable for the cultivation of this plant. The insolation also affects the emission of the inflorescences, photosynthesis, ripening of the clusters and oil content of the mesocarp. As for relative humidity, a monthly average of more than 75% is required.

It tolerates moderately acidic soils (pH 5,5–6,5), although these generally have nutrient deficiencies. In light soils, from sandy to loamy sandy, there are problems of washing and leaching of nutrients, so that their consistency is insufficient for the support of the plant. The heavy soils, of clay texture, have limitations for their handling, the difficulty to drain them and ease with which they are compacted.

Therefore, the optimal soils for African palm cultivation are deep soils with good drainage, slightly clayey texture, good organic matter content, flat to slightly undulating topography with slopes lower than 2% and with a level of fertility from medium to high.

2.2 Bud Rot

Bud rot (BR) is one of most present diseases in oil palm plantations, this disease has been responsible for large economic losses and has generated social instability. This disease has been investigated in all Latin America by different institutes, the investigation results coincide in that it has biotic and abiotic origins.

The symptoms show the destruction of the young arrows, without presenting damage to meristematic area in initial stages of the disease; That is, meristematic area is the point of growth of the plant, if this is affected the plant dies. Experts indicate that when the disease is detected in time, it is sufficient to prune the affected young arrow together with a chemical control may be sufficient to control the disease. But if the attack is severe there is destruction of the arrows and the meristematic area, consequently, the emission and maturation of the new arrows is stopped, causing the death of the plant.

There is no clarity about the etiology of the disease and has been related to several abiotic factors of a physiological nature and biotic factors of a pathogenic nature. Preliminary studies have reported that no nutrients are associated with the incidence of BR, although it is possible that nutrient imbalance in the soil, together with excess water, may become predisposing factors to the disease.

2.3 Hardware Design

Figure 1 shows the block diagram of the designed node: Data Processing, Sensors, GPS, Transmission and Energy.

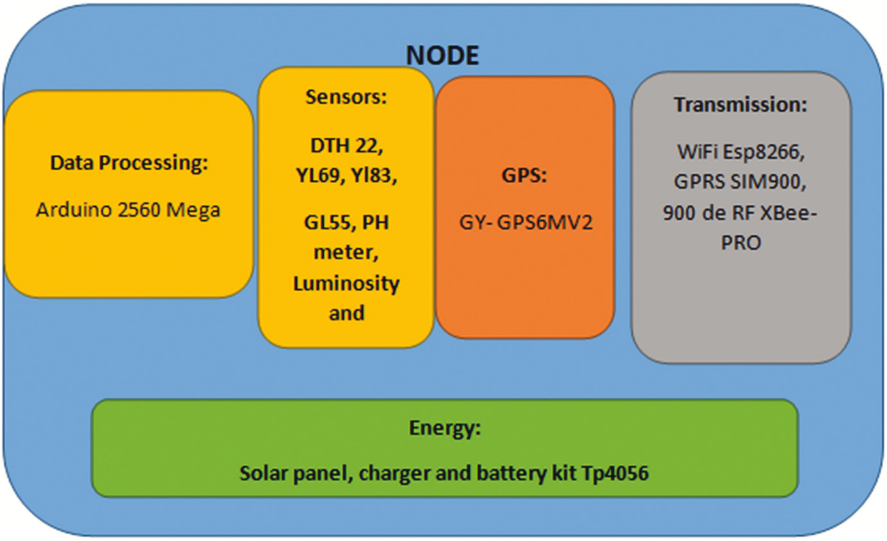


Fig. 1. Designed node

Data Processing: To perform data processing, the Mega Arduino has been chosen. This device has the following characteristics: 54 digital input/output pins, 14 can be used as analog PWM outputs, 16 analog inputs and 4 TTL-UART series receivers/transmitters. It has a 256 Kbytes flash memory, 8 SRAM memory and 4 KB EPROM.

Sensors: The sensors have been chosen according to the variable to be measured. Table 1 describes their main characteristics:

All these sensors are compatible with Arduino, in the case of pH measurement it is done using a tester that sends to Arduino the analog input of acquired pH value.

A GY-GPS6MV2 easy to program module has been used, designed to be compatible with Arduino and integrated as a module to the system.

Table 1. Sensors to be used to measure the variables

Variable	Sensor
Air humidity	DTH 22
Soil moisture	YL69
Temperature	DTH22
GPS	GY-GPS6MV2
pH	Tester pH, temperature and luminosity
Luminosity	GL55

Transmission: Three types of communication technologies are used: GSM, WiFi and Xbee. GSM: When some nodes are out of range, because of the topography or the existing distances in the terrain, in this project GSM SIM900 model has been used. WiFi: In cases where nodes are near an access point (wireless router or another), wireless technology is the most economic form of connectivity. Here, the Wi-Fi module ESP8266 has been used. Xbee: Preferred transmission in this type of networks by its low power consumption vs. data transmission.

Energy: The node it is designed to use the charger and battery kit tp4056, a 5 v 600 ma solar panel and it is possible to connect it to the electrical grid network.

2.4 Software Design

The software has been developed using LAMP architecture, which is a solution of great performance and availability. It has great advantages since it has complete on-line documentation and is a open source way to be able to develop Big Data techniques desired to be implemented.

Operating System: Ubuntu server 12.04.4. Web Server: Apache Database: MySQL Web Programing language: PHP

Table 2 shows the decomposition of the data into four arrays in order to calculate the average value of the humidity in each of the hours.

2.5 Performed Tests

First, the Arduino 2560 has been set up with dth22, YL-69 and YL-38 sensors in conjunction with an Ethernet module. For 24 h data has been registered every 30 s, the system worked without any problem showing the variations expected by the day/night change.

Secondly, the Arduino 2560 has been configured with dth 22, ph (tester), brightness, YL-69, YL-38, ESP8266 WiFi module and GPS module. Also for 24 h the system functions correctly. With an approximate consumption of 700mA. Lime dissolved in water has been applied at 4 h of test start point and the expected pH change has been detected properly. It has been verified that the

Table 2. Database table of edafoclimatic variables and information fields

Array 1	Array 2
Data Humidity soil from 08:00 to 14:00	Data Humidity soil from 14:00 to 18:00
[	[
Hum_S8 D1 - Hum_S14 D1	Hum_S14 D1 - Hum_S18 D1
Hum_S8 D2 - Hum_S14 D1	Hum_S14 D2 - Hum_S18 D1
Hum_S8 D3 - Hum_S14 D1	Hum_S14 D3 - Hum_S18 D1
.	.
.	.
]	]
Array 3	**Array 4**
Data Humidity soil from 18:00 to 01:00	Data Humidity soil from de 01:00 to 08:00
[	[
Hum_S18 D1 - Hum_S1 D1	Hum_S1 D1 - Hum_S1 D1
Hum_S18 D2 - Hum_S1 D1	Hum_S1 D2 - Hum_S8 D1
Hum_S18 D3 - Hum_S1 D1	Hum_S1 D3 - Hum_S8 D1
.	.
.	.
]	]

GPS module gives the correct position with a margin of error of 20 meters. No data losses has been detected and the corresponding database entry has been recorded for all submissions.

Thirdly, the Arduino 2560 has been configured with dth 22, ph (tester), luminance, YL-69, YL-38, GSM/GPRS SIM900 module and GPS module. Using the TELEBUCARAMANGA operator and configuring SMS sending mechanism every hour. For 24 h the test has been done obtaining a 100% reception of the data.

Fourth, two nodes have been configured. Node1 (emitter): Arduino 2560 with dth 22, ph (tester), luminance, YL-69, YL-38, GPS module and XBee pro modules. Node 2 (receiver): Arduino 2560 with Ethernet and XBee pro modules. For 24 h the test has been performed without observing data losses. The power consumption of both nodes has been reported to be approximately the same, about 600 mA.

Finally, the installation of the software (LAMP) has been performed on a Raspberry PI 3 (in the other tests it was installed on a laptop). In this case the node had: Arduino 2560 with dth 22, ph (tester), luminance, YL-69, YL-38, GSM/GPRS SIM900 module and GPS module. For 24 h data has been taken, data loss has been observed and it has been found that access to it is slow because performance limitations of Raspberrry. Also the approximate energy consumptions has been calculated obtaining 600mA for the Arduino and 2.5 A for the Raspberry.

3 Results

3.1 A Single Transmitting Node

The data acquisition was performed 4 times a day, at following times 8 am, 2 pm, 6 pm and 1 am; Since in these hours is when data have a significant variation. A soil sample from an African oil palm crop was analyzed where bud rot was not present.

In the database a table was created to store edafoclimatic variables and three information fields: Control ID, the number of node that is transmitting and timestamp.

3.2 Two Nodes Transmitting

Figure 2 shows the connection scheme used in two nodes that performed the acquisition of edafoclimatic variables, without presence of bud rot (Node 1) and presence of bud rot (Node 2).

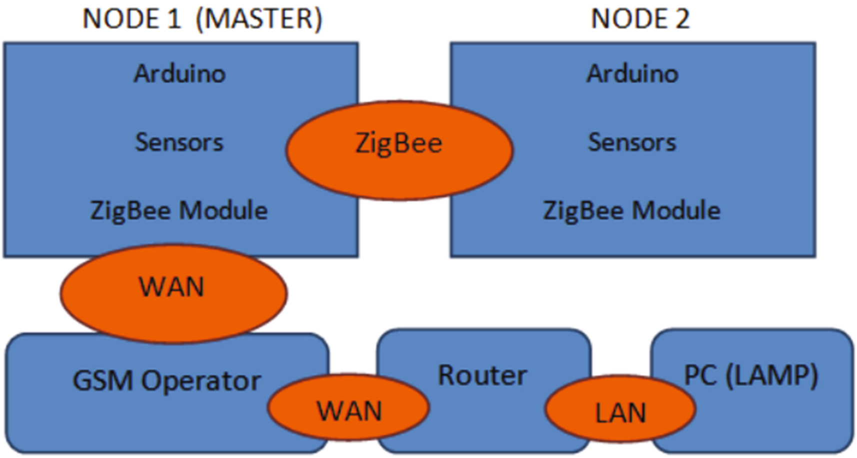


Fig. 2. Connection scheme

3.3 Tests Made with Soil of a Crop Without Presence of Bud Rot

The acquisition was performed 4 times a day, at 8 am, 2 pm, 6 pm and 1 am, Since in these hours is when data have a significant variation. The transmission of acquired data will be done 3 times a day. The acquisition at 1 am will be transmitted, to minimize the energy consumption, along with the data acquired at 8 am.

The nodes were interconnected with ZigBee protocol. One of them acted as master and sent data to the server through GSM network of CLARO operator.

Tests Performed on Crops Without Bud Rot Transmitted by Node 1. Figure 3 shown that the soil where palms that does not suffer from bud rot (BR) have pH that is neither alkaline nor acidic soils, in addition it can be observed that there is no significant variation of pH.

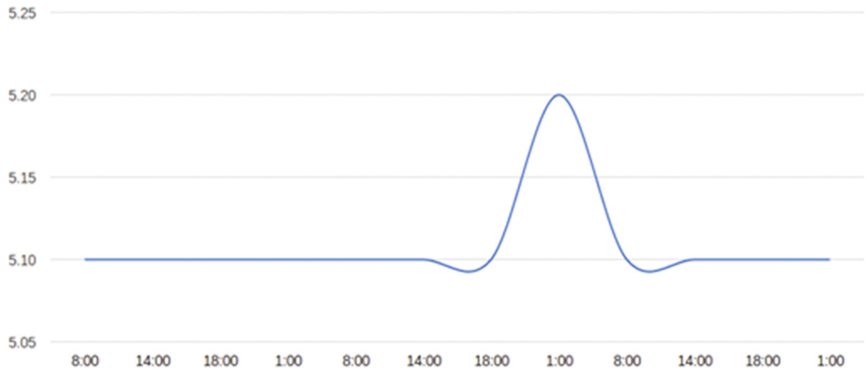


Fig. 3. Soil pH variation of crop without bud rot [pH/s]

The results corresponding to luminosity are shown in Fig. 4 You can see day and night cycle.

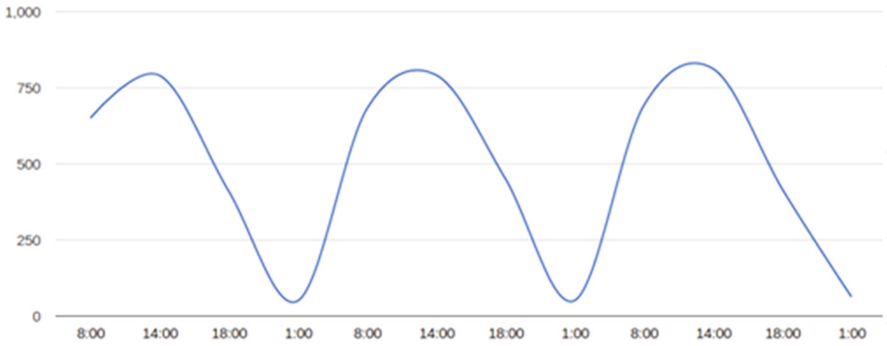


Fig. 4. Luminosity variation of crop without bud rot [lux/s]

Figure 5 shows temperature changes that occur during the 72 h of the evaluation in a soil sample without but rot presence.

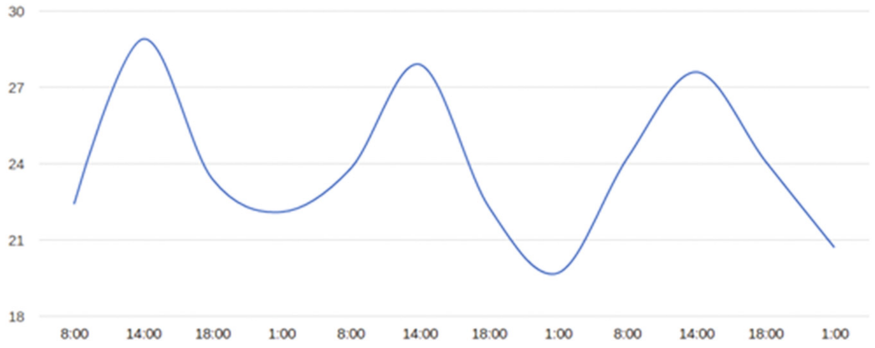


Fig. 5. Temperature variation of crop without bud rot [°C/s]

Figure 6 shows ambient humidity taken by the node installed on the ground that does not have bud rot presence.

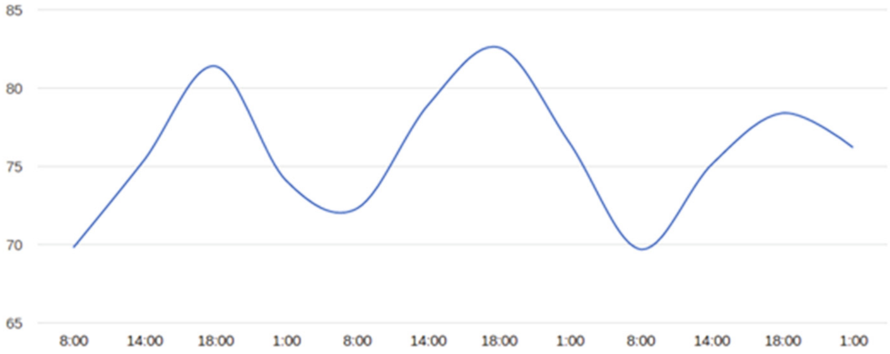


Fig. 6. Humidity variation of crop without bud rot [%/s]

Tests Performed on Crops with Bud Rot. The next node is installed in a container containing soil taken from a crop having bud rot.

In order to verify that there was a variation of pH in soil with bud rot presence, twenty-four (24) h after the transmission began, an application of lime dissolved in water was performed.

In Fig. 7 it is possible to observe ambient temperature which was measured during a period of 72 h.

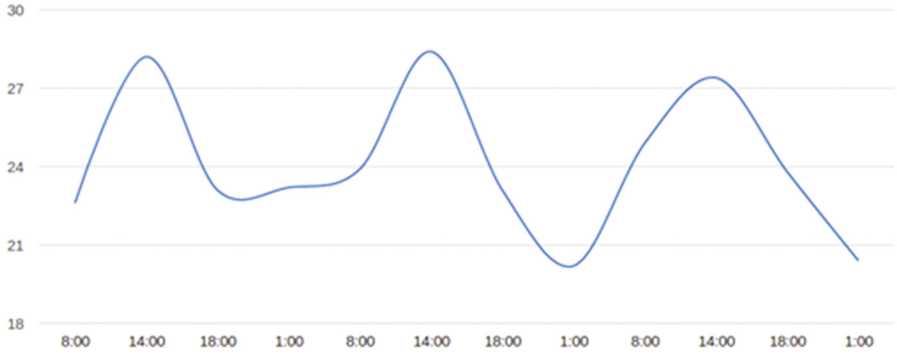


Fig. 7. Temperature variation of crop with bud rot [°C/s]

Figure 8 shows the behavior of the pH. As can be observed the soil pH is acid, it is transformed to neutral pH soil using lime.

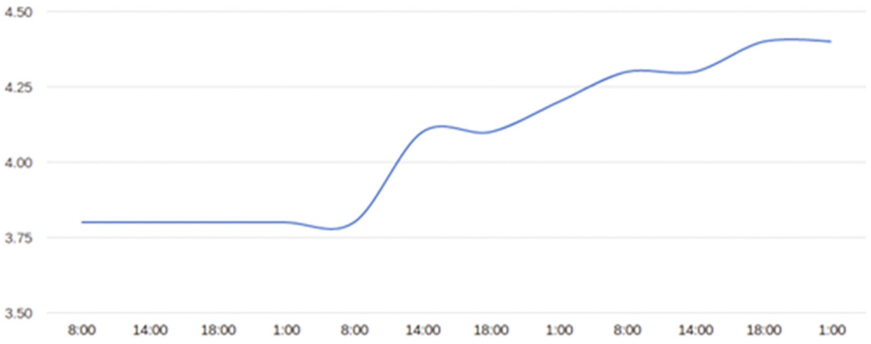


Fig. 8. pH variation of crop with bud rot [pH/s]

Figure 9 shows luminosity variable for soil with bud rot presence.

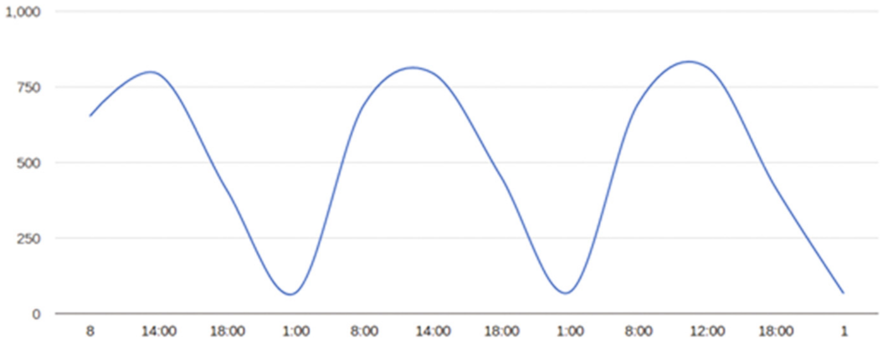


Fig. 9. Luminosity variation of crop with bud rot [lux/s]

Figure 10 shows humidity present in the container containing soil of a crop with presence of but rot. Here it is necessary to take into account that after 12 h of test start a lime dissolved in water is applied, for this reason humidity increases abruptly in about 10%.

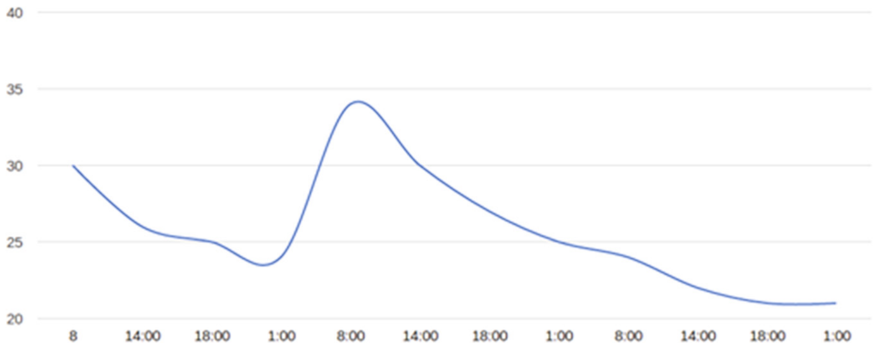


Fig. 10. Humidity variation of crop with bud rot [%/s]

4 Conclusions

In this paper a WSN system and node design has been presented. The WNS nodes that make possible acquire, process and transmit sensor measurements has been developed. The prototype can send measurements in real time mode to a central server process. In this project the prototype in set up to capture variables such as: pH, humidity, temperature and luminosity from African Palm trees. Additionally, the WSN device is equipped with different source of power supply and georeferencing system. The results indicate that the proposed system makes possible an early detection of abiotic factors that may trigger bud rot in African Palm trees. Thus, enforce early detection of triggered environment for bud rot, farmers could apply strategies to minimize bud rot impact on crops.

Acknowledgment. The authors gratefully acknowledge the support of the Thematic Network: RiegoNets (514RT0486)-CYTED.

References

1. Fedepalma, Comunicaciones and others: Mejores prácticas agroindustriales para una excelente palmicultura: Boletín El Palmicultor **537**, 17–18 (2016)
2. SISPA, Fedepalma: Anuario Estadístico (2014)
3. Ceballos, M.R., Gorricho, J.L., Gamboa, O.P., Huerta, M.K., Rivas, D., Rodas, M.E.: Fuzzy system of irrigation applied to the growth of Habanero Pepper (Capsicum Chinense Jacq.) under protected conditions in Yucatan, Mexico. Int. J. Distrib. Sens. Netw. (2015). SAGE Publications, London, England, UK

4. Erazo, M., Rivas, D., Pérez, M., Galarza, O., Bautista, V., Huerta, M., Rojo, J.L.: Design and implementation of a wireless sensor network for rose greenhouses monitoring. In: 2015 6th International Conference on Automation, Robotics and Applications (ICARA), pp. 256–261. IEEE (2015)
5. Fernandez, L., Huerta, M., Sagbay, G., Clotet, R., Soto, A: Sensing climatic variables in a orchid greenhouse. In: 2017 International Caribbean Conference on Devices, Circuits and Systems (ICCDCS), pp. 101–104. IEEE (2017)
6. Ibayashi, H., Kaneda, Y., Imahara, J., Oishi, N., Kuroda, M., Mineno, H.: A reliable wireless control system for tomato hydroponics. Sensors **16**(5), 644 (2016)
7. Laing, D.: La causa de pudrición de cogollo (PC) en palma de aceitehipótesis abiótica-edáfica. (2009)
8. de Franqueville, H.: La pudrición del cogollo de la palma aceitera en América LatinaRevisión preliminar de los hechos y logros alcanzados. BuroTrop Cirad-Cp Departamento de Cultivos Perennes (2001)
9. Corley, R.H.V.: How much palm oil do we need? Environ. Sci. Pol. **12**(2), 134–139 (2009). Elsevier

A Cloud-Based Platform for Decision Making Support in Colombian Agriculture: A Study Case in Coffee Rust

Emmanuel Lasso(✉), Oscar Valencia, David Camilo Corrales,
Iván Darío López, Apolinar Figueroa, and Juan Carlos Corrales

Telematics Engineering Group, University of Cauca, Popayán, Colombia
{eglasso,ovalencia,dcorrales,navis,apolinar,jcorral}@unicauca.edu.co
http://www.unicauca.edu.co

Abstract. In the last years, the yield of Colombian crops has been affected by climate change. The weather variation affects the Colombian crops with the occurrence of diseases as coffee rust. To address the coffee rust control, we proposed a cloud-based platform for decision making support named AgroCloud. The coffee crop weather of 100 municipalities from upper basin of the Cauca river were monitored. This information was used to improve the disease control process. User Acceptance Test carried out with domain end users show that the platform is useful and is easily usable.

Keywords: Coffee rust · DSS · Decision support · AgroCloud · Diseases control · Agriculture

1 Introduction

In Colombia, one of the biggest challenges is the marketing of agricultural products. The crops have increased twelve-fold compared with the crops in the last 20 years. 42.3 millions of hectares are committed to agriculture yield and 7.1 millions hectares in crops [1]. The Colombian farmers play an important role; they produce 78.8% of agricultural products and 60% are products of the basic food basket for Colombian people [2].

In the last years, the yield of Colombian crops has been affected by climate change. The increase of temperatures and rainfall variation affect the Colombian crops with the occurrence of diseases and pest invasions [3].

The rust is an example of a disease. The rust attacks to coffee crops where the weather is a key factor for its germination. The Rust disease has reduced considerably the coffee production in Colombia (by 31% on average during the epidemic years compared with 2007). These reductions have had direct impacts on the livelihoods of thousands of small holders and harvesters [4]. More than 350.000 Colombian families depend on coffee harvest for their sole income.

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_14

As such, the coffee rust impacts terribly on the economic and social aspects of the main coffee-growing regions [5].

To tackle the aforementioned problem, we propose a cloud-based platform for decision making support in Colombian agriculture named AgroCloud. The study case is the coffee rust. The remainder of this paper is organized as follows: Sect. 2 presents the study case and related works; Sect. 3 the AgroCloud platform; Sect. 4 presents results and Sect. 5 conclusions and future work.

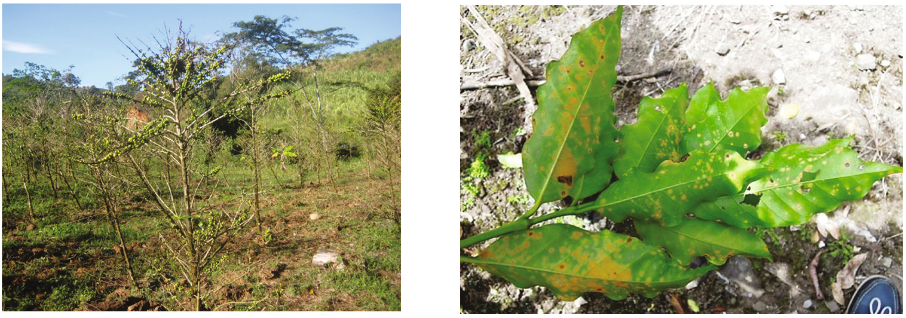
2 Background

In this section, we explained the study area and the concepts that are employed in AgroCloud.

2.1 Study Case

The study case of AgroCloud is the coffee rust. The weather conditions of the coffee rust are monitored in 100 municipalities from upper basin of the Cauca river (ubCr). The disease and the monitored area are explained next.

Coffee Rust is caused by the fungus *Hemileia vastatrix*, a parasite that affects the coffee leaves. Among the cultivated species, Coffea arabica is the most severely attacked. The disease causes defoliation, in the worst-case scenario (Fig. 1a), death of branches and crop losses [4]. The first symptoms are yellowish spots that appear on the underside of leaves (Fig. 1b). These spots then grow and produce uredospores with a orange colour. Chlorotic spots can be observed on the upper surface of the leaves [6,7].



(a) Symptoms of strong defoliation, in Libano, Tolima, Colombia.

(b) Small yellowish lesions on the underside of a coffee leaf.

Fig. 1. Coffee rust. Source: [4]

Below is described the Colombian region where the weather conditions of the coffee rust are monitored.

Monitoring the Coffee Crop Weather 100 municipalities from upper basin of the Cauca river (ubCr) are monitored by AgroCloud. ubCr is composed by four Departaments: 32% Cauca, 47% Valle of Cauca, 13% Risaralda and 8% Quindio. The total area of ubCr is represented by 23.000 Km2 with a population of 7.122.518 people. The main crops seeded in ubCr is the coffee[1]. Figure 2 shows the 100 municipalities on upper basin of the Cauca river.

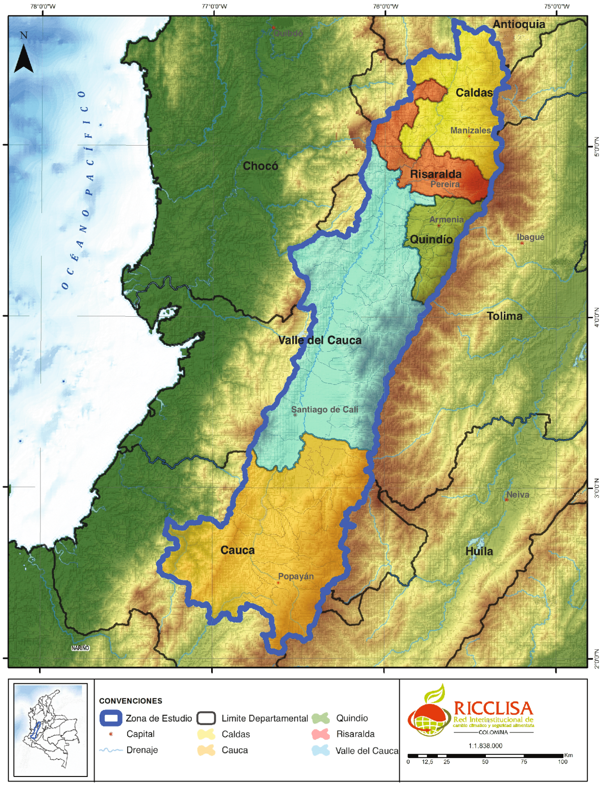


Fig. 2. Upper basin of the Cauca river. Source: www.ricclisa.org

2.2 Decision Support Systems

One important concept to understand AgroCloud are the Decision Support Systems (DSS). Scott et al. [8] a recognized researchers in DSSs field define them as systems that combine individual intellectual resources and the capabilities of a computer to improve the decisions quality.

From agriculture, a DSS is a mechanism that collects, organizes, and integrates all types of information required for producing a crop; The first step

[1] www.ricclisa.org.

consists in the analysis and interpretation of the information; subsequently the analysis is used to recommend the most appropriate action choices [9]. Expert knowledge is a key element of DSS and it is used to assist producers with both daily operational and long-range strategic decisions [10].

In this work, a cloud-based platform for decision making support in Colombian agriculture is proposed. The coffee rust was the disease taken as study case.

2.3 Related Works

Although our proposal is focused in a decision support system for chemical control in crops with coffee rust, we consider important describe the works that address the coffee rust detection from computer science, since that task is the starting point for disease control. Also, works that propose the use of DSS in diseases crops.

Coffee Rust Detection within computer science is addressed from data mining. Colombian and Brazilian researchers have in recent times attempted to detect the coffee rust through Decision Trees (DT), K Nearest Neighbor (K-NN), Bayesian Networks (BN), Support Vector Machines (SVM), Artificial Neural Networks (ANN) and Ensemble Methods. Table 1 shows a summary of the related works for coffee rust detection.

Table 1. Related works: coffee rust detection

Work	Publication year	Algorithm	Country of coffee crops
[5,7,11–15]	2014 – 2017	DT, K-NN, BN, SVM, Ensemble Method	Colombia
[16–20]	2009 – 2012	DT, BN, SVM	Brazil

In Colombian coffee crops, two datasets were used. The first one was collected trimonthly for 18 plots, closest to weather station at the Technical Farm (Naranjos) of the Supracafé, in Cajibio, Cauca, Colombia (21°35′08″ N, 76°32′53″W), during 3 years (2011–2013). The dataset contains variables related with Weather conditions, Physic crop properties, and crop management [12]. The second dataset was obtained from Jazmín Village which is a coffee growing area sowing with *Caturra* variety in 45 farms approximately, monitored by Cenicafé and located in Santa Rosa de Cabal, Colombia (4°55’00”N, 75°38’0”W). The dataset contains samples for six daily meteorological attributes around 26/02/1986 and 15/12/1988 [7].

From Brazilian coffee crops, a dataset was built with information of the experimental farm Procafé (South latitude 21°34′00″ longitude West 45°24′22″

and altitude 940 m) located in Varginha, Minas Gerais, during the years 1998 – 2006. This dataset contains physic crop properties and weather conditions [18].

The main problem of the related works mentioned above is the low number of samples of Incidence Rate of Rust; if the available examples are few, the dataset does not represent a sample trustworthy of the population, then the data mining algorithms will be not inaccurate [7,11].

Decision Support Systems for Crops Several DSSs for crops have been developed. We reviewed works from 2012 until present year. DSSs for control of diseases in crops of potato, tomatoes, grapes and wheat have been built, while DSSs for crops of citrus, soybean, sorghum, rapeseed, cardoon and sugarcane are focused in crop management. Table 2 presents a brief summary of related works of DSS in crops.

Table 2. Related works: DSS for crops

Work	Publication year	Crop	Disease
[21]	2015	Potato and tomatoes	Late blight
[22]	2014	Grapes	Downy, Powdery mildews, Grey mould, and Black rot
[23]	2016	Citrus	-
[24]	2015	Soybean	-
[25]	2017	Tomatoes	Gray mold
[26]	2016	Grapes	Downy mildew, Powdery mildew, Black rot, Anthracnose
[27]	2012	Maize, Soybean, Sorghum, Rapeseed, and Cardoon	-
[28]	2012	Sugarcane	-
[29]	2015	Wheat	Powdery mildew and eyespot, Flag leaf from Septoria leaf blotch, Fusarium head blight
[30]	2016	Grapes	Black rot

Although numerous works propose DSSs for control of diseases and crops management, at the present time, these are not focused in coffee crops to control the rust. In the next section we explain AgroCloud: a cloud-based platform for decision making support for control of coffee rust.

3 AgroCloud Architecture

3.1 Conceptual Diagram

The DSS conceptual diagram for the detection and control of coffee rust is shown in Fig. 3.

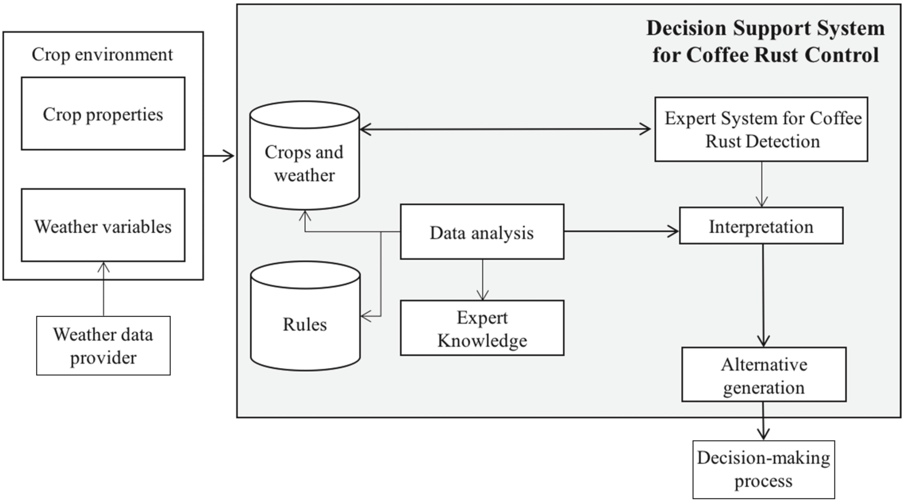


Fig. 3. Conceptual diagram of decision support system for coffee rust control

Crop environment is composed of information obtained from the weather data provider METEOBLUE[2] and crop properties entered by farmers. This information is stored and constitutes the main resource of the expert system for the detection of coffee rust favorable conditions in crops. This system identifies crop conditions for a given infection rate of the disease, as detailed in [31,32]. Once the DSS is consulted, the expert system checks whether the crop has favorable conditions for the disease. On the other hand, the Data Analysis component also makes use of weather and crop data stored, processed through rules and expert knowledge, in order to recognize the properties of the elements that make up the disease control. The Interpretation module identifies the state of the crop against the disease and combines this information with the suggested process for its control, resulting in the generation of alternatives that the farmer can take.

3.2 Deployment

AgroCloud[3] represents a cloud-based platform to support the development of information services for the Colombian agricultural sector and an early warning

[2] www.meteoblue.com.
[3] Available in https://agrocloudcolombia.com.

system to reduce vulnerability to variability and climate change phenomena. The platform is focused on the municipalities located in the upper basin of the Cauca River. The main components of the platform are: weather monitoring, weather forecasting, support for decision-making, Expert System (ES) and reports.

Weather monitoring is carried out from a subscription to a climate data provider, obtaining values of weather variables such as: air temperature, relative humidity, wind speed and direction, rainfall and solar radiation. The information is queried through the data provider API and stored in the AgroCloud databases.

The Decision Support System (DSS) makes use of computational tools to analyze the variables that intervene in crop diseases and to generate a decision that implements the alternative with more probability to be successful. As a specific case, the DSS has been developed for the management of Coffee Rust control and the costs of its application, as is described by Lasso and Corrales in [33].

The ES integrates the information that has been collected by the previous components (weather monitoring and forecasting, support for decision-making) for alerting on the presence of favorable conditions for the occurrence of rust in coffee crops. Once the parameters corresponding to the validity of the data are verified, an alert is communicated to the involved actors. This system corresponds to an implementation of the expert system for coffee rust proposed by Lasso and Corrales in [31,32], which makes use of graph patterns for the disease presented by the same author in [14].

Finally, **the agroclimatic reports** correspond to structured documents on the climatic conditions presented in a municipality, identifying significant events in different time periods. Additionally, reports may also characterize conditions for historically identified diseases.

The layered view of the architecture is shown in Fig. 4 and its components are described below.

– Weather data provider: AgroCloud obtains the weather information in the municipalities from the weather data provider METEOBLUE, which delivers local weather data for any point in the world. It offers a web-based access interface consulted periodically, providing the values of different climate variables such as: temperature, rainfall, humidity, wind, among others.
– Persistence layer: Databases for the storage of several elements in crops environment, such as: Weather data, obtained from data provider described previously; Crops, that contains crop properties, agro-production management; and Users database, which contains the information of AgroCloud users according to the organizational structure of the production system. Additionally, in this layer the knowledge base for disease detection (early warning services) and control (decision-making support), obtained from experts knowledge is stored.
– Service layer: This layer is comprised of high granularity services (HGS), an enterprise service bus (ESB), and value-added services (VAS). The HGS are divided according to the nature of the functions that they offer. In this way, the agroclimatic (relationship between crop adaptation and climate)

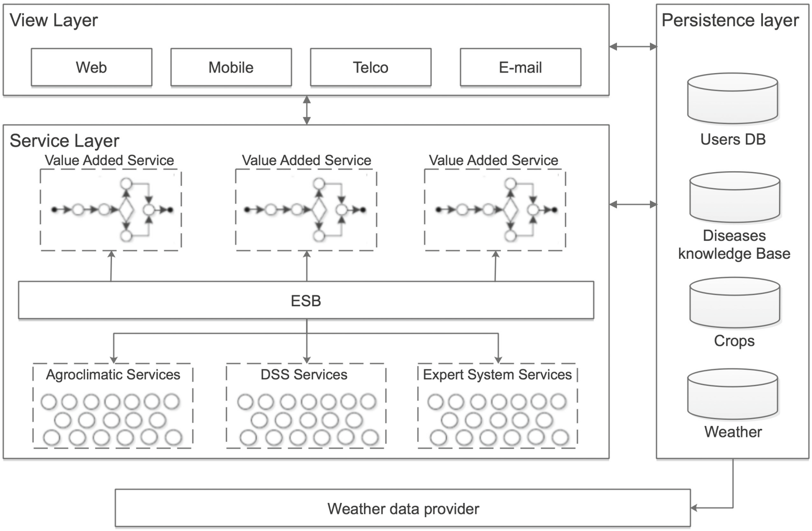


Fig. 4. Layered view of AgroCloud architecture

services provide the visualization and recovery of the weather monitoring data, forecast services and the report generation. The ESB allows the reuse of functions offered by the HGS, enabling the integration between different areas covered by AgroCloud (weather monitoring, disease detection and control). As a result, several VAS are obtained modeled as business process.

– View layer: Represents the interfaces for user interaction according the different services mentioned above. The Web platform and the Mobile application represent the main interaction, built based on elements of easy use and understanding for the different user roles. Telco refers to traditional telecommunication services, such as automatic voice calls and SMS (Short Message Service), as a communication channel for sensitive events for crops such as identification of favorable disease conditions. E-mail component is an element mainly used by the reporting system, allowing periodic and cost-free communication of significant events found in the influence area covered by the platform.

4 Results

End users are the main actors who interact with a product daily and constantly. Therefore, their satisfaction is one of the most important measures of the final product success; in particular those systems that support decision making and provide important information from the expert knowledge in a productive sector such as agribusiness. This section presents the User Acceptance Testing (UAT) methodology [34] and the main elements to take into account for validation of AgroCloud which comprises weather services, detection of favorable conditions for diseases, and the DSSEx component.

UAT is the last phase of the validation process of a software solution. This methodology intends to analyze, from the use of the application by end users, if there are failures in responses and input forms, usability, correspondence with the problem that tries to solve, and the impact on it. From the UAT theoretical basis, the application of this methodology specifically in the AgroCloud platform is presented below.

4.1 Type of Test

In the execution of UAT was used the concept of "black box", which is commonly categorized as a functional test, but can also be used for UAT. Users only know and interact with the system inputs and outputs, without being able to see the code and internal flow of operation. In addition, the end user knows the business requirements. This test was developed from meetings where each user used the system through a computer. In cases where the user group exceeded 10 people, the use of the system was projected so that everyone could perceive the process and answers obtained.

4.2 UAT Users

In the coffee production environment, there are 3 types of users directly related to the treatment and control of the crop. UAT was applied to 53 users (Fig. 5) and their characteristics are described below.

- Producers. Refers to the small, medium or large farmer, depending on the level of production. In addition, there are associations that seek to join efforts to carry out an orderly and articulated production. 5 medium-sized producers from Cajibío (Cauca), an association of organic coffee producers (Popayán) (12 producers), and an association of coffee growers from Los Andes - Corinto (ASPROCCAN) (20 producers).
- Agricultural Extension Officers (AEO). Staff with technical knowledge in coffee production, delegates by the Coffee Growers Federation to advise producers on different problems around the process. A technical assistance group of the National Federation of Coffee Growers (Colombia) was interviewed (15 users).
- Researchers. They are usually agronomists who study the phenomena that affect coffee production and how to solve them. An agronomist was interviewed for UAT.

4.3 Business Requirements

Provide relevant climatic information such as current conditions, historical data, weather forecast, among others. Additionally, support the selection of type of fungicide and spray system to be used to counteract the impact of coffee rust

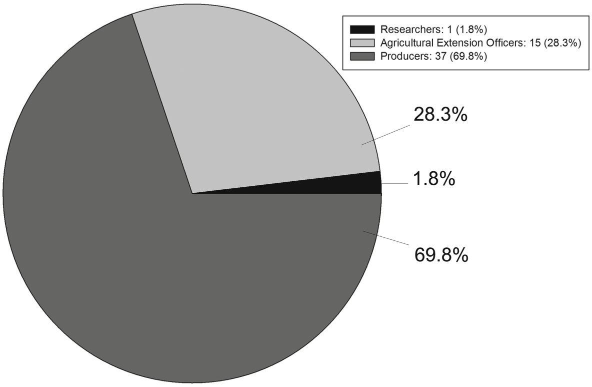


Fig. 5. Percentage of interviewed users by user group

disease. This choice was based on costs comparison of possible available combinations in the market by budget per application and year. On the other hand, the expert system requirement is to detect favorable conditions for diseases (case study of coffee rust) in the upper basin of Cauca River municipalities.

4.4 Test Cases

Before defining the test cases that allowed to validate the AgroCloud platform, it is necessary to mention the lack of climatic data in the study area for the periods when the test was executed. Taking into account this, it was necessary to define an adaptation scheme of the systems and their sources of information to carry out the simulation of the different possible scenarios around rust epidemics. In this way, data from a weather station and disease monitoring in a pilot coffee farm were used to simulate scenarios that favored rust, based on knowledge stored in the expert system. Based on the above considerations, test cases are shown below.

Favorable Conditions

- Listing of favorable conditions for diseases.
- View details of a favorable condition.
- Access rust control management from the details of a favorable condition.

Control Management

- Get recommendation of the fungicide application moment according to the flowering period.
- Access to cost management from the recommendation of the type of fungicide.
- Get recommendation of the fungicide application moment according to the fixed calendar system.

- Access to cost management from the recommendation of implementation dates by fixed calendar system.
- Obtain recommendation of the fungicide application moment according to the level of rust infection.
- Get current recommendations on the fungicide application technology for rust control.

Cost Management

- Obtain costs for a known control system.
- Get cost comparison for different control system options (fungicide and spraying equipment).
- View details of a control system configuration costs.

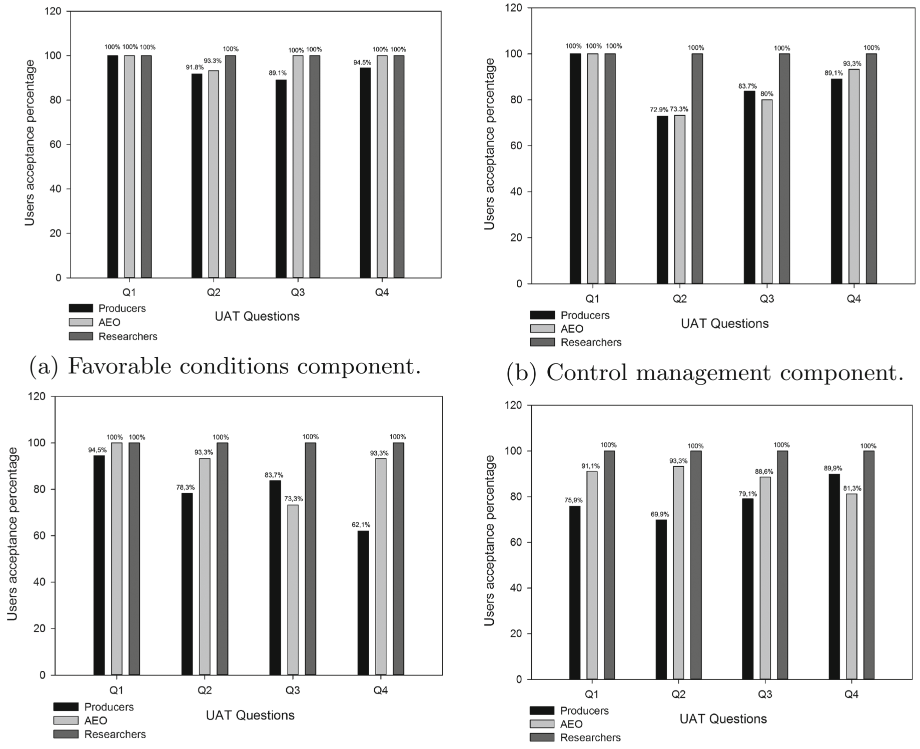
Weather Services

- Get current conditions of a municipality.
- Get the historical data of a municipality.
- Get the weather forecast of a municipality.

Additionally, four questions about each test case were applied. The number of users by user type who answered affirmatively to each of these questions are shown in Fig. 6.

- Q1: Was the task completed satisfactorily?
- Q2: Did the user require assistance to complete the task satisfactorily?
- Q3: Were the information request forms satisfactorily understood?
- Q4: Users consider the tool useful?

From results shown in the previous figure, it is possible to establish that the AgroCloud platform presents relevant information for users at a high level. However, cost management and weather services were not completely understandable by producers and AEO user groups. In this sense, improving the usability of these services becomes a key aspect for all user groups to obtain the greatest benefit in their decision-making processes. Similarly, some users needed technical assistance to complete the tasks of all AgroCloud services, in particular the producer group to complete the tasks of all AgroCloud components. These aspects evidenced the need to improve the platform help menus according to the obtained feedback. Therefore, it is important to take into account the users recommendations and comments in order to analyze and determine which changes allow to increase the comprehensibility level of the platform.



(a) Favorable conditions component.

(b) Control management component.

(c) Cost management component.

(d) Weather services component.

Fig. 6. UAT results for three different user groups in the AgroCloud components

5 Conclusions and Future Works

Systems aimed at solving problems in agricultural production environments are challenged not only to generate the best recommendation, but also to generate an environment of easy understanding and usability for all possible roles and actors present. The interaction between the development team and producers allows us to know the concerns of the main users and generate ideas to further develop new functionalities in order to improve the system. Recommendations and comments obtained in the UAT allowed to define different guidelines to improve the platform (utility and usability) for the end users in an agricultural environment.

Among the main recommendations obtained from UAT that will be taken into ac-count for the improvement of the platform, are: (i) take into account other diseases such as South America leaf spot and brown leaf spot, in addition to pests, (ii) recommendations to improve the fertilization of crops, (iii) recommendation of plan nutrition from digital analysis of its leaves photos, (iv) take into account other crops, (v) address nutrition and fertilization in crops, (vi) producers want to take advantage of the soil analysis they have on their farms. The above considerations must be accompanied by the incorporation of an autochthonous written and celestial language that can be understood and assimilated by the group of producers (words, phrases, symbols, images, among others that are handled within a rural community).

As a future work, real and close to real time data are required in order to improve the decisions-making process. Similarly, significant weather events should be exclusively related to the selected station or municipality. On the other hand, add a comparison section of hourly, daily, monthly, and annual data; and to show the maximum and minimum values of the meteorological variables. Finally, regarding the detection of favorable conditions for diseases, users recommended that detections should be sent by Short Message Service (SMS) or, if possible, by a message from a social network like WhatsApp, Facebook, or Twitter.

Acknowledgments. The authors are grateful to the University of Cauca and its Telematics Engineering Group (GIT), the Colombian Administrative Department of Science, Technology and Innovation (Colciencias) and AgroCloud project of The Interinstitutional Network of Climate Change and Food Security of Colombia (RICCLISA) for supporting this research.

References

1. Colombia siembra. Cientific DE-DEI-11, Ministerio de Agricultura y Desarrollo Rural (2016)
2. Piza, C., Diaz, L.P., Pulido, N., Rincon, R.J.D.: Agricultura familiar: una alternativa para la seguridad alimentaria. Conexion Agropecuaria JDC, **6**(1), September 2016
3. Fernandez, M.: Efectos del cambio climtico en la produccion y rendimiento de cultivos por sectores. Cientific 2130628, IDEAM (2013)
4. Avelino, J., Cristancho, M., Georgiou, S., Imbach, P., Aguilar, L., Bornemann, G., Läderach, P., Anzueto, F., Hruska, A.J., Morales, C.: The coffee rust crises in colombia and central america (2008–2013): impacts, plausible causes and proposed solutions. Food Secur. **7**(2), 303–321 (2015)
5. Corrales, D.C., Pena, A., Leon, C., Figueroa, A., Corrales, J.C.: Early warning system for coffee rust disease based on error correcting output codes: a proposal. Revista Ingenierias Universidad de Medellin **13**, 57–64 (2014)
6. Rivillas, C., Serna, C., Cristancho, M., Gaitan, A.: La Roya del Cafeto en Colombia. Impacto, manejo y costos de control. Cientific bot036, Cenicafe (2011)
7. Corrales, D.C., Gutierrez, G., Rodriguez, J.P., Ledezma, A., Corrales, J.C.: Lack of Data: Is It Enough Estimating the Coffee Rust with Meteorological Time Series? pp. 3–16. Springer International Publishing, Cham (2017)
8. Keen, P.G.W., Morton, M.S.S.: Decision Support Systems: An Organizational Perspective. Addison-Wesley series on decision support, Addison-Wesley Pub. Co. (1978)
9. Agrios, G.N.: Plant Pathology. Academic press, New York (1997)

10. Plumb, R.: Precision agriculture in the 21st century: geospatial and information technologies in crop management, committee on assessing crop yield: site-specific farming, information systems and research opportunities, board on agriculture. Pest Management Science **56**(8), 723–723 (2000). National research council, National academy press, Washington DC, USA 1997, xii+ 149 pp, price£ 32.95. ISBN 0-309-05893-7
11. Corrales, D.C., Figueroa, A., Ledezma, A., Corrales, J.C.: An Empirical Multi-classifier for Coffee Rust Detection in Colombian Crops, pp. 60–74. Springer International Publishing, Cham (2015)
12. Corrales, D.C., Ledezma, A., Pea, Q.A.J., Hoyos, J., Figueroa, A., Corrales, J.C.: A new dataset for coffee rust detection in Colombian crops base on classifiers. Sistemas y Telemtica **12**(29), 9–23 (2014)
13. Corrales, D.C., Casas, A.F., Ledezma, A., Corrales, J.C.: Two-level classifier ensembles for coffee rust estimation in colombian crops. Int. J. Agric. Environ. Inf. Syst. (IJAEIS) **7**(3), 41–59 (2016)
14. Lasso, E., Thamada, T.T., Meira, C.A.A., Corrales, J.C.: Graph Patterns as Representation of Rules Extracted from Decision Trees for Coffee Rust Detection, pp. 405–414. Springer International Publishing, Cham (2015)
15. Plazas, J.E., Rojas, J.S., Corrales, D.C., Corrales, J.C.: Validation of Coffee Rust Warnings Based on Complex Event Processing, pp. 684–699. Springer International Publishing, Cham (2016)
16. Meira, C.A.A., Rodrigues, L.H.A., Almeida, S., de Moraes., S.A.: Warning models for coffee rust control in growing areas with large fruit load. Pesquisa Agropecuaria Brasileira **44**(3), 233–242 (2009)
17. Luaces, O., Rodrigues, L.H.A., Meira, C.A.A., Quevedo, J.R., Bahamonde, A.: Viability of an alarm predictor for coffee rust disease using interval regression. In: Garca-Pedrajas, N., Herrera, F., Fyfe, C., Bentez, J.M., Ali, M. (eds.) Trends in Applied Intelligent Systems. Lecture Notes in Computer Science, vol. 6097, pp. 337–346. Springer, Heidelberg (2010)
18. Cintra, M.E., Meira, C.A.A., Monard, M.C., Camargo, H.A., Rodrigues, L.H.A.: The use of fuzzy decision trees for coffee rust warning in Brazilian crops. In: 2011 11th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 1347–1352, November 2011
19. Luaces, O., Rodrigues, L.H.A., Meira, C.A.A., Bahamonde, A.: Using nondeterministic learners to alert on coffee rust disease. Expert Syst. Appl. **38**(11), 14276–14283 (2011)
20. Perez-Ariza, C., Nicholson, A., Flores, M.: Prediction of coffee rust disease using Bayesian networks. pp. 259–266. DECSAI University of Granada (2012)
21. Small, I.M., Joseph, L., Fry, W.E.: Development and implementation of the blight-pro decision support system for potato and tomato late blight management. Comput. Electron. Agric. **115**, 57–65 (2015)
22. Rossi, V., Salinari, F., Poni, S., Caffi, T., Bettati, T.: Addressing the implementation problem in agricultural decision support systems: the example of vite. net®. Comput. Electron. Agric. **100**, 88–99 (2014)
23. Navarro-Hellín, H., Martínez-del Rincon, J., Domingo-Miguel, R., Soto-Valles, F., Torres-Sánchez, R.: A decision support system for managing irrigation in agriculture. Comput. Electron. Agric. **124**, 121–131 (2016)
24. Dandawate, Y., Kokare, R.: An automated approach for classification of plant diseases towards development of futuristic decision support system in indian perspective. In: 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 794–799. IEEE (2015)

25. Cañadas, J., Sánchez-Molina, J.A., Rodríguez, F., del Águila, I.M.: Improving automatic climate control with decision support techniques to minimize disease effects in greenhouse tomatoes. Inf. Process. Agric. **4**(1), 50–63 (2017)
26. Waghmare, H., Kokare, R., Dandawate, Y.: Detection and classification of diseases of grape plant using opposite colour local binary pattern feature and machine learning for automated decision support system. In: 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), pp. 513–518. IEEE (2016)
27. Antonopoulou, E., Karetsos, S.T., Maliappis, M., Sideridis, A.B.: Web and mobile technologies in a prototype DSS for major field crops. Comput. Electron. Agric. **70**(2), 292–301 (2010). Special issue on Information and Communication Technologies in Bio and Earth Sciences
28. Stray, B.J., Van Vuuren, J.H., Bezuidenhout, C.N.: An optimisation-based seasonal sugarcane harvest scheduling decision support system for commercial growers in south africa. Comput. Electron. Agric. **83**, 21–31 (2012)
29. Jarroudi, M.E.L., Kouadio, L., Beyer, M., Junk, J., Hoffmann, L., Tychon, B., Maraite, H., Bock, C.H., Delfosse, P.: Economics of a decision-support system for managing the main fungal diseases of winter wheat in the grand-duchy of luxembourg. Field Crops Res. **172**, 32–41 (2015)
30. Molitor, D., Augenstein, B., Mugnai, L., Rinaldi, P.A., Sofia, J., Hed, B., Dubuis, P.-H., Jermini, M., Kührer, E., Bleyer, G., et al.: Composition and evaluation of a novel web-based decision support system for grape black rot control. Eur. J. Plant Pathol. **144**(4), 785–798 (2016)
31. Lasso, E., Thamada, T.T., Meira, C.A.A., Corrales, J.C.: Expert system for coffee rust detection based on supervised learning and graph pattern matching. Int. J. Metadata Semant. Ontol. (2017, to appear)
32. Lasso, E., Corrales, J.C.: Expert system for crop disease based on graph pattern matching: a proposal. Revista Ingenieras Universidad de Medellin **15**(29), 81–98 (2016)
33. Lasso, E., Valencia, Ó., Corrales, J.C.: Decision Support System for Coffee Rust Control Based on Expert Knowledge and Value-Added Services, pp. 70–83. Springer International Publishing, Cham (2017)
34. Cimperman, R.: UAT Defined: A Guide to Practical User Acceptance Testing, 1st edn. Addison-Wesley Professional, Boston (2006)

Potential Yields of Maize and Barley with Climate Change Scenarios and Adaptive Actions in Two Sites in Mexico

Antonio R. Arce-Romero, Alejandro I. Monterroso-Rivas(✉), Jesús D. Gómez-Díaz, and Miguel A. Palacios-Mendoza

Departamento de Suelos, Universidad Autónoma Chapingo, Carretera Mexico - Texcoco Km 38.5, 56230 Texcoco de Mora, Estado de México, Mexico
antonio.arce93@gmail.com, aimrivas@correo.chapingo.mx, dgomez1059@yahoo.com.mx, angelpalacios1291@gmail.com

Abstract. Maize is a grain of great importance for Mexican food; and along with barley and its industrial applications represent two crops economically important for the country. With climate change scenarios, distribution and yields of grains could be affected, putting the food security of much of the population at risk. In this work, potential yields of maize and barley were modeled with climate change scenarios and two adaptation actions were evaluated. In order to model yields, the AquaCrop FAO model was used. This model has a water approach and has been widely used in Mexico. Two study cases were chosen and their climate, soil, phenological and management information was compiled. Baseline models were calibrated using four concordance indexes. Once calibrated, we tested the response of yields to 28 climate change scenarios; considering five General Circulation Models, two RCP and three time horizons. Two adaptation actions were evaluated: changing planting date and increase of organic mulches. Results show that yield of maize in the near future (2015–2039) would fall 50% average, while barley yields would decrease in 40%. Adaptation measure based on changing planting date was as effective as increasing mulches in both cases. Maize could gain 1 ton/ha by taking this action, while barley could gain 250 kg/ha in the place studied.

Keywords: Food security · AquaCrop · Climate change

1 Introduction

Climate change has highlighted the susceptibility of the food security of Mexico (Gay Garcia and Rueda Abad 2015). There have been many studies on this subject, most of them focused on knowing the impacts of climate change on agricultural productivity and distribution of crops (Gay Garcia and Rueda Abad 2015). In this sense, Cervantes et al. (2014) report a literature review of research carried out regarding the impacts of climate change in Mexico highlighting that biggest impacts are reduced yields and reduced areas suitable for crops. Under a climate change focus, Ojeda-Bustamante et al. (2011) proposed a study of future water needs and changes in the phenological cycles for maize, beans, tomatoes, potatoes, sorghum, sugar cane, alfalfa and mango; finding

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_15

that crop water demand will decline 13% in the autumn-winter cycle; 6% for the spring-summer cycle and will increase by 7% for perennial crops.

Those studies highlight the importance of analyze crop yields with climate change scenarios especially because yield is a variable directly related to economic benefit of agriculture. Despite the multitude of approaches to evaluate crop yields, AquaCrop FAO model (Steduto et al. 2009) has shown acceptable results in terms of simplicity, robustness and precision in a national and international context (Soddu et al. 2013; Garcia et al. 2015; Flores Gallardo et al. 2013; Heng et al. 2009). The number of input variables is lower compared with models like WOFOST or CropSyst, which makes AquaCrop more accessible without losing its predictive ability (Todorovic et al. 2009).

AquaCrop is a model that uses the amount of water available to simulate current and potential biomass production. It was developed by the Land and Water Division of FAO and is the result of the revision of irrigation and drainage document FAO No. 33 entitled "Yield Response to Water" (Smith and Steduto 2012). In a study conducted in the United States validating AquaCrop for maize, it was found that the software was able to correctly simulate the progression of the accumulation of grain and fodder (Mebane et al. 2013). Also Garcia et al. (2015) calibrated and executed AquaCrop for maize at three sites in Colombia, finding a good fit of the modeled and observed yield. In Mexico, Flores-Gallardo et al. (2013) used AquaCrop software to validate and model maize yields in Sinaloa for the years 2003 to 2009, getting a good prediction performance.

In addition to its applications in water management of crops, it is known to have been widely used to model potential impacts of climate change on different agricultural systems. Regarding climate change focuses, (Vanuytrecht et al. 2015) modeled yields of winter wheat, maize, potatoes and sugar beets for 2050 in Belgium, finding increased variability of yield. Some increases were found due to the high amount of CO2 and the lengthening of the growing season, but decrements were associated with water stress.

Based on the above, the objective of this work was to model the potential yield of maize and barley under climate change scenarios in two study cases using software AquaCrop to propose and model adaptation actions.

2 Methodology

2.1 Crops and Biophysical Information

Two crops of economic importance in Mexico were chosen to model the potential yield in two study sites. The case studies presented in this paper are maize in Chapingo, Estado de Mexico and barley in Perote, Veracruz.

Maize is a cereal of great importance to Mexico. White maize grain is mainly used for making traditional tortillas and tamales, but it can also be obtained oil for the manufacture of varnishes, paints, artificial rubbers and soaps (SIAP 2016).

On the other hand, barley has had an increasing economic importance from its use in the brewing industry. The main use of barley in Mexico is to obtain malt grain required for brewing, and in a lesser proportion for the production of fodder. According to official statistics, in our country cultivation is practiced at least in one of the two cycles of the year in 23 states in Mexico (INIFAP 2011).

Since the software used requires the inclusion of specific site parameters; climatic, soil, crop and management information was referenced to weather stations. Chapingo weather station was used to model maize, it is located in the municipality of Texcoco, State of Mexico (19.48° North, −98.88° West). Perote weather station was used to model barley; it is located in the municipality of Perote, Veracruz (19.58° North, −97.24° West).

For each station we extracted climate information for 10 years, starting from 2000. Climatic variables used were precipitation, maximum and minimum temperature. Evapotranspiration was calculated following the Penman equation modified by FAO (Allen 2006).

The soil module was filled using information from digital cartography provided by the National Institute of Statistics and Geography (INEGI, *Instituto Nacional de Estadística y Geografía*), which are at a scale of 1: 250 000. The soil-related variables were the number and depth of horizons, texture, curve number and depth of the aquifer. To determine the soil parameters of each season, the official classification of textural classes were used to define the soil moisture regime, which is available in the default files of AquaCrop.

Information of the phenology and crop management was obtained from scientific literature and official sources of agricultural information in México (INEGI 2005; INIFAP 2011; SAGARPA 2009). The phenological variables used were density and planting date; days from seed to germination, flowering, maturity, senescence; root depth, harvest index, susceptibility to high and low temperatures and salinity.

2.2 Calibration and Validation of the Baseline Scenario

Crop yields were modeled using software FAO AquaCrop (Steduto et al. 2009). The input variables of the software retrieves climatic, agronomic and soil information, which was collected in previous steps. Nevertheless, a transformation of the data was required in order to match the appropriate format of AquaCrop software.

Once the input information was captured, the software was run and the results were recorded in a spreadsheet in Excel format. We recorded the following variables: dry yield, biomass produced and potential biomass (tn/ha), water productivity (kg/m3), harvest index (dimensionless), temperature stress and water stress (% of damage) and precipitation of the growing period of the crop (mm). Then we carried out several comparative analysis, which focused on contrast modeled and observed yield, as well as its relationship with precipitation and stresses.

In the context of crop modeling, performance evaluation is especially useful to consolidate a model which is then used with climate change scenarios (Soddu et al. 2013; Deb et al. 2014). Thus, models were evaluated by comparing observed and modeled yields. Historical records of crop yields were obtained from the Agricultural Information Service and Fisheries (SAGARPA, *Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación*). The indexes used for validation were the correlation coefficient between precipitation and yield modeling, the determination coefficient, the mean square error and the normalized mean square error. These indexes were used in accordance with the manual of AquaCrop software.

2.3 Performance with Climate Change Scenarios and Adaptation Actions

Once crop models were calibrated in a baseline scenario, we proceeded to incorporate climate change scenarios. The variables used in future climate scenarios were the maximum and minimum temperature, evapotranspiration, precipitation and the concentration of carbon dioxide. To obtained the first four variables from results of the Fifth National Communication of Mexico to the United Nations Framework Convention on Climate Change (Cavazos et al. 2013). Our scenarios considered five General Circulation Models (HADGEM, GFDL, REA Assembly, MPI and CNRM), which were obtained and processed from UNIATMOS platform (Fernandez Eguiarte et al. 2015).

Then we considered two Representative Concentration Pathways (4.5 and 8.5 W/m^2) and three time horizons (2015–2039, 2045–2069, and 2075–2099). In total, we tested 28 climate change scenarios.

The variable carbon dioxide under climate change scenarios was obtained directly from the program, as it provides projections considering the RCP 4.5, 6.5 and 8.5 W/m^2.

Once the data with climate change variables were captured in the program format, they were replaced by the baseline scenario and the software was executed to obtain potential future yields.

2.4 Adaptation Actions

They were proposed and evaluated two actions to adapt yields to the impact of climate change. A literature review was used to support the selection of those adaptation actions. In accordance with Ojeda-Bustamante et al. (2011), good adaptation actions for various crops is the use of hybrids or varieties resistant to heat cycle especially stress and compaction sowing period to colder months. On the other hand, Conde et al. (2006) propose the change in planting date, increased fertilizer application, the change in the variety of seeds, the application of irrigation and combinations of those actions.

Since this study is focused to model potential yields of maize and barley temporary under a rainfed regime, we selected the following adaptation actions: (1) change in planting date and (2) application or increase of organic mulches.

The first measure is based on changes in precipitation patterns, climate change scenarios showed a delay in the start of the rain season at the study sites analyzed (Fig. 1).

On the other hand, the use of organic mulches can be an effective measure to increase the soil fertility and decrease soil evaporation rates.

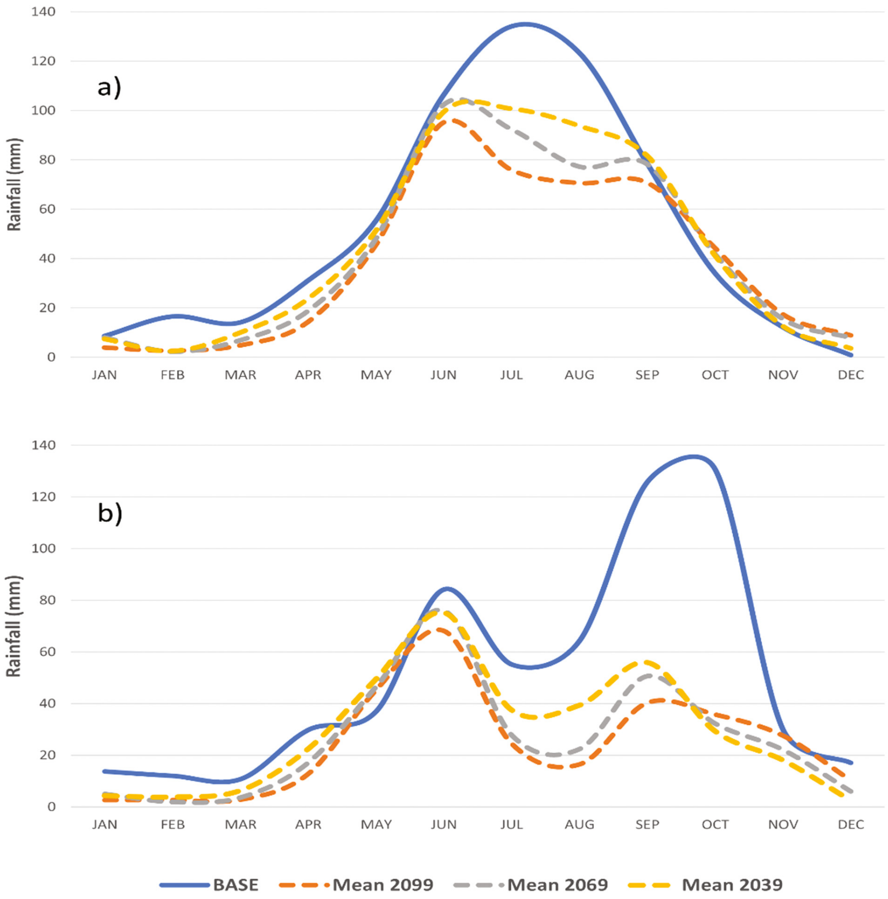


Fig. 1. Change in annual precipitation under climate change scenarios at two study sites. (A) Chapingo, Edo. Mex. (B) Perote, Ver.

3 Results

3.1 Maize at Chapingo, Estado de México

Calibration of Baseline. According to the phenological information obtained for this site, the growing cycle was of 150 days, with planting date on May 15. The beginning of flowering was set at 60 days from sowing and coincided with the point of maximum coverage and the maximum depth of the roots. The harvest index was 40%, while literature reported this index close to 50% for herbaceous crops.

The model generated by AquaCrop fitted with annual precipitation with (correlation coefficient of 0.85).

The mean squared error was 0.59, implying that the model is wrong about 590 kg/ha average for the selected time period. Figure 2 shows the modeled and observed yield, and some other variables of importance as rainfall and evapotranspiration. Some years had less rain than others, but its performance is higher. In this case, it is possible to explain this variation from the interaction of yield with the evapotranspiration. For 2008, for example, yield is higher than in 2005 and 2006, although rainfall was lower. However, the evapotranspiration was also lower, which meant less evaporative demand and a consequent increase in production efficiency of water precipitated, resulting in higher performance.

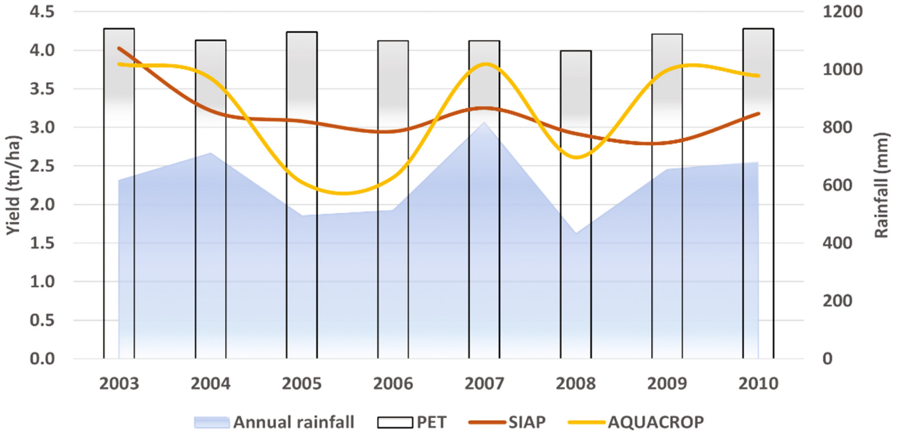


Fig. 2. Yield comparison between observed and modeled data for maize at Chapingo.

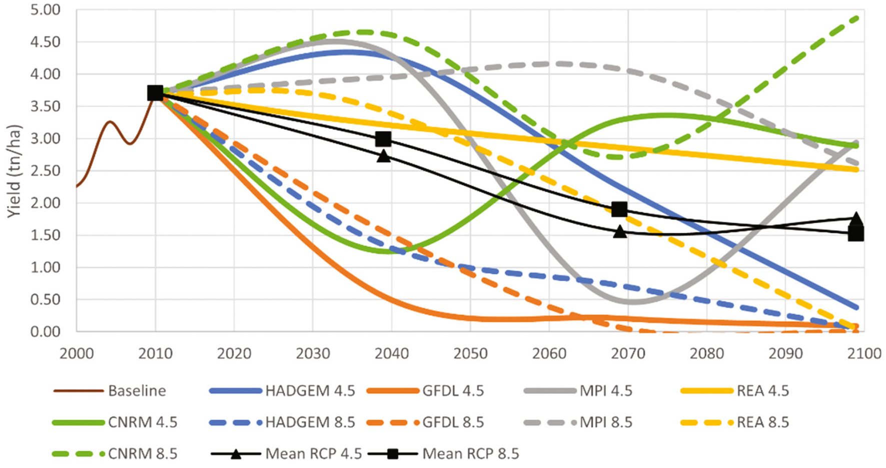


Fig. 3. Performance maize Chapingo, Edo. Mex with climate change scenarios

Scenarios of Climate Change and Adaptation Actions. Figure 3 shows the projections of maize yield under different climate change scenarios. The projected decrease in yields is observed differentially, depending on the model of climate change and the representative concentration pathway. In the near horizon (2039), yield under RCP 8.5 W/m^2 is higher than yield expected by the RCP 4.5 W/m^2, suggesting that rising temperatures could benefit crops in short-term.

For the middle horizon (2045–2069) there is more consensus among the models, because the difference between the RCP's 4.5 and 8.5 is 337 kg per hectare. Together of all models by the period 2045–2069, a general reduction in yield is observed with 1.5 tn/ha less than the baseline scenario, which implies a decrease of 46.7%.

In the far horizon (2075–2099) the more optimistic model CNRM RCP 8.5 provides increased yield of 1.6 tn/ha more than the base scenario. However, 5 of the 10 scenarios predict yields less than 500 kg/ha. The GFDL RCP 8.5 showed yields of 0 tn/ha. In average of all models, yield could decrease 49.3% relative to the baseline, thus harvesting 1.6 tn/ha or less than what is currently obtained.

Historical planting date for this area was set for 15th May, according to the moment where there is adequate moisture for the development of maize. Under climate change scenarios, rains could be delayed, as shown in Fig. 1, so the planting date should also be modified. Furthermore, the implementation of organic mulches covering 100% of the soil can reduce soil evaporation by 50%, thus benefiting the crop with higher water availability. Table 1 shows the characteristics of each scenario proposed and its adaptation actions.

Table 1. Adaptation actions proposed for maize in Chapingo, Edo. Mex.

Scenario number	Seedtime	Mulches	Description of the actions
1	15th of May	75% organic materials	Baseline parameters
2	15th of May	100% organic materials	Same planting date, increase in mulches
3	May 30	75% organic materials	Planting date is modified but mulches remain as baseline

Figure 4 shows the results in yields applying the three scenarios described above. In the figure, averages, maximum and minimum are shown.

Without changing any management practice that is currently performed (Scenario 1) decreased 49.3% return for the year 2099 would on average. Applying padded with organic matter (Stage 2) that would decrease 18.7%, while that by modifying the sowing date (Stage 3) 19.9% less be expected. In terms of performance without modifying any current management 1.6 tn/ha on climate change for the year 2099 they would be lost; modifying the planting date they would be lost 650 kg/ha and applying padding 100% would lower the yield at 610 kg/ha.

3.2 Barley in Perote, Veracruz

Calibration of Baseline. According to the calibrated parameters, the total growth cycle was established in 95 days, considering 7 days for emergency, 55 days from the planting date to the start of flowering, 65 days to maximum leaf coverage, 85 days to senescence and 95 days to maturity. The seeding density was adjusted to 150 plants/m^2 with a maximum coverage of 80% and a depth of 1.2 m radical. Harvest index was 35%.

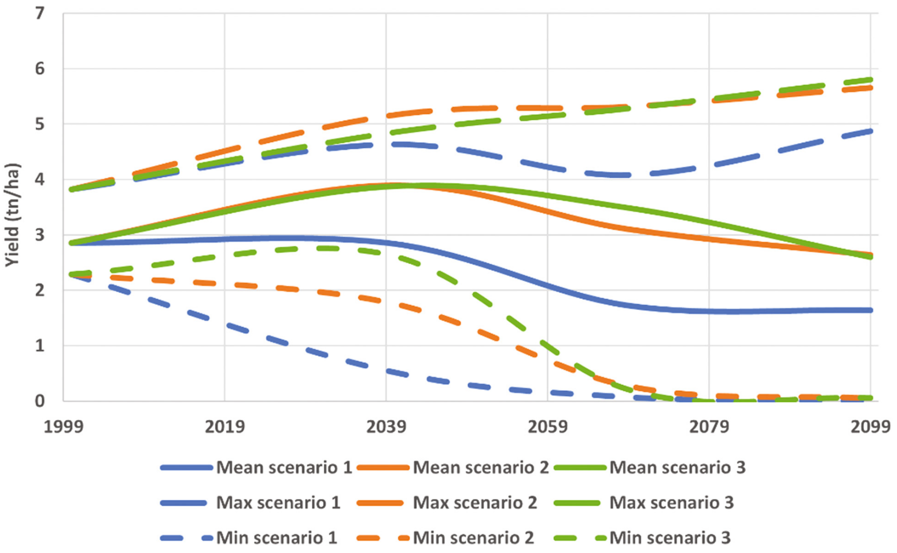


Fig. 4. Adaptation actions for maize in Chapingo, Edo. Mex

The model generated by AquaCrop fitted with data reported by the SIAP. Modeled yield and annual precipitation had a correlation coefficient of 0.73, suggesting strong relationship.

The mean squared error was 0.25, implying that the model is potentially wrong by 250 kg/ha. The normalized mean squared error was 28.7%, corresponding to an acceptable fit. The efficiency ratio Nash Sutcliffe is especially useful to verify if the trend of observed versus modeled data fits a 1:1 line, its value was 0.81 considered excellent for the purposes of this work. On the other hand, the Willmott index was 0.95, which indicates that the defined potential error between observed and modeled average was minimal. The observed average yield of 7 years was 0.86 tn/ha while modeled average yield was 0.84 tn/ha, only 20 kg/ha difference.

Behavior of modeled and observed yield is shown in Fig. 5. The fit between these yields is strong in the years where precipitation increased.

Rain of the growing season better explained the trend followed by performance than total annual rainfall. In future studies it is possible to assert that the use of the linear models is feasible, since it is possible to explain yields using rainfall for this study case.

These results confirm the importance of precipitation for growing barley in Perote, highlighting its distribution throughout the growing season.

Scenarios of Climate Change and Adaptation Actions. Figure 6 shows the yields of barley with climate change scenarios. Calibrated yield in the baseline scenario was 0.84 tn/ha. For the period 2015–2039, it is expected to obtain 0.43 tn/ha in average of all scenarios, which means 49% less than baseline. In the middle horizon (2045–2069) it is expected to harvest 0.51 tn/ha with a RCP of 8.5 W/m^2, while the RCP 4.5 W/m^2 0.34 tn/ha are projected. For the horizon 2075–2099 a slight improvement in yields is

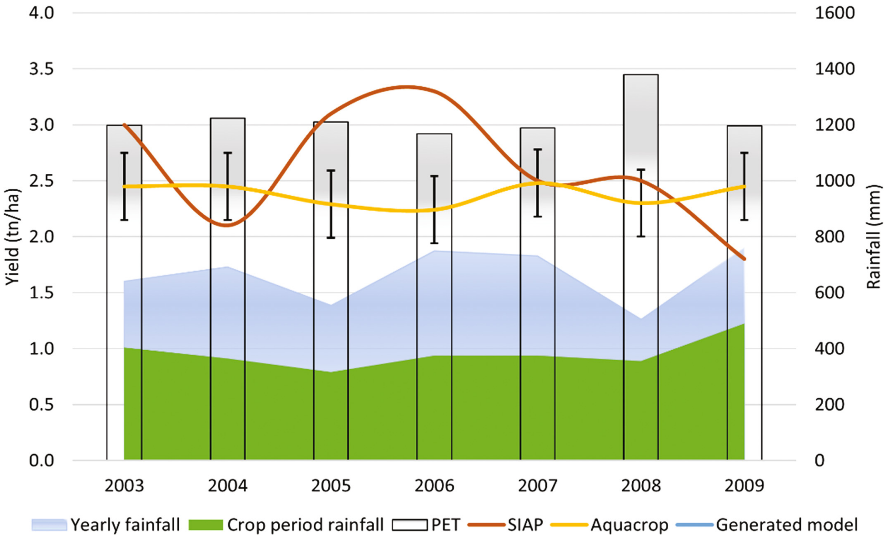


Fig. 5. Yield comparison between observed and modeled data for barley at Perote.

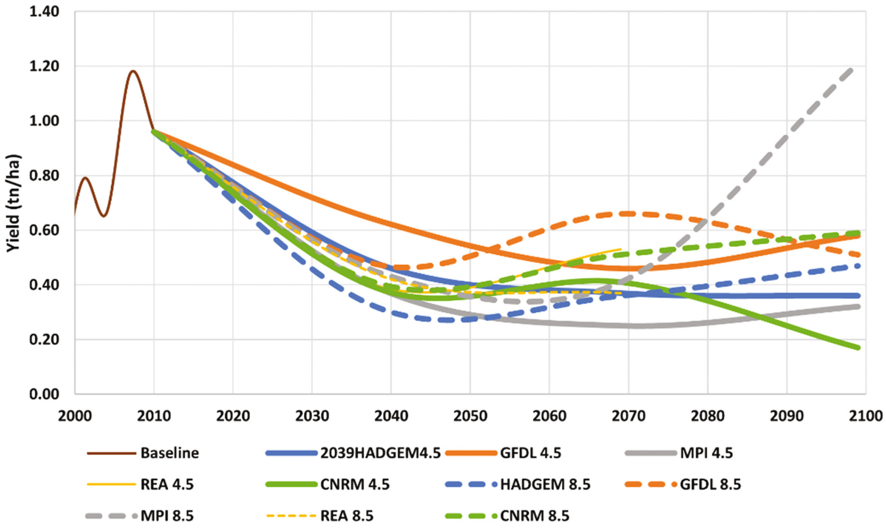


Fig. 6. Yield of barley Perote, Ver. under climate change scenarios.

recorded compared to the period 2045–2069. However, it involves a decrease of 40% compared to the baseline scenario.

Adaptive actions were similar to those used in the case of maize. Based on the distribution of rainfall with climate change scenarios, in this site we moved the planting date as an adaptation measure assuming that can be easily implemented. Since water stress had the greatest impact on yields under climate change, it was considered to apply organic mulches covering 25% of the soil, with the aim to reduce evaporation by 12%. Table 2 shows a comparative table of the proposed actions is presented.

Table 2. Adaptation actions proposed for barley in Perote, Ver.

Scenario number	Seedtime	Mulches	Description of the actions
1	April 15	0% organic materials	Baseline parameters
2	April 30th	0% organic materials	Planting date is modified but mulches remain as baseline
3	April 15	25% organic materials	Same planting date, increase in mulches

The results show an improvement in yields for the two proposed adaptation actions. Scenario 2 had the best performance for the period 2015–2039, which reports 240 and 120 kg/ha more than Scenario 1 and 2 respectively. For the middle horizon the two proposed adaptation actions show similar performance, reaching 150 kg/ha more than Scenario 1. For the far horizon adding mulches has a performance to reach 220 kg/ha above the Scenario 1. Scenario 1 corresponds to the same parameters as baseline, so the fact of implementing mulches or changing planting date could reduce yield loss by 203 kg/ha in average of all scenarios and time horizons (Fig. 7).

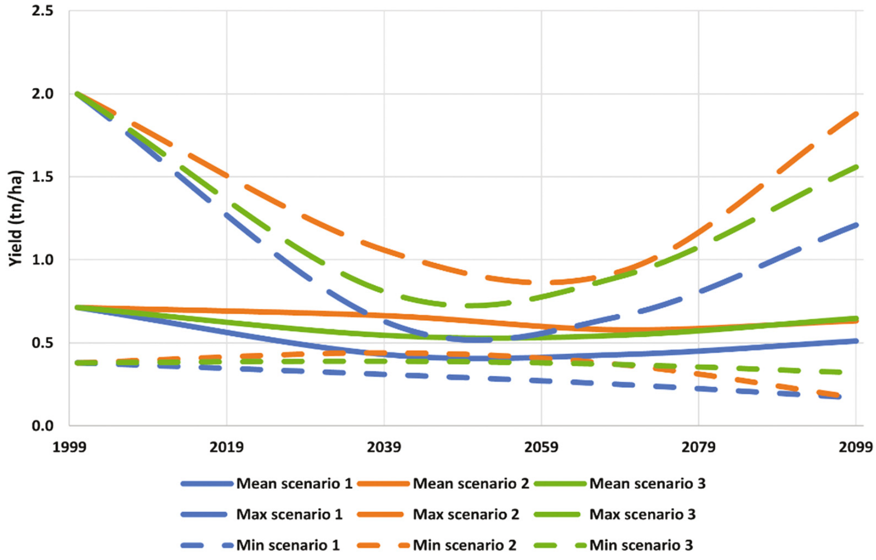


Fig. 7. Adaptation actions for barley in Perote, Ver.

Changing planting date to April 30, a yield of 0.63 tn/ha would be expected for the period 2075–2099, which implies more than 120 kg compared to baseline planting date. Modifying mulches, yields for horizon 2075–2099 would be 0.65 tn/ha, which puts only 20 kg above the measure of change of planting date. In terms of cost, it could be easier to adopt the change of planting date because it does not generate additional costs as Scenario 3.

4 Conclusions

The AquaCrop model proved to be a good method to estimate potential yields for crops under our study sites. The sensitivity of the program to the variation in precipitation was confirmed, which had been previously reported in the literature.

Crop yields had good fit compared to those observed by SIAP. Barley model in Perote had the best adjustment considering statistical indexes.

Under climate change scenarios, potential yields would suffer a generalized decrease. Yields of maize could fall by 50% at a distant horizon, according to the average of all models, while for barley this decline is by 40%.

Adaptation actions based on changing planting date were more effective in the case of maize. Although using organic mulches represents a good measure that can restore the natural fertility of the soil. For barley, yields under climate change scenarios could be around 0.51 tn/ha, while proposed actions can reach 0.65 tn/ha, it means 140 kg/ha more.

It is necessary to continue studying and exploring the potential impacts of climate change on crops, and thereby propose adaptation actions to decrease vulnerability. Future work should focus on assessing the performance patterns of all crops in the country. The information should assist decision-makers to guide and support the adaptation process to climate change.

References

Allen, R.G.: Evapotranspiración del cultivo: Guías para la determinación de agua de los cultivos. FAO, p. 297 (2006)

Cavazos, T., et al.: Actualización de escenarios de cambio climático para México como parte de los productos de la Quinta Comunicación Nacional. Informe Final del Proyecto al INECC, 150 pp. (2013). http://escenarios.inecc.gob.mx/index2.html

Cervantes, A., Angulo, V., Tavizón, F., González, R.: Impactos potenciales del cambio climático en la producción de maíz. Investigación y Ciencia (2014). http://www.investigacionyciencia.es/revistas/investigacion-y-ciencia/numero/429/el-futuro-de-la-energa-elica-8652

Conde, C., Ferrer, R., Orozco, S.: Climate change and climate variability impacts on rainfed agricultural activities and possible adaptation measures. A Mexican case study. Atmósfera **19**(3), 181–194 (2006)

Deb, P., Shrestha, S., Babel, M.S.: Forecasting climate change impacts and evaluation of adaptation options for maize cropping in the hilly terrain of Himalayas: Sikkim, India. Theoret. Appl. Climatol. 649–667 (2014). https://doi.org/10.1007/s00704-014-1262-4

Fernández Eguiarte, A., et al.: Actualización de los escenarios de cambio climático para estudios de impactos, vulnerabilidad y adaptación (2015)

Flores-Gallardo, H., Ojeda-Bustamante, W., Flores-Magdaleno, H., Sifuentes-Ibarra, E., Mejía-Saénz, E.: Simulación del rendimiento de maíz (Zea mays L.) en el norte de Sinaloa usando el modelo Aquacrop. Agrociencia **47**(4), 347–359 (2013)

García, Á.J., Riaño, H.N., Magnitskiy, S.: Simulation of corn (Zea mays L.) production in different agricultural zones of Colombia using the AquaCrop model. Agronomía Colombiana **32**(3), 358–366 (2015). http://revistas.unal.edu.co/index.php/agrocol/article/view/45939

Gay Garcia, C., Rueda Abad, J.: Reporte Mexicano de Cambio Climatico. UNAM-PINCC, Ciudad de México (2015). http://www.pincc.unam.mx/reportemexico/index.php

Heng, L.K., Hsiao, T., Evett, S., Howell, T., Steduto, P.: Validating the FAO aquacrop model for irrigated and water deficient field maize. Agron. J. **101**(3), 488–498 (2009). https://doi.org/10.2134/agronj2008.0029xs

INEGI: Conjunto de datos vectoriales edafológicos Serie II (2005)

INIFAP: Manejo integrado del cultivo de cebada en condiciones de temporal en San Luis Potosí (2011)

Mebane, V.J., Day, R.L., Hamlett, J.M., Watson, J.E., Roth, G.W.: Validating the FAO aquacrop model for rainfed maize in Pennsylvania. Agron. J. **105**(2), 419–427 (2013). https://doi.org/10.2134/agronj2012.0337

Ojeda-Bustamante, W., et al.: Impacto del cambio climático en el desarrollo y requerimientos hídricos de los cultivos. Agrociencia **45**(1), 1–11 (2011)

SAGARPA: SIAP, Sistema de Información Agropecuaria y Pesquera de México. Secretaría de Agricultura, Ganadería y Pesca (2009). www.siap.sagarpa.gob.mx

Smith, M., Steduto, P.: Yield response to water: the original FAO water production function, Rome, 10 p. (2012)

SIAP - SAGARPA: SIAP, Sistema de Información Agropecuaria y Pesquera de México. Secretaría de Agricultura, Ganadería y Pesca (2016). www.siap.sagarpa.gob.mx

Soddu, A., et al.: Climate variability and durum wheat adaptation using the AquaCrop model in Southern Sardinia. Procedia Environ. Sci. **19**, 830–835 (2013). http://linkinghub.elsevier.com/retrieve/pii/S1878029613003629

Steduto, P., et al.: Aquacrop-the FAO crop model to simulate yield response to water: I. concepts and underlying principles. Agron. J. **101**(3), 426–437 (2009)

Todorovic, M., et al.: Assessment of aquacrop, cropsyst, and WOFOST models in the simulation of sun flower growth under different water regimes. Agron. J. **101**(3), 509–521 (2009)

Vanuytrecht, E., Raes, D., Willems, P.: Regional and global climate projections increase mid-century yield variability and crop productivity in Belgium. Reg. Environ. Change **16**, 659–672 (2015). https://doi.org/10.1007/s10113-015-0773-6

Multitemporal Land Cover Change Analysis in the Upper Cauca Basin in Colombia

Cristian Valencia-Payan(✉), Edgar Leonairo Pencue-Fierro, Apolinar Figueroa-Casas, and Juan Carlos Corrales

Universidad del Cauca, Popayán, Colombia
{chpayan,leonairo,apolinar,jcorral}@unicauca.edu.co

Abstract. Reliable information about land cover is very important in environmental conservation processes and natural resource management. Establishing the types of land cover, how these covers are distributed, and the changes they have undergone over a period of time, are vital in revealing whether or not activities carried out in the region have affected the territory and how the landscape has changed. These in turn are essential to the generation of alternative land-use planning and broader environmental conservation processes. This study therefore presents an analysis of land cover in the Upper Cauca Basin in southwestern Colombia using land cover classifications based on remote sensing data obtained by Landsat satellites.

Keywords: Land covers · Change · Permanent crops · Transitory crops · Natural forest

1 Introduction

Information about land cover for sustainable management and policy development is more and more required [1] as this information helps in environmental protection, resource use, environmental change monitoring, and so on. In regions such as the Upper Cauca Basin (UCB) this information can be scarce. This is due to the effort demanded by supervised testing in carrying out field measurements, while satellite measurements are often complicated due to high cloud cover in the Andean region.

The UCB is located in the Inter-Andean region and has evolved as a region with one of the greatest agricultural productivities in Colombia [2]. The UCB is an area of no small strategic importance because it extends from the Colombian Massif to the limits of Caldas and Antioquia departments, between the western and central mountain ranges. Its altitude ranges from 5,400 to 950 m above mean sea level (m.a.m.s.l.). The UCB incorporates 99 municipalities and has 191 sub-basins that flow into the Cauca River. It is currently experiencing substantial deterioration, mainly due to inadequate land use, deforestation, water pollution from anthropogenic sources, and draining of swamps and wetlands for the expansion of the agricultural frontier, especially in the flat area.

The main challenges in the basin are related to efficient water resource management. For the departments of Cauca and Valle del Cauca the use of water is essential for

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_16

carrying out agricultural activities, which demand 42% and 75% of the resources in each department, respectively [3], where such a situation is strongly linked to land use and land cover changes in the basin.

Changes in land use and land cover are important in understanding the natural dynamics of the land [4]. Research into change dynamics can help to give us a better understanding of how the ecosystems and the land cover conservation process are affected by human activities. Land cover classification is one of the most relevant tasks to understand the current state of a region [4]. Using this classification, studies can be carried out related to the conservation of natural systems, assessing the possibilities of maintaining ecosystem services, and evaluating the capacity for productive activities. For these reasons and to evaluate the existing environmental supply, multitemporal analysis using information obtained through remote sensing was carried out. It covered a wide geographical area of strategic interest, being able to determine the changes undergone in 26 years. This length of period was established so that reliable measures could be taken [5]. This information plays a determinant role in the quantification and understanding [6] of landcover change, the effects caused by nature and man in the expansion-production-transformation processes in the settlement areas.

In this study, using four classification sets from Landsat 4, 5 and 7 data, in a window of 25 years, the changes presented in the vegetation cover attributed to natural forest and to crops (permanent and Transitional) in the departments of Cauca, Caldas, Risaralda, Quindío, and Valle del Cauca, are analyzed. The UCB covers about 3 million hectares in this region (1 million in Cauca, 1.1 million in Valle del Cauca, 400,000 in Caldas, 239,000 in Risaralda and 193,000 in Quindío). The importance of this study is that it has not been done before for this region. In addition, it has a high degree of specificity compared to the data obtained by the country’s governmental authorities and may be used as input for territorial planning and environmental conservation processes.

2 Related Work

Changes in natural systems can be determinants for extreme events such as droughts, floods, landslides and population displacement. Such events may be triggered by natural or anthropic processes affecting vegetation cover that would otherwise act as a regulator to avoid these unwanted phenomena [4]. These events can be predicted by tracking the landcover changes in recent years to determine whether or not these might be a future hazard. In [4] a new method to perform land cover classifications from the Landsat spectral bands and vegetation indices as a byproduct of the processing was presented. In [5] a method for classification and analysis of land cover changes using machine learning algorithms and Principal Component Analysis was presented- The researchers make a comparison between the area occupied by the cover detected in the two years studied.

In [6] a land cover change detection method using remote sensing data from the Landsat, IKONOS and SPOT platforms was presented. This study was conducted between 1986 and 2013 in two Saudi Arabian cities, principally to detect changes due to the expansion of these cities in the surrounding areas. In [7] a land cover change analysis between 1975 and 2007 using Landsat images 2, 5 and 7 was performed. In addition, auxiliary data from the ALOS and SPOT platforms were used, finding strong changes in the 9 classified images. In [8] a study of the changes in mangrove cover in a 25-year time window from radar images was performed, finding that the changes in this vegetation is governed by the tide cycles of flood and retreat.

In [9] a land cover classification using multiple sources for the management of forest fires was presented. Nine types of land covers were differentiated, grouping those that serve as fuel during forest fires. In [10] a model for the detection of coverage changes in a theoretical manner, considering several fundamental aspects that lead to land cover changes, the detection of the focus of change, and other related aspects, was presented. In [11] a mapping of land cover from Landsat images was presented, achieving a high separability between soil types and other types of vegetation cover.

As can be observed, there are few studies in which a comparison between the covers present in the area in a given period of time has been made. Most of the work has focused on the classification of the present land covers, leaving aside the study of the land cover changes. Those who have done studies on land cover changes have done so in regions of relatively small size or focused on a single cover. In this study, we will perform the analysis on regions of a considerably large size, in multiple land covers and in multiple years based on the classification model presented in [4].

3 Materials and Methods

For the data extraction for classification, a specific technique was developed supported in the OpenCV and Gdal libraries. With this, spectral data and the georeferenced information are loaded and processed. This technique is oriented to the temporal comparison of land cover and the obtaining of spatial and temporal statistics. Figure 1 shows the general diagram of the process carried out.

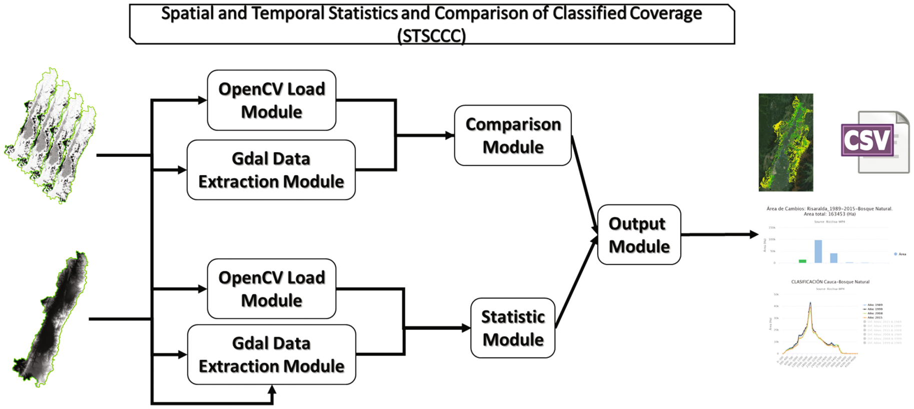

Spatial and Temporal Statistics and Comparison of Classified Coverage (STSCCC)
OpenCV Load Module
Gdal Data Extraction Module
Comparison Module
OpenCV Load Module
Gdal Data Extraction Module
Statistic Module
Output Module
CSV
Área de Cambios: Risaralda_1989-2015-Bosque Natural.
Area total: 163453 (Ha)
Source: Ricclisa-WP4
Área
CLASIFICACIÓN Cauca-Bosque Natural
Source: Ricclisa-WP4
Año: 1989
Año: 1999
Año: 2008
Año: 2015

Fig. 1. General scheme for the proposed technique.

3.1 Techniques

As shown in Fig. 1, the technique developed to obtain the information that will be used in this study starts loading the required data. It is then processed to extract the information of interest. That data is subsequently exported in different formats for later analysis.

Spatial and Temporal Statistics and Comparison of Classified Coverage (STSCCC)
The Statistic Module performs the disaggregation by altitudes for each year of the area occupied by each cover in user-defined steps. Multi-year comparison of a selected classified cover with the remaining classified cover is also done. As a result, graphics with data discriminated by altitudes and CSV files containing the collected information are presented, as are: a scenario of a possible future state of land cover according to the behavior presented in the years analyzed; a color-coded map with the cover changes; and statistics graphs of the changes found. To perform this process a mask is created from a Digital Elevation Map (DEM) of the region studied. This mask contains the pixels that are within the range of altitudes according to the step selected. Then, pixel-to-pixel information of land cover information sorted for each year and the altitude mask data are multiplied. Once this is done, information that does not match the range of altitudes is eliminated and the statistical information for each year is obtained.

The Comparison Module performs a comparison of two-time points of the same land cover of interest by discriminating between: (1) the area gained by each land cover, (2) the area of the three land covers that gained most over the cover of interest, (3) the area of conservation, (4) the area that became bare soil, and (5) the area which became the remaining covers. The comparison is done pixel by pixel taking into account that clouds, cloud shadows, and bodies of water for each classification are not part of obtaining statistics.

3.2 Data

Land Cover Classification Data
The UCB land cover classification was obtained from the classification method of [4], corresponding to years 1989, 1999, 2008 and 2015, which was calculated from Landsat 4, 5 and 7 satellites.

In Fig. 2, we can see the location of the UCB, in the Republic of Colombia, that comprises part of the departments of Cauca, Valle del Cauca, Quindío, Risaralda, and Caldas. The limits of the UCB are observed in fluorescent green color. Table 1 lists the Landsat images used in the UCB classification process. Because Path/Row 009/056 does not have an adequate image in the years 1999 and 2008, the departments of Caldas and Risaralda will have comparisons only between the years 1989 and 2015.

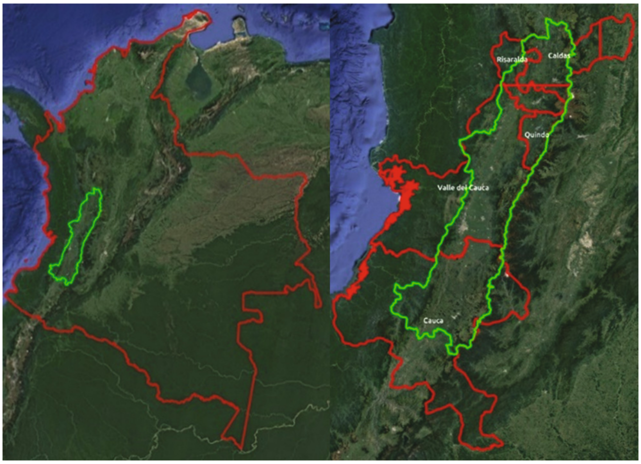


Fig. 2. Location of the UCB in Colombia and its departments.

Table 1. List of Landsat images.

Year	Path/Row		
	009/056	009/057	009/058
1989	LT4/1989219	LT4/1989219	LT4/1989219
1999	None	LT5/1999191	LT5/1999191
2008	None	LE7/2008256	LE7/2008256
2015	LE7/2015003	LE7/2015003	LE7/2015003

Validation Data

As validation data, the official information of the Colombian Environmental Information System (SIAC, Spanish acronym) [12] was used. This information was collected by the different agencies and govern entities whose objectives are the monitoring and environmental control of regions in the country.

Digital Elevation Map (DEM)

To obtain the disaggregated by altitude the DEM generated by NASA from Shuttle Radar Topography Mission data were used [13]. This has a resolution of approximately 30 m. These data are freely accessible and have global coverage. The DEM was produced in 2000 and released globally in 2015. This has various post processing stages applied by the NASA to ensure the quality of the final product.

4 Results

STSCCC

In the following, the most representative results are presented when applying the technique developed in the regions that are part of the UCB for the Natural Forest and Crops (Permanent and Transitional) land covers.

Natural Forest

As can be seen in Fig. 3, the area of Natural Forest has decreased in the department of Caldas. The graph demonstrates that all areas of the department have experienced deforestation processes, losses of forest being greater between 1000 and 2800 m.a.m.s.l.. The changes were found mainly to be with Grassland, Planted Forest and Bare Soil.

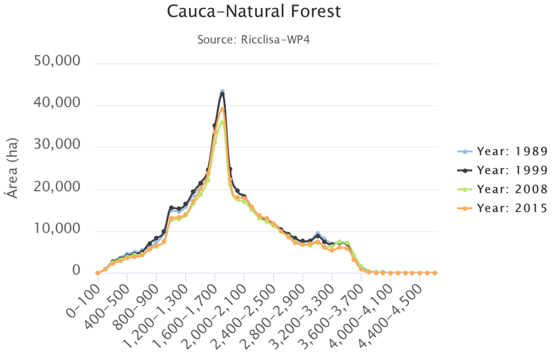


Fig. 3. Discrimination by altitude for Natural Forest land cover in the department of Caldas.

Figure 4 shows the behavior of Natural Forest in the department of Cauca. Between 1989 and 2008, there was a progressive decline in this land cover, which can be seen most clearly at the peak between 1700 and 1800 m.a.m.s.l., however, in the years between 2008 and 2015 there was a slight recovery in all the altitude ranges

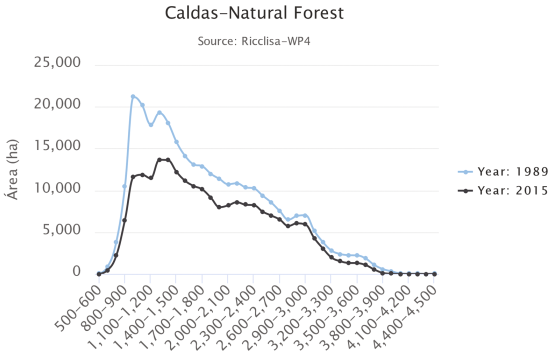
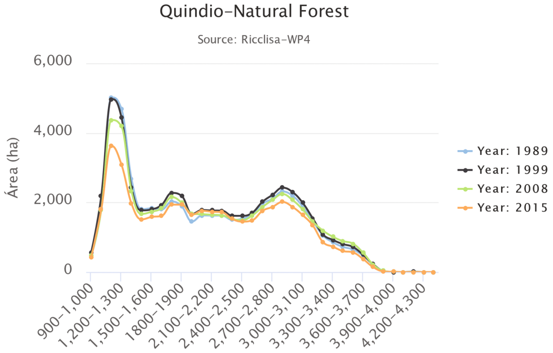


Fig. 4. Discrimination by altitude for Natural Forest land cover in the department of Cauca.

discriminated, except for the values registered between 3100 and 4000 m.a.m.s.l.. The area gained has been mainly in Grasslands (79,000 ha), Bare Soil (9,000 ha) and lastly the Paramos (3,600 ha). Paramos, however, expanded in 5,000 ha, replacing Natural Forest.

For the department of Quindío, Fig. 5 reveals how there has been a progressive decrease in Natural Forest in the four years selected. It is also observed that between 1989 and 1999 this land cover was almost unchanged. Between 2008 and 2015, there were changes in Natural Forest, suggesting that during this period the department was undergoing a strong transformation in its productive activities. Much of the area of change has occurred with Grasslands, which could indicate the increase in activities related to livestock. In addition, the Paramos have gained more than 1,000 ha of forest over the same period.

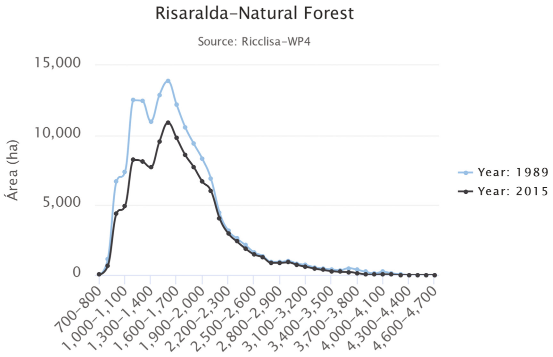


Quindio–Natural Forest

Source: Ricclisa–WP4

Fig. 5. Discrimination by altitude for Natural Forest land cover in the department of Quindío.

In the department of Risaralda, Natural Forest (Fig. 6) has decreased mainly in the area of its most significant altitudes, between 1000 and 2300 m.a.m.s.l.. The distribution of this land cover is seen to be similar regardless of the amount of area lost between



Risaralda–Natural Forest

Source: Ricclisa–WP4

Fig. 6. Discrimination by altitude for Natural Forest land cover in the department of Risaralda.

1989 and 2015. This is an abnormal behavior for which to be able to find a reason would require additional information.

Again, the greatest interaction occurs between Grasslands and Natural Forest, indicating the increase in livestock activities, or deforestation to use land in agricultural activities, the latter because Permanent Crops in the department increased by 60%, as will be seen later.

Figure 7 shows Natural Forest behavior in the department of Valle del Cauca, whose concentration zones are located at 1000–1100 m.a.m.s.l.. Here it is observed that between 1989 and 1999, land cover decreased by around 8,000 ha and continues to decline, losing about 5,000 ha in 2015. It is further observed that in the other regions its behavior was stable, except for the range of 2800–3300 m.a.m.s.l., where there was a fall in 2015 of the number of hectares of this land cover. Interaction between Grasslands and Natural Forest is also very strong in this case, having very close values in the area gained from each cover. However, it is Permanent Crops and Planted Forest that have gained most from Natural Forest. This is a strong indicator of the increase of sowing activities in the department both of crops and forests to obtain raw materials.

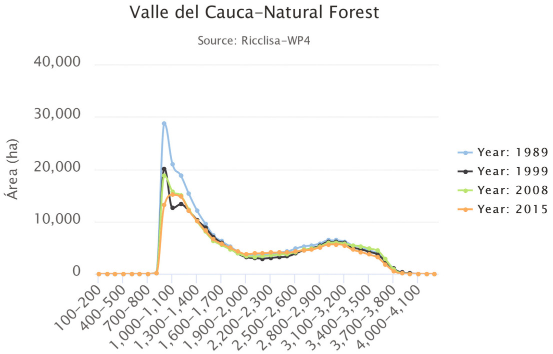


Fig. 7. Discrimination by altitude for Natural Forest land cover in the department of Valle del Cauca.

Crops (Permanent and Transitional)

As we can see in Fig. 8, the area of Transitional Crops at each step of altitude selected has decreased in the department of Caldas, a behavior related to that shown in Fig. 3a, except for the high zones (almost 4000 m.a.m.s.l.) where it was observed that there had been a gain in the Paramos. Permanent Crops was observed to be concentrated between 900 and 1000 m.a.m.s.l. and increased in this region by around 1,300 ha.

For Permanent Crops (Fig. 9a), peak activity is between 900 and 1000 m.a.m.s.l., decreasing toward 1200 m.a.m.s.l. It is observed that it has been increasing since 1989, which is in line with the data obtained previously for this land cover. Figure 9b shows the variability of Transitional Crops compared to 1989. In all ranges discriminated, no year was able to surpass the number of hectares associated with this land cover in 1989. Transitional Crops decreased considerably in 2015, except for some regions, compared to 1999 and 2008, where there was a similar behavior in the different discriminated ranges.

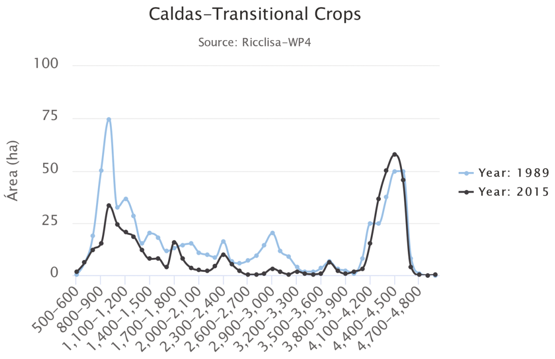
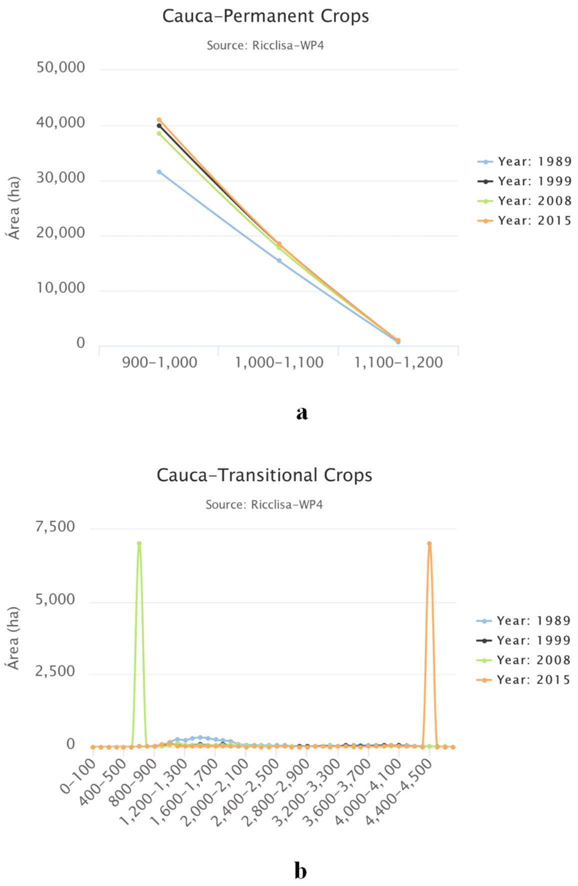


Fig. 8. Discrimination by altitude for Transitional Crop land cover in the department of Caldas.



a

b

Fig. 9. Discrimination by altitude for (a) Permanent Crops and (b) Transitional Crops in the department of Cauca.

For the department of Quindío (Fig. 10) a behavior with continuous changes was observed in the year of greatest area dedicated to Transitional Crops, concentrating between 1000 and 1200 m.a.m.s.l. and 3,500 and 4500 m.a.m.s.l., No information on Permanent Crops was found for this department in the classification carried out, likely due to high cloud cover in the region.

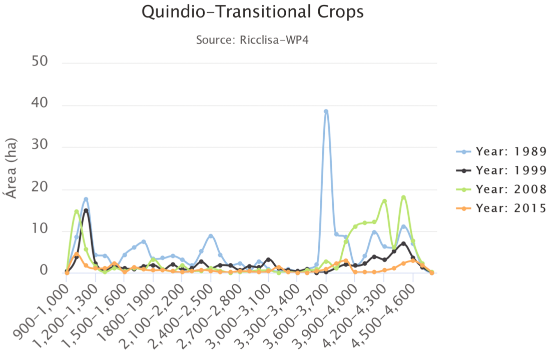


Fig. 10. Discrimination by altitude for Transitional Crops in the department of Quindío.

In the department of Risaralda, for Transitional and Permanent Crops (Fig. 11), similar behavior occurs: a first concentration zone between 900 and 1700 m.a.m.s.l. in the case of Transitional Crops, decreasing in almost all discriminated ranges and increasing between 1500 and 1700 m.a.m.s.l.. In the same way, Permanent Crops presented a behavior similar to that observed in the department of Caldas, concentrating between 900 and 1000 m.a.m.s.l. and increasing in the selected period of time.

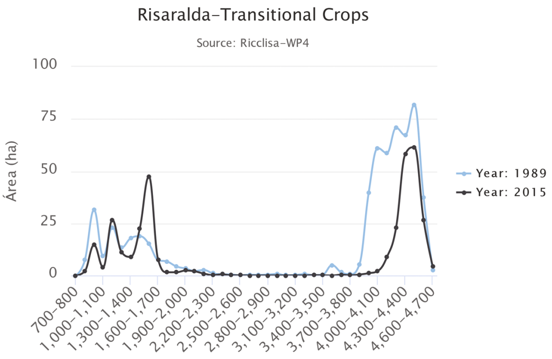
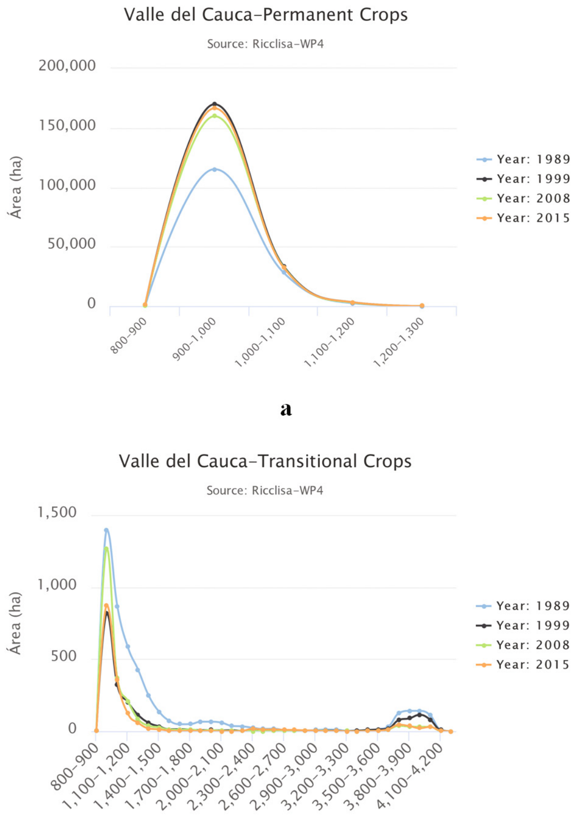


Fig. 11. Discrimination by altitude for (a) Permanent Crops and (b) Transitional Crops in the department of Cauca.

Permanent Crops (Fig. 12a) are located in the Valle del Cauca mainly between 800 and 1100 m.a.m.s.l.. Since 1989, the area occupied can be seen to have increased, peaking in 1999, a slight decrease in 2008 with a subsequent recovery in 2015. In other discrimination ranges, behavior after 1989 appears to be constant. The difference presented in the concentration zone could be due to the fact that, during the periods of image capture, the region was at rest after harvest, or at the beginning of the planting process. In the case of Transitional Crops (Fig. 12b), the peak is observed between 1000 and 1100 m.a.m.s.l.. Declining progressively since 1989, a slight peak is also found between 3600 and 4200 with a behavior similar to that observed in Fig. 8.



a

b

Fig. 12. Discrimination by altitude for (a) Permanent Crops and (b) Transitional Crops in the department of Valle del Cauca.

As could be seen, information discriminated by altitude helps to corroborate the results obtained in the process of comparison between land covers, showing behavior previously estimated in terms of the losses and gains of land covers of interest. Finally, to verify the results obtained in the classifications and to determine if the information presented achieved a greater degree of specificity, the information of land cover change of Natural Forest was compared during 1989–1999 with the exchange information for this cover in the window 1990–2000 obtained by the SIAC in Valle del Cauca.

Validation Process

Validation was performed by pixel-to-pixel comparison between the change map obtained from the SIAC and the change map generated by the STSCCC technique. Initially it can be observed that the SIAC map discriminates just five types of changes: Forest points, deforestation points, points without information, regeneration points and a no stable forest point, the latter occupying most of the department of Valle del Cauca. In the data obtained from STSCCC, six types of changes are obtained, as detailed above. Figure 13 shows images of SIAC and STSCCC with their labels.

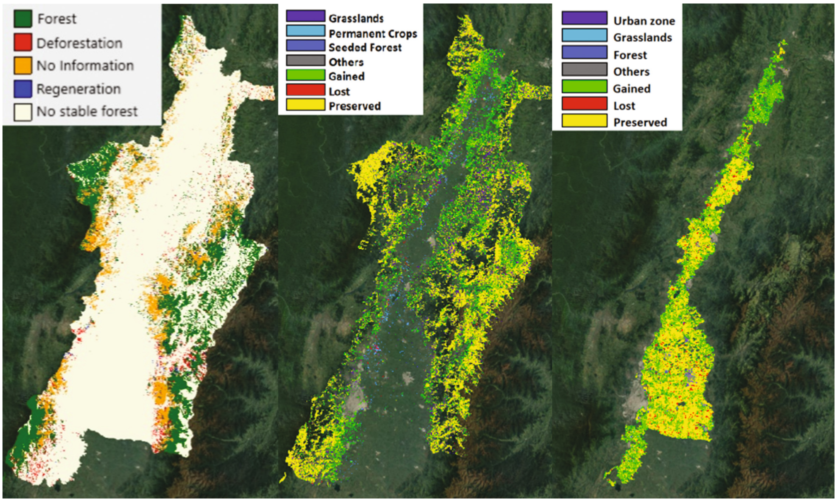


Fig. 13. Comparison of change information in Forest of (a) SIAC (b) STSCCC Natural Forest and (c) STSCCC Permanent Crops.

In the SIAC data for the department of Valle del Cauca, Fig. 10, comparison found that the classification of Forest and Regeneration of SIAC coincide with the classes Gained and Preserved of STSCCC, a high coincidence with the Deforestation classification of SIAC joining the ranks Grasslands, Permanent Crops, Planted Forest, Others and Lost. In the case of STSCCC, the No Stable Forest label of SIAC coincided with information on hectares gained by Natural Forest, Grasslands, and Permanent Crops as shown in Fig. 13, detailing the classification of SIAC and the STSCCC classification for Permanent Crops. Similarly, the SIAC No Information label was matched with hectares by the aforementioned covers. Similar behaviors were observed in the other departments of UCB.

5 Conclusions

This study conducted a multi-year comparison for three of the most important land covers in the Upper Cauca Basis (UCB) in Colombia. The resulting information could be used, in a complementary way with information on land use, for decision making or in the creation of new environmental policies by government entities. The data can also be used to obtain a quantitative indicator of the environmental policies previously applied.

The information obtained reveals how land use has increased for agricultural activities in the five departments in the UCB, with important impacts on Natural Forest in some departments. Globally, close gains have been the norm, so that it could be viewed that statistically land covers have experienced a relatively stable behavior in the UCB. Nevertheless, this does not mean these changes have not presented significant environmental consequences. In addition, it was observed how Transitional Crops have decreased in the altitudes of more than 3,500 m.a.m.s.l., making way for other land covers. The displacement of cover attributed to the Paramos ecosystem was further observed, Paramos currently yielding space to Transitional Crops across all departments. Such practice puts this fragile ecosystem at serious risk.

In the time window, a clear focus of all the departments towards the planting of Permanent Crops was observed. This is related to the increase particularly of cane and coffee crops in the region. Likewise, livestock activities or the planting of Grasslands are an important part of the observed change.

The ability to observe changes in land covers selected for this study as well as other covers not shown, by discrimination by altitude, will allow entities in charge of environmental control to have a better vision of the land cover distribution by departments, by municipalities, or by watershed areas. This will improve decision making on the use of resources and the protection of regions, among others.

Finally, when comparing the available information obtained from the SIAC, the presented study was found to achieve a higher degree of discrimination, distinguishing between multiple land cover changes observed. This allows a greater understanding of the changes that have occurred in the selected period of time.

Acknowledgments. The Authors are grateful to the Optics and Laser Group (GOL), the Telematics Engineering Group (GIT), the Environmental Studies Group (GEA) of the University of Cauca, The University of Cauca, the RICCLISA program and the WP4–Covers project for supporting this research. We are especially grateful to Colin McLachlan for suggestions related to the English text.

References

1. Jansen, L.J., Di Gregorio, A.: Obtaining land-use information from a remotely sensed land cover map: results from a case study in Lebanon. Int. J. Appl. Earth Obs. Geoinf. **5**(2), 141–157 (2004). doi:10.1016/j.jag.2004.02.001
2. Colombia: sus principales cultivos y zonas agroecológicas. http://www.agro20.com/group/agricultura-america-latina-y-caribe/forum/topics/colombia-sus-principales-cultivos-y-zonas-agroecol-gicas
3. Departamento Nacional de Planeación: Programa para el Saneamiento,. Manejo y Recuperación. Ambiental de la Cuenca Alta del. Río Cauca. Bogotá, pp. 9–10 (2009)

4. Pencue-Fierro, E.L., Solano-Correa, Y.T., Corrales-Munoz, J.C., Figueroa-Casas, A.: A semi-supervised hybrid approach for multitemporal multi-region multisensor landsat data classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **9**(12), 5424–5435 (2016). doi:10.1109/jstars.2016.2623567
5. An, J., Zhou, L., Yang, Z., Zhang, H.: Land-cover classification and change analysis of Qinghai-Tibet Highway by remote sensing data. In: MIPPR 2009: Remote Sensing and GIS Data Processing and Other Applications (2009). doi:10.1117/12.832877
6. Alqurashi, A.F., Kumar, L.: Land use and land cover change detection in the Saudi Arabian desert cities of Makkah and Al-Taif using satellite data. Adv. Remote Sens. **03**(03), 106–119 (2014). doi:10.4236/ars.2014.33009
7. Bagan, H., Takeuchi, W., Kinoshita, T., Bao, Y., Yamagata, Y.: Land cover classification and change analysis in the Horqin sandy land from 1975 to 2007. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sen. **3**(2), 168–177 (2010). doi:10.1109/jstars.2010.2046627
8. Cohen, M.C., Lara, R.J.: Temporal changes of mangrove vegetation boundaries in Amazonia: application of GIS and remote sensing techniques. Wetlands Ecol. Manage. **11**(4), 223–231 (2003)
9. Koetz, B., Morsdorf, F., Van der Linden, S., Curt, T., Allgöwer, B.: Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. For. Ecol. Manage. **256**(3), 263–271 (2008). doi:10.1016/j.foreco.2008.04.025
10. Lambin, E.F.: Modelling and monitoring land-cover change processes in tropical regions. Prog. Phys. Geogr. **21**(3), 375–393 (1997). doi:10.1177/030913339702100303
11. Milne, A.K., O'Neill, A.L.: Mapping and monitoring land cover in the Willandra Lakes World Heritage Region. Int. J. Remote Sens. **11**(11), 2035–2049 (1990). doi: 10.1080/01431169008955159
12. Sistema de Información Ambiental de Colombia. http://sig.anla.gov.co:8083/
13. Shuttle Radar Topography Mission. https://www2.jpl.nasa.gov/srtm/

A Smart Farming Approach in Automatic Detection of Favorable Conditions for Planting and Crop Production in the Upper Basin of Cauca River

Iván Darío López(✉) and Juan Carlos Corrales

Grupo de Ingeniería Telemática (GIT), Universidad Del Cauca, Popayán, Colombia
{navis,jcorral}@unicauca.edu.co

Abstract. The upper basin of Cauca River is one of the most important agricultural and industrial corridors in Colombia. In addition, Colombian relief has a variety of soils and climatic zones, in which the slightest modification of agro-climatic conditions could imply a great impact on sowing, production, yield, crop quality, and hence, the farmers' income sources. However, these conditions are not the only factors that influence a crop growth; dimensions like biophysical, economic-productive, political-institutional, and socio-cultural establish interrelations that affect the viability of planting and produce a crop in a geographic territory. For a more efficient information management, Smart Farming represents the application of information and communication technologies in order to support more intelligent decisions in the agricultural sector. This study presents a theoretical proposal based on Big Data Analytics to detect appropriate conditions for sowing and crop production in the upper basin of Cauca River. Similarly, the preliminary public datasets (corresponding to biophysical and economic-productive) and main proposal components are described in a general view. Finally, a future implementation and evaluation of this research work is proposed.

Keywords: Agro-climatic conditions · Automatic detection · Big Data Analytics · Crop planting · Crop production · Data fusion · Smart Farming

1 Introduction

The agricultural sector represents approximately 40% of Colombian exports, where about 21% of the population depends directly on agriculture as a source of employment [1]. Furthermore, climatic variability impacts mostly in countries located within tropical and subtropical regions, which report the highest rates of poverty and hunger [2]. Colombian relief has a variety of soils and climatic zones; in this way, the slightest modification of agro-climatic conditions could imply a great impact on sowing, production, yield, and crop quality. This increases the vulnerability of a crop and becomes a direct threat to food security and income sources of a farmer.

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_17

In order to determine the effects of climatic variability on production and yields of some selected crops, the National Climate Change Adaptation Plan [3] assessed agro-climatic risk by sector. Results showed that production and yields of most crops tend to decrease significantly due to high temperatures as a consequence of thermal and water stress, shortening of the growing season, and the increase of pests and diseases. However, these are not the only factors that influence a crop; there are different dimensions such as biophysical, economic-productive, political-institutional, and socio-cultural, which establish interrelations that affect the viability of planting a crop in a geographic territory, as well as the expected level of production and crop yield [4]. In this sense, for better models performance, capture of strong spatio-temporal database of various biophysical and socioeconomic aspects is necessary [5].

For a more efficient management of the information available in different productive processes developed in agriculture, the concept of Smart Farming [6] arises. This concept represents the application of Information and Communication Technologies (ICT) in the agricultural sector. These new strategies include different technological paradigms such as GPS services, sensors, processing of satellite images, Big Data, among others. This theoretical proposal focuses on the application of the Big Data value chain to biophysical and economic-productive dimensions mainly due to the availability of datasets. In these two dimensions, data availability is more significant than the political-institutional and socio-cultural, where the amount of data is small and difficult to obtain.

Based on these assessments, the present study focuses on detecting appropriate conditions for sowing and crop production in Colombia, considering characteristics such as variety, value, and viability of the data in the context of Smart Farming. The remainder of this paper is organized as followed: Sect. 2 presents the related works. Section 3 defines the study area and the possible data sources for testing. Subsequently, in Sect. 4, the general Smart Farming proposal is described and finally, Sect. 5 highlights the conclusions and future work of this study.

2 Related Works

2.1 Data Fusion for Sowing and Agricultural Production

Different studies use fusion and data integration techniques with the aim of optimizing the planning and monitoring of agricultural activities and practices based on crop growth systems. In [7], a case study for barley cultivation in Finland is presented. The authors propose an architecture around a Database Management System (DBMS) to store data generated by a Wireless Sensor Network (WSN). The WSN links the DBMS to three sources of heterogeneous data: (a) a static sensor located at a given point, (b) portable sensors that can be moved in a given zone, and (c) static telemetry systems with multiple sensor readings. In the same line of research, [8] proposes a method for data fusion process in the case study of mushroom cultivation. The data sources are basically multiple climate sensors which measure temperature and humidity, and multiple soil sensors which measure resistivity.

On the other hand, in [9] medium and high resolution satellite images are used to determine crop yield. The data fusion procedure is based on combining time series of variables such as temperature and precipitation, with climate data expressed in satellite images of the National Oceanic and Atmospheric Administration (NOAA). In the same way, in [10] a multi-sensor imaging and data fusion system is developed to manage oil palm plantations in Malaysia. In this case, data fusion is relevant taking into account the types of data sources, which are complementary and simultaneously contribute to obtain a more general vision of the plantation.

2.2 Machine Learning for Sowing and Agricultural Production

The study proposed in [11], presents a decision support system prototype for yield prediction of rice crops in Maharashtra, India. The implemented model corresponds to a set of association rules. It allows to select variables such as precipitation, evapotranspiration, minimum, average, and maximum temperature, to predict the target class; in this case, low, moderate, or high yield. Studies like [12, 13] use Artificial Neural Networks (ANN) to estimate climatic variables in a geographic zone. Similarly, in [14] an ANN is used to predict crop yields by detecting various soil and climate parameters. The model parameters were water depth, soil type, temperature, atmospheric pressure, rainfall, humidity, nitrogen, phosphate, potassium and organic carbon.

On the other hand, in [15] a model of forecast of food shortage in Uganda is presented. In this study, more accurate results were obtained using Support Vector Machines (SVM) and k Nearest Neighbors (kNN) techniques. Similarly, in [16] an intelligent tool for rice yield prediction was developed. The SVM was based on a kernel method and it was used to find clusters based on weather data. Gandhi [17] proposes a model for predicting rice yield in India. This study describes the application of the Support Vector Machines using Sequential Minimal Optimization (SMO). In conclusion, it was established that classifiers such as Naive Bayes, and Multilayer Perceptron obtained better performance than SMO using a dataset of rice cultivation.

Some studies apply Multiple Classifier Systems (MCS) as in [18], a bootstrap-based assembly method called *UChooBoost*. The objective was to determine which seed lots of eucalyptus are the best for soil conservation in a seasonally dry soil. The observed results showed that there was no significant difference in the accuracy of this new algorithm over traditional bootstrap assemblies. Finally, in [19] *AdaSVM* and *AdaNaive* methods are proposed as models of assemblies to project the production of rice, cotton, sugar cane, among other crops in a period of time.

2.3 Big Data Analytics for Sowing and Agricultural Production

Big Data applied in Smart Farming environments is a recent focus of research [20], several works have begun to generate knowledge in this subject. The study referenced in [21], uses simulation models to establish recommendations on adaptation of crop management practices. In [22] an analysis of vulnerability and adaptation opportunities is made taking into account the specific biodiversity of maize crop (genetic variability between individuals of the same species). Similarly, Badr [23] establishes a proof-of-concept case study, and investigates the coupling degree of a scalable geospatial data platform, the Physical Analytics Integrated Repository and Services (PAIRS), to DSSAT system. For geospatial analysis, PAIRS provide a way to retrieve heterogeneous data sources in order to simulate crop models using hundreds of terabytes of data.

3 Study Area

Cauca River Basin is located at an altitude of approximately 3200 m above sea level (masl). It has a length of 1360 km and flows into the Magdalena River, in the Department of Bolivar. It crosses from south to north nine departments (Cauca, Valle del Cauca, Quindío, Risaralda, Caldas, Antioquia, Cordoba, Sucre, and Bolivar). This basin is the second largest waterway of Colombia, representing approximately 41% of the Colombian population. It is subdivided into three large regions, the upper, middle and lower basins. The upper basin (which is the subject of this study) comprises an area of 3031488 ha (Fig. 1). This zone extends from the *Macizo Colombiano* to the boundaries of Caldas and Antioquia departments, between the peaks of the western and central mountain ranges. Its altitude ranges from 5400 masl in the perpetual snows of the central mountain range and 950 in the Cauca alluvial valley. The study area comprises 191 sub-basins that flow into the Cauca River, with 99 municipalities and a population of approximately 7056000 inhabitants [4].

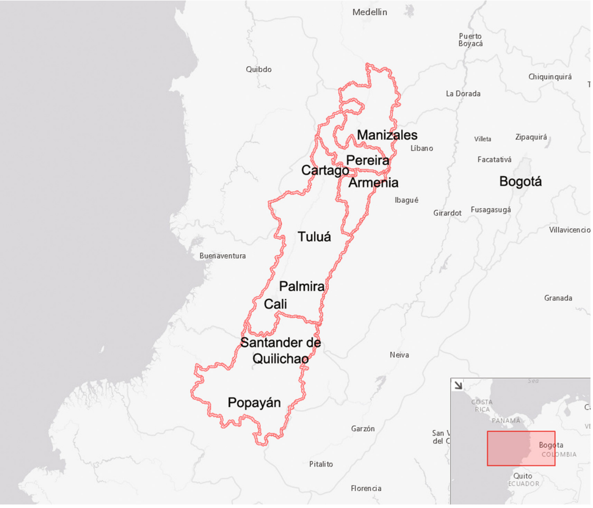


Fig. 1. Study area – upper basin of Cauca river. Taken from [4].

Representative economic activities include agriculture, livestock, fish farming, human consumption, among others. The agricultural and industrial corridor located in this area develops traditional agriculture, coffee cultivation, sugar cane, bean, banana, citrus, cacao, maize, avocado, among others. In addition, approximately 20% of the national population benefits from the environmental supply of this basin resources.

Furthermore, to develop the present proposal different data sources are required as a fundamental input. Each data set must be preprocessed and analyzed in order to determine its viability of incorporation into the model. In this way, the currently available preliminary datasets have public access and these are described below.

- **Agronet.** Ministry of Agriculture database. It collects data on production and yield of more than 200 crops in Colombia [24].
- **AVA.** Agriculture, Vulnerability and Adaptation (AVA) is a methodology for measuring the vulnerability of the agriculture sector. It is a project from the Climate and Development Knowledge Network (CDKN) [4].
- **CVC.** The Valle del Cauca Autonomous Regional Corporation is one of the main environmental authorities in the Upper Cauca river basin. This institution pro-vides water quality data for environmental sustainability [25].
- **DANE.** The National Administrative Department of Statistics applies different censuses to obtain data from different aspects of the population in Colombia, among them, the national agricultural census [26].
- **Datos Abiertos Colombia.** This governmental web platform collects public data in Colombia from different sectors, without restrictions and for specific use in research projects [27].
- **DNP.** The National Planning Department is an entity that defines the implementation of a strategic vision of the country in the social, economic and environmental fields [28].
- **FAOSTAT.** This web platform contains data on agriculture producer prices. These are prices received by farmers for primary crops as collected at the point of initial sale. Monthly data are provided from January 2010 to December of the previous year for over 60 countries and about 200 commodities [29].
- **SAC.** The Farmers Society of Colombia promotes national agricultural development and the welfare of the Colombian farmer. In addition, it promotes the scientific investigation and the diffusion of more advanced and new methods of cultivation [30].
- **Meteoblue.** It is a meteorological service created at the University of Basel, Switzerland, in cooperation with the U.S. National Oceanic and Atmospheric Administration and the National Centers for Environmental Prediction. It provides satellite weather data for any length and latitude within a given territory [31].

4 Automatic Detection of Favorable Conditions for Planting and Crop Production

The agricultural sector focuses its efforts on the use of ICT as support of decision-making in its productive processes. Therefore, Smart Farming is the main element in the present proposal in order to benefit from the potential of its different technological paradigms and generate a positive impact on the sowing and crops production in Colombia. One of these paradigms is Big Data, a concept widely disseminated today by the scientific and technological community, but in which many authors differ in giving a concerted definition. Around this theme there is an extensive documentary reference that allows forming an appropriate knowledge base to provide an alternative solution to the problem previously raised. Figure 2 presents the general architecture of this proposal and its main components are described below.

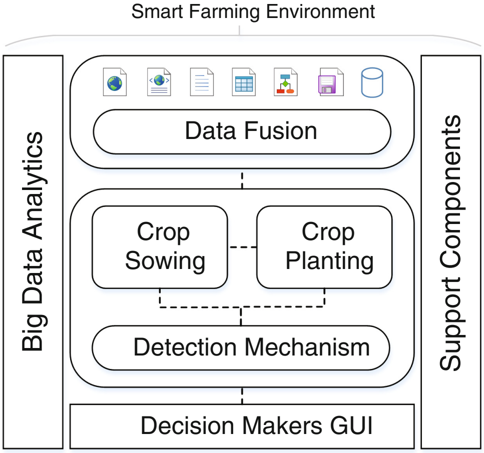


Fig. 2. General architecture for automatic detection of planting and crop production.

4.1 Big Data Analytics

The Big Data value chain is composed of four key elements: Generation, Acquisition, Storage, and Analysis [32]. Although some items belonging to the previous elements may be addressed within the development of this proposal, the main focus is the analysis component, which includes different tools or analytical methods for inspecting, transforming and modeling data. In this way, it is possible to extract value, useful conclusions, and support decision making.

Taking into account that agriculture is one of the areas where more data is currently being generated, it is necessary to focus different research efforts towards the development of analysis mechanisms that allow the integration of the most important characteristics of Big Data. In order to optimize systems for decision-making in the agricultural sector, some of these characteristics are established in the so-called seven V model: volume, velocity, variety, veracity, value, viability, and visualization. It is important to clarify that to delimit the scope of this work, only the characteristics of variety, value and viability will be addressed.

On the other hand, data analysis can be classified into three levels according to the depth of the analysis processes, these levels are described below.

- **Descriptive Analysis.** It is based on historical data to describe what has happened. It is often useful for finding basic trends in data sets.

- **Predictive Analysis.** It focuses on predicting probabilities or future trends, as well as extracting patterns to provide forecasts.
- **Prescriptive Analysis.** It directs decision-making in an efficient way. Optimization techniques are used to find optimal solutions under a series of given constraints.

4.2 Data Fusion

Integration and data fusion can be described as the synergistic use of information from multiple sources to aid in the general understanding of a phenomenon, to measure evidence or to combine decisions [33]. This component allows one or more correlated datasets to be generated, with coherent information obtained from different and varied data sources. These sources belong primarily to the biophysical and economic-productive dimensions; in case of finding data that belong to other dimensions, these will be analyzed to determine their possible incorporation in the data fusion component.

On the other hand, the demand for data fusion processes within a wide range of applications has made it possible to propose different data fusion models [34]. These models seek to provide a well-defined set of guidelines for identifying the system under study, regardless of the nature of the data collected, and thus contribute efficiently to the subsequent selection and implementation of the most suitable data fusion model. One of the most widely used frameworks is the JDL (Joint Directors of Laboratories) model, which identifies three levels of fusion:

- **Level 1:** objective refinement
- **Level 2:** situation assessment
- **Level 3:** threat assessment

Although this model is currently used in many applications, it has often been criticized for its levels of development in a strict order, and also for its lack of adequate representation of a human's expert knowledge within the process. With this in mind, the present proposal is based on the evolution of JDL, which has refined the development of new models such as DFIG (Data Fusion Information Group) [35]. It focuses primarily on the implications of situational awareness, user refinement and mission management. This model consists of six levels:

- Level 0: pre-processing of sources and topic assessment
- Level 1: objective assessment
- Level 2: situation assessment
- Level 3: impact assessment or threat refinement
- Level 4: process refinement
- Level 5: user refinement or cognitive refinement

4.3 Detection Mechanism

This component is closely related to Big Data Analysis component, which data analysis is around traditional methods and Big Data methods, such as data mining and machine learning. In this component different algorithms or techniques of machine learning are applied for both the data preprocessing and the detection process. Specifically, the detection is to determine which new crops could be planted and produced in a geographical area. The aim of this component is to find non-obvious correlations with traditional techniques for modelling data and learning their patterns.

4.4 Decision Makers GUI

This GUI shows the user all the information that has been analyzed from the different datasets. This tool allows the farmer, technician, researcher, or decision maker to consult the best options for planting different crops in their area, as well as other information that involves the management of their crops so that they can be produced properly and obtain acceptable yields.

4.5 Support Components

These components provide additional information about the datasets, characteristics of the dimensions, characterization of the agricultural area, among others. Similarly, information about the calibration of the algorithms is also obtained in this component.

5 Conclusions and Future Work

This paper has presented an approach in automatic detection of favorable conditions for planting and crop production in the upper basin of Cauca River. This research aims to establish a reference around the field of Smart Farming and the sustainability of crops in Colombia. These outcomes will allow farmers, technicians, farmers federations, and decisions makers in general, to optimize their courses of action when handling their sowing and crop production. Furthermore, prescriptive analysis represents a significative method for establish new options of crops which can be appropriately adapted to the conditions of a territory.

Meanwhile in the current state of knowledge, most of the studies reviewed in data fusion perform the integration of data at the multi-sensor level, i.e. only taking into account environmental data. In addition, an analysis of the data is not performed taking into account the variety of the same from different dimensions (economic, political, social, among others). Furthermore, in machine learning field, agricultural production estimation models only are applied to a specific dataset pertaining to a crop and a particular area. Therefore, production estimations are based only on the historical climate data. Similarly, estimations of the crop adaptive capacity are based on future scenarios of climate change. As a consequence, the period of time in these scenarios is too wide to obtain an adequate model validation. Finally, in Big Data Analytics paradigm, none of the reviewed studies addressed the issue of sowing and crop production from a formal data analysis perspective in Big Data and Smart Farming. Most are approximations that try to provide solutions from different paradigms that conceptually could be grouped in Big Data.

The selection of variety, value, and viability characteristics lies in the intrinsic needs of the agricultural domain; i.e. in addition to volume, there is a wide variety of data sources that require a thorough analysis to determine the interrelationships of those variables that generate relevant knowledge for the decision maker.

As future work, we propose the implementation of this research work and evaluate them in a case study. For this purpose, the collection of data from different sources is important taking into account that each one represents a particular dimension. The merging of data of different dimensions will allow the refinement and enrichment of each dataset, improving the accuracy of the model.

Acknowledgements. The authors would like to thank Universidad del Cauca, AgroCloud project of the RICCLISA program for supporting this research, and Colciencias (Colombia) for PhD scholarship granted to MSc. Iván Darío López.

References

1. Lau, C., Jarvis, A., Ramírez, J.: Centro Internacional de Agricultura Tropical (CIAT) - Agricultura Colombiana: Adaptación al Cambio Climático. Centro Internacional de Agricultura Tropical (CIAT), 1 (2013)
2. Solomon, S.: Climate Change 2007 - The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC. Cambridge University Press (2007)
3. DNP: Plan Nacional de Adaptación al Cambio Clímatico PNACC. Ministerio de Ambiente y Desarrollo Sostenible, Resúmen Ejecutivo (2014)
4. Agricultura: Vulnerabilidad y Adaptación: metodología para medir la vulnerabilidad del sector agrícola. Climate and Development Knowledge Network (2011). http://cdkn.org/project/agricultura-vulnerabilidad-adaptacion-cuenca-alta-cauca/
5. Kalra, N.: Advanced computational procedures for the understanding of agricultural processes. Curr. Sci. **113**(2), 208–209 (2017)
6. Sundmaeker, H., Verdouw, C.N., Wolfert, J., Freire, L.P.: Internet of Food and Farm 2020. In: Digitising the Industry, vol. 49, pp. 129–150. River Publishers (2016)
7. Kaur, B., Owusu, R.: Inverse problems and data fusion for crop production applications targeting optimal growth – Fertilization. In: 2015 26th International Workshop on Database and Expert Systems Applications (DEXA), pp. 108–114 (2015)
8. Malviya, S., Mittal, D., Birle, A.: Agriculture multi sensor data fusion and analysis system. Int. J. Adv. Res. Electron. Commun. Eng. (IJARECE) **4**(5), 1334–1337 (2015)
9. Griffin, S., Kunz, E.: Data fusion/integration of high and medium resolution imagery for crop analysis. In: Presented at the American Society for Photogrammetry and Remote Sensing Annual Conference (ASPRS), Baltimore, Maryland, p. 4 (2009)
10. Pohl, C., Kanniah, K.D., Loong, C.K.: Monitoring oil palm plantations in Malaysia. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 2556–2559 (2016)
11. Gandhi, N., Armstrong, L. J., Petkar, O.: Proposed decision support system (DSS) for Indian rice crop yield prediction. In: 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), pp. 13–18 (2016)

12. Moosavi, V., Talebi, A., Mokhtari, M., Shamsi, S., Niazi, Y.: A wavelet-artificial intelligence fusion approach (WAIFA) for blending Landsat and MODIS surface temperature. Remote Sens. Environ. **169**, 243–254 (2015)
13. Rötter, R., Carter, T., Olesen, J., Porter, J.: Crop-climate models need an overhaul. Nature Clim. Change **1**(4), 175–177 (2011)
14. Snehal, M., Dahikar, S., Sandeep, V., Rode, V.: Agricultural crop yield prediction using artificial neural network approach. Int. J. Innovative Res. Electr. Electron. Instrum. Control Eng. **2**(1) (2014)
15. Okori, W., Obua, J.: Machine Learning Classification Technique for Famine Prediction. In: Presented at the Proceedings of the World Congress on Engineering 2011, London, UK, vol. 2 (2011)
16. Sap, M., Awan, A.: Development of an intelligent prediction tool for rice yield based on machine learning techniques. Jurnal Teknologi Maklumat **18**(2), 73–74 (2006)
17. Gandhi, N., Armstrong, L., Petkar, O., Tripathy, A.: Rice crop yield prediction in India using support vector machines. In: 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 1–5 (2016)
18. Kolesnikova, A., Song, C.-H., Lee, W.: Applying UChooBoost algorithm in precision agriculture. In: Proceedings of the International Conference on Advances in Computing, Communication and Control, New York, NY, USA, pp. 30–34 (2009)
19. Balakrishnan, N., Muthukumarasamy, G.: Crop production-ensemble machine learning model for prediction. IJCSSE **5**(7), 148–153 (2016)
20. Kempenaar, C.: Big data analysis for smart farming. Wageningen University and Research, Wageningen (2016)
21. Aggarwal, P.: Food, fibre and forest products. In: Current, pp. 273–313 (2007)
22. Monaco, E., Bonfante, A., Alfieri, S., Basile, A., Menenti, M., De Lorenzi, F.: Climate change, effective water use for irrigation and adaptability of maize: a case study in southern Italy. Biosyst. Eng. **128**, 82–99 (2014)
23. Badr, G.: Toward large-scale crop production forecasts for global food security. IBM J. Res. Dev. **60**(5/6), 5:1–5:11 (2016)
24. MinAgricultura: Agronet Estadísticas (2016). http://www.agronet.gov.co/estadistica/Paginas/default.aspx
25. CVC: Segunda campaña de muestreo con propositos de calibracion del modelo de calidad del agua del rio cauca (Second sampling campaign with model calibration purposes of water quality in the Cauca river). Corporacion Autonoma Regional del Valle del Cauca Caracterizacion y Modelacion Matematica del Río Cauca -PMC- Fase II. Convenio Interadministrativo 0168 de Noviembre 27 de 2002, vol. XV (2005)
26. DANE: ANDA - Archivo Nacional de Datos (2006). https://formularios.dane.gov.co/Anda_4_1/index.php/home
27. MinTIC: Datos Abiertos Colombia. Plataforma de datos abiertos del gobierno Colombiano (2017). https://www.datos.gov.co/
28. DNP: Datos y Estadisticas DNP (2015). https://www.dnp.gov.co/programas/inversiones-y-finanzas-publicas/Datos-y-Estadisticas/Paginas/Datos-y-Estadisticas.aspx
29. FAO, FAOSTAT (2017). http://www.fao.org/faostat/en/#data/RP
30. SAC: Estadísticas - SAC - Sociedad de Agricultores de Colombia (2012). http://www.sac.org.co/es/estudios-economicos/estadisticas.html
31. Meteoblue: Meteoblue - Weather close to you (2006). https://www.meteoblue.com/es/tiempo/pronostico/semana/canalete_colombia_3779399
32. Hu, H., Wen, Y., Chua, T., Li, X.: Toward scalable systems for Big Data Analytics: a technology tutorial. IEEE Access **2**, 652–687 (2014)

33. Haghighat, M., Abdel-Mottaleb, M., Alhalabi, W.: Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition. IEEE Trans. Inf. Forensics Secur. **11**(9), 1984–1996 (2016)
34. Sidek, O., Quadri, S.: A review of data fusion models and systems. Int. J. Image Data Fusion **3**(1), 3–21 (2012)
35. Blasch, E., Bossé, É., Lambert, D.: High-Level Information Fusion Management and System Design, 1st edn. Artech House, Inc., Norwood (2012)

An Online Learning Method for Embedded Decision Support in Agriculture Irrigation

Harold Murcia-Moreno(✉), Brigete González-Quintero, and Jeison López-Gaona

Department of Electronics, Faculty of Engineering, Universidad de Ibagué,
Carrera 22 calle 67, Ibagué, Colombia
harold.murcia@unibague.edu.co,
{2420132005,2420121002}@estudiantesunibague.edu.co

Abstract. In view of the advantages of Wireless Sensor Networks (WSNs), acquisition devices, environmental user-interfaces and low-cost monitoring systems in the agricultural and farming domain, an innovative use of automatic learning decision support is proposed to manage the irrigation process. The aim of this work is to develop a low computational cost technique for a smart irrigation support system, which can be implemented into a simple microcontroller. The classical methods in automatic irrigation use basic on-off controller or complex models with a large number of variables and focus to replace the farmer in the control scheme. Conversely, this proposal uses the farmer experience as the center of the closed loop interpreting its irrigation rules and adapting to the crop changes depending on growth cycle, weather and soil sensors' signals without involving any model. The three weeks dataset for the testing was constructed from a one-week experimental setup with soil-water potential, soil temperature and sunlight sensors by using approximation functions. Moreover, an online algorithm (AdaDelta) based on gradient descent was tested in an adaptive binary classification with a single layer neuron via MATLAB simulation. Preliminary results of this application have shown its potential with an accuracy of 97% and 6.5% mean square error over the reference method, which poses new possibilities to work in this approach and generate precision agriculture applications for low cost and common irrigation plants by using the new age technologies.

Keywords: Precision agriculture · Water management · Automated irrigation · Adaptive learning

1 Introduction

In recent years, the global food demand has increased. Some research estimates that food production must increase about 70% for 2050, given the population growth and the constant limitations in crop area [1], which implies a necessary increase of agricultural production per square meter. In other words, the methods of cultivation should improve every day, even though sowing surfaces and water resources are growing increasingly scarce. Moreover, the climate changes have a negative impact on agriculture practices. In some countries, the climate trends from 1980 to 2008 had a large impact on average

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_18

yields, and other agricultural factors [2]. For example, the increase in temperatures recorded in the last 25 years, has led to a decrease in rice production ranging from 10% to 20% in various parts of the world, and while daytime minimum temperatures rise, and nights are warmer, production will continue to fall [3]. In this way, modern day farming requires new techniques in agriculture to improve planting planning, efficient water management and mitigation of the climate changes on crops' yields. Irrigation is an essential component of water management and according to recent reports, agriculture irrigation accounts for 50–60% of freshwater usage from sources in the natural environment and up to more than 90% in some developing countries [4]. In Colombia, recent studies have shown that agriculture uses the majority of water demand: 46.6% of the hydric demand with 16,760.3 millions of cubic meters and this percentage will continue to dominate in water consumption due to population growth and increased food demand [5]. From this principle, there is a notable need to develop effective actions, science and technology-based strategies for sustainable water use, including technical, agronomic, administrative and public policy improvements [6].

With the current drought conditions in our country due to climate changes, the need for irrigation systems in some regions is evident. When soils do not have the amount of water needed to supply a crop and rain is not available, it is vital to supply water using irrigation processes to ensure crop health. Weather is the one of the first key factors to estimate the water requirements of the crops [7]. Weather measurement stations are a useful tool to obtain reference evapotranspiration, (ETo) which is an important factor in determining water consumption and water requirement models [8, 9]. Based on meteorological variables, these methods can determine the current and future water conditions to make decisions regarding the irrigation of crops [10]. However, not all farmers have access to a network of weather stations or advanced irrigation systems. In other cases, the number of stations is lower than what is necessary given their cost and because of insufficient information, the predictions may fail; besides, the executions of the systems are accurate but complex with the need for specialized personnel to manage the system [11]. On the other hand, although many innovative techniques have been employed towards automated irrigation, in real implementation, the vast majority of techniques used are based on the basic on-off controller given that most irrigators on farms are bi-state, consisting of switching the controller output between maximum and minimum output according to the sign of the error [12]. Some works present the combining of weather data and crop variables as a potential method to irrigate efficiently. However, the presence of an agronomist expert (a human) is an important agent in the closed loop decisions, due to the complexity of factors such as: Soil types, plants varieties, water requirements according to growth cycle and presence of biologic agents like plagues.

Currently, the constant search for the decision irrigation systems, is the development of techniques and technologies to improve the agricultural water management and to provide the precise tools for irrigation in crops, with an easily understanding for the farmers by using techniques from the automatic control and Artificial Intelligence that may provide solutions in an efficient way. The use of online learning techniques about automatic making decisions could be a potential solution to improve agriculture irrigation in any variety of crops and reducing water systems' extreme costs without displacing farmer labor. The study of online learning algorithms is an important domain in machine learning, and it has both interesting theoretical properties and practical applications. The goal of online learning is to make a sequence of accurate predictions given knowledge of the correct answer to previous prediction tasks and possibly additional available information [13]. Applied to our study problem context the online learning algorithms can tune the irrigation decision constants based on environmental data and the feedback information given by the farmer. Meanwhile, the adaptive learning features would ensure the correct parameter tuning whatever the crop may be in any of its phenological stages.

The motivation of this article is to develop a simple algorithm with a recursive update parameter, which can be implemented on a low-cost microcontroller, based on the agronomist experience and a few of environmental variables to support the decisions-making about irrigation. The paper continues as follows, Sect. 2 will introduce to their related works, then the problem statement and methodology are presented, to be followed by the results, their discussion and the conclusions.

2 Literature Review and Related Work

Over the last few years, different intelligent irrigation techniques have emerged, along with advances in monitoring networks, systems composed of a set of sensors, responsible for measuring variables such as soil moisture and temperature, and communicate its levels real-time to a central computer where it is analyzed by farmers and agronomists experts [14]. Among recent work are applications such as: Efficient energy control of irrigation in agriculture through the use of Wireless Sensor Networks (WSN), which are capable of capturing the irrigation processes quickly and transmit the data in real time and sensors that can be used by experts or be processed by automated systems for decision-making [15]. Generally, these irrigation models calculate the timing and amounts of water to be scattered, that is, it selects historical data, evaluates climatic changes and calculates the amount of water needed for irrigation [16, 17]. They are also capable of continuous measurement of the state of the water from the plant as well as the soil for the programming of the irrigation of plum. The measurement is made with sensors and supported by linear transformers of variable displacement [18].

Most automatic irrigation systems are based on evapotranspiration (ETo) calculations, historical climate data (ED) and their relationship to the type of crop involved in [19]. The efficiency of several standard “intelligent irrigation” controllers, with an additional soil moisture sensor (SMS) system is analyzed by keeping in mind two important characteristics: the amount of water applied and the quality of the earth/ground. These characteristics were analyzed in residential environments. These controllers examine weekly the use of water and perform their readings by irrigation meters. On the other hand, great research efforts have been made to improve the efficiency of water consumption, trying to establish the irrigation bounds in situations of limited water supply. These methods of theoretical analysis determine the lower boundary of dynamic irrigation managing to innovate against agricultural irrigation. For its development, an integration of the following models is carried out: dynamic simulation of soil moisture and nitrogen, crop growth and the optimization of the irrigation program. The first two models reflect the influence of fertirrigation, the meteorological factors and the physiological characteristics of the crops in the lower limit of dynamic irrigation. The latter model reveals the effect of the social economy and the level of management of water conservation in the lower limit of dynamic irrigation. However, these methods require a precise estimation of ETo indexes and the validation of complex models involving a large number of variables [20]. There are also systems to support decisions for the management of irrigation in agriculture. For instance, SIDSS is a system based on soil and climate measurements, being counted by several automatic nodes extended in the field, estimating the irrigation needs of a plantation, being carried out on a weekly basis. A control is made against local disturbances and errors that are presented in front of the estimates, which paves the way for two techniques of automatic learning tested with decisions made by a human expert [21]. Besides, there are also works with low cost microcontrollers but simple decision making based on a system to monitor the temperature and state of the water in a crop. Its function is to store and retrieve data using analog soil moisture sensors, and digital temperature sensors. A maize crop is constructed and tested to evaluate performance and suitability in local conditions [22, 23].

In Colombia, the development and implementation of irrigation systems has an equitable progress in hardware and software, which is not focused on irrigation decision systems, but on designing of network of wireless sensors (WSN) that has been realized with different types of nodes. Therefore, some communication devices and virtual tools that work together to form an environment offer to the user a greater control of crop irrigation based on evapotranspiration. Finally, the decision for the actuators states is done from the user through a web application, this allows to turn on and off the different elements that are in the system authorizing the irrigation of the crop by using sprinklers [24]. The work carried out recently concludes that irrigation systems that use biophysical properties are systems for the automatic acquisition of knowledge that focus on expert systems in charge of irrigation and fertilization, which seek to optimize decision making. There must be an expert system for the configuration of irrigation programming based on operations research and qualified tools in the construction of the irrigation expert system. In recent years, no work has been reported on a method that adapts to crop phenology (growth station), that is, intelligent methods that adapt to the phenological changes and its requirements. Besides, it should be noted that all the methods used so far require precise and efficient but complex models involving many variables, which make the possibility to work with the low-cost equipment or easy access elements for most farmers difficult. Numerous automatic control reports have been focused on efficiency by using linear actuators with variable flow devices. However, although this configurations present better performances, few farms have flow control elements for irrigation on crops. This does not mean that the efficiency of the basic on-off controllers cannot be improved by the use of intelligent irrigation systems.

3 Problem Statement

We consider a scenario of a farm or greenhouse as a place where environmental sensors are located in a crop, an active element (actuator) defines the irrigation process and both sensors' data (inputs) and the state of the active element (output) are registered. The core of this closed-loop scheme is a farmer or an agronomist expert, who define when and for how long to irrigate the crops through activation of the active element (pump or gate) with a constant water flow. A two-state actuator was defined given its simplicity, low cost and popularity among predominantly real implementations [20]. Figure 1 illustrates the raised scenario where input/output data is managed by agro-meteorological stations which have a strong background in this kind of monitoring applications. The collected information could be saved on a computer, or sent to an online server via internet and a user interface could be the bridge between the agriculture and the crop.

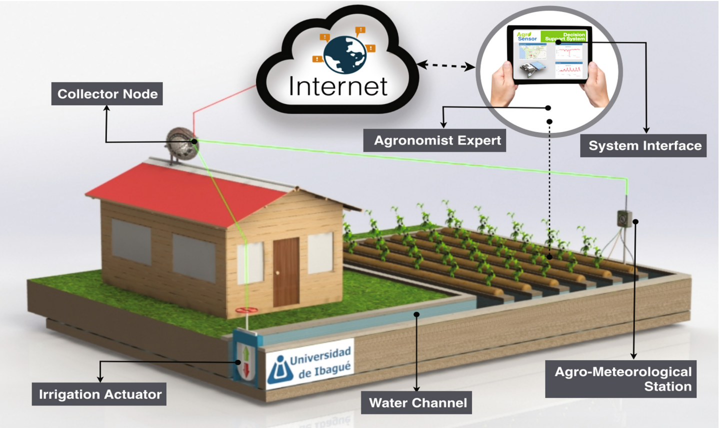


Fig. 1. A Cloud Sensor Network (CSN) diagram deployed for agricultural applications.

Moreover, usually on farms, people observe environmental features as the color and texture of soil and leaves, sun light, time of day, weather conditions, crop information, ranges of moisture type of soil or phenology [25] and based on those experiences, determine some irrigation rules. These rules are static for a period of time, but given that the sensitivity to water requirements change with cycle of the crop [26], the irrigation rules must change from one cycle to another, i.e. the irrigations rules are dynamic conditions throughout the crop, e.g., in corn crops the hydric stress has a bigger negative impact on production during the flowering stage, than in the other growth cycles [27, 28]. Currently, the automatic irrigation devices do not have flexible decision elements, whereas the development of online techniques learn from the human, that adapt recursively with the conditional changes and with a low computational cost are a potential solution on automatic systems for irrigation in crops and greenhouses taking into account the proliferation of agricultural projects powered by low-cost microcontrollers. In further expert observations, sensors can quantify the variables information variables and to ease the decision making. In agriculture, temperature sensors and light sensors are frequently used on telemetry applications. However, Soil Matric Potential SMP sensors are one of the most relevant variables in water applications and are widely used by the agronomists to determine the availability of soil water and the ability of plants to extract it [29]. Figure 2 shows the flow diagram for irrigation process starting from classical decision-making (continuous line) and add the proposed advice system (dashed line). In this diagram, the smart support decision system is always working in series or parallel topology, in regards to the normal loop, learning from expert actions through the sensors data. Once the learning is complete, the system must generate a control state advice for the actuator or assume the control if the user approves the automatic mode. When the user change the irrigation rules, the system must to learn again to continue with the support. Therefore, we propose a recursive binary classification tasks in two classes as "1" (positive class) and "0" (negative class) for simplicity. The inputs **X** are defined as: soil temperature **ST**, visible sunlight indicator **SIV** and Soil Matric Potential **SMP**. Finally, an activation function $\phi_{(z)}$ generates advice state for the actuator based on inputs data **X**.

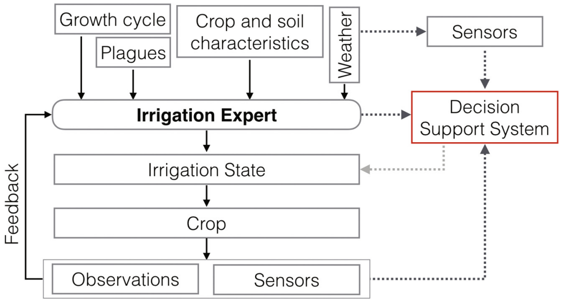


Fig. 2. Scheme of the proposed system. Continuous line represents the typical irrigation scheme and dashed line represents the proposed learning system.

In this schema, the expert determines the amount of water from technical data (weather reports, technical schedules), soil and characteristics (plant variety, size, age, color, texture, cycle, etc.) and define the actuator state (On-Off). Based on this idea, it is possible to establish irrigation rules to define the actuator state, which depends on multiples factors. In some cases, these factors are measurable (inputs), however in other cases it could be difficult e.g. plagues or specific characteristics in color or shape of plants. The directly related-variable with the irrigation is the water potential which normally reaches its maximum value at about −10 kPa when the soil is totally wet and presents values from −23 kPa to −200 kPa according to the soil type, how wet or dry the soil should be depending on soil type, the plants cycle and cultural practices for managing the field [30]. Taking into account the above, for our hypothetical case study, we are considering a crop with three cycles with different sensibility to irrigation and humidity reference. Moreover, we assume that for the second cycle the farmer irrigates the crop only during mornings or nights, and during the last cycle the farmer irrigates the crop only when the sun is out as a strategy to compensate for the temperature, so sunlight is a determinant factor of decision (see Table 1).

Table 1. Irrigation reference values based on the irrigation rules for a case study with hypothetical conditions. SIV numbers represent an indicator and SMP represents the soil suction in Kpa.

Crop cycle	Cycle time	Max. SIV	Min. SIV	SIV reference	Max. SMP	Min. SMP	SMP reference
1	8 weeks	None	None	None	−18	−22	−20
2	8 weeks	10	0	None	−12	−16	−14
3	8 weeks	1000	300	None	−14	−18	−16

4 Methodology

To solve the problem defined in Sect. 3, we proposed to decompose the methodology in three stages (see Fig. 3), and then to solve them successively:

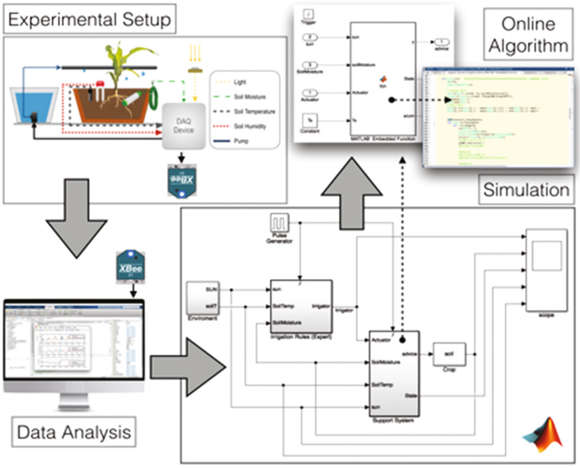


Fig. 3. General diagram of the methodology.

1. Experimental Setup: The first stage consists of the design and implementation of an experimental setup with real conditions to construct the dataset for soil temperature, soil moisture, and sunlight with the sensors pt100, watermark 200ss (Installed at a depth of 20 cm) and grove si1145 respectively. A maize plant was planted in a sandy loam soil into a 20 × 20 × 40 cm pot. The actuator was emulated with a small pump controlled from an Arduino Mega which has programed the irrigation rules and send the data to a computer each minute via ZigBee communication.
2. Approximation Functions: One week of data was saved with 10080 samples per signal and sample rate of one minute; the measurement noise was removed with a low-pass Butterworth and first order filter at 10 Hz in a digital implementation. Once the database was created, simple mathematical models were obtained from acquired signals to create the simulation scenarios.

3. Simulation: Finally, the algorithm was tested according to the irrigation rules of Table 1 using an embedded function in MATLAB Simulink with a sample time of sixty minutes. The environmental signals were created from mathematical approximation functions and random factors, the farmer was simulated with a set of conditional functions depending of the inputs (Soil Moisture, sunlight and soil temperature) and the support system is connected in series between the expert and the actuator.

5 An Algorithm to Decision Support Based on Gradient Descent

In this section, we introduce our adaptation learning algorithm for irrigation activation based on farmer experience. Its methodology is based on extracting the knowledge of the experts (farmers) and being able to process it by a system. An event based diagram of the system is depicted in Fig. 4, as can be observed, there are three states: Initial state for acquiring and normalizing of the data which jump to learning state once the window data are complete, learning state where the coefficients are adapting and auto state, when the learning is complete and the user decided to switch across the user selector μ. The selection of the operational mode is designated by ß that change from false to true in convergence cases and zero when the average error between the support advice and the desired value is higher than a comparison value γ e.g.: changes of crop cycles.

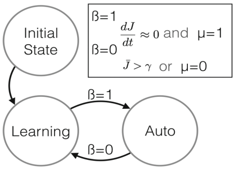

Initial State
ß=1 $\frac{dJ}{dt} \approx 0$ and μ=1
ß=0 $\bar{J} > \gamma$ or μ=0
ß=1
Learning
Auto
ß=0

Fig. 4. General diagram state of the proposed system.

Given the computer resources of low cost microcontrollers, a single layer neural network is an appropriate configuration that can be trained recursively with the sensors data as inputs and the farmer actions as the target in learning state. As seen in Fig. 5, the artificial neurons have inputs (X), connection weights (W), a prediction according to the activation function which is used in training to calculate the weight coefficients and finally a quantizer that determines the output state according to the problem context. The main difference between the presented neural network and the simplest classical perceptron is that the weights are updated based on a nonlinear activation function (Sigmoid for this case) rather than a unit step function like in the simplest perceptron configuration, which have a hard threshold with an output either a 0 or a 1. Sigmoid has the advantage of being differentiable, with derivatives that are easy to calculate and convert into output values between 0 and 1 (soft threshold), which is helpful for calculating the weight updates in certain training algorithms [14].

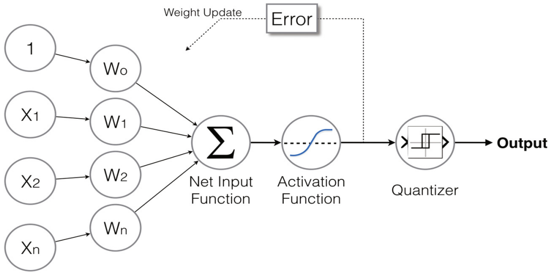


Fig. 5. General diagram state of the proposed system.

Gradient descent is an iterative algorithm and the most common method to optimize neural networks from an objective function J(w) which can be implemented in the batch mode, online mode or using mini-batches with adaptive variants. Besides, its robustness makes it highly applicable in online optimization applications. If the analyzed function is differentiable with respect to its parameters, the gradient descent is an efficient optimization option respect to other methods, since the computation of first-order partial derivatives respect to all the parameters is of the same computational complexity as just evaluating the function [31]. During the learning process, the goal is to minimize the objective function which is parameterized by a model's parameters $w \in R^d$ by updating the parameters in the opposite direction of the gradient of the objective function ∇_w J(w) with respect to the parameters. In typical algorithms, a constant factor called learning rate η determinates the size of the adjustments made to the weights at each iteration of the algorithm, the learning rate value and the initial conditions of the weights have a significant impact on how the gradient descent converges to the optimal point. Bellow the respectively pseudocode for the stochastic gradient descent:

Algorithm 1 The gradient descent algorithm for training linear regression models

Requirement: Set of training instances with length T
Requirement: A fixed learning rate η
Requirement: A function, *gt*, that determines the direction in which to adjust a given weight, $\boldsymbol{w_t}$
Requirement: A convergence criterion which indicates that the algorithm has completed

```
Begin w with a random starting point
repeat
  for t=1:T do %% Loop over # of updates
     Compute gradient: gt
     Δx= gt. X
     w = w - η Δx
  end for
until convergence occurs
```

There are some variants of gradient descent, which differ in how much data we use to compute the gradient of the objective function and the time it takes to perform an update. Some of these adaptive algorithms are AdaM, AdaGrad and AdaDelta, which can adapt the learning rate to the parameters, performing larger updates for infrequent and smaller updates for frequent parameters [32]. Dean et al. [33] have found that AdaGrad greatly improved the robustness of SGD. However, AdaDelta is an extension of AdaGrad that seeks to reduce its aggressive, monotonically decreasing learning rate, instead of accumulating all past squared gradients, Adadelta restricts the window of accumulated past gradients. We use an AdaDelta solver for the optimization process because it is a type of gradient descent solver that is very robust and automatically updates not only the parameters of the function but also adapts its own learning rate [34, 35]. We have found this solver to perform better than a simple stochastic gradient descent which requires tuning of the learning rate to converge. Compared with the classical gradient descent, the sum of gradients is recursively defined as a decaying average of all past squared gradients in AdaDelta. Assume at time t this running average is $E\left[g^2\right]t$ then we compute:

$$E\left[g^2\right]t = \rho E\left[g^2\right]t - 1 + (1-\rho)gt^2 \tag{1}$$

where ρ is a decay constant. Since we require the square root of this quantity in the parameter updates, this effectively becomes the Root Means Square RMS of previous squared gradients up to time t:

$$RMS\left[g\right]t = E\left[g^2\right]t + \varepsilon \tag{2}$$

where a constant ε is added to better condition the denominator, the resulting parameter update is then:

$$\Delta xt = \frac{-\eta}{RMS\left[g\right]t} gt \tag{3}$$

Bellow the respectively pseudocode for the AdaDelta method:

```
Algorithm 2 Computing AdaDelta update at time t
Requirement: Decay rate ρ, Constant ε
Requirement: Initial parameter x1

Initialize accumulation variables $E[g]^2 0 = 0, E[\Delta x^2]0 = 0$
repeat
  for t=1:T do %% Loop over # of updates
     Compute Gradient: gt
     Accumulate Gradient: $E[g^2]t = \rho E[g^2]t\text{-}1 + (1-\rho)gt^3$
     Compute Update: Δxt = $-\frac{RMS[\Delta x]t-1}{RMS[g]t} gt$
     Accumulate Updates: $E[\Delta x^2]t = \rho E[\Delta x^2]t\text{-}1 + (1-\rho)\Delta xt^2$
     Apply Update: xt+1 = xt + Δxt
end for
until convergence occurs
```

6 Results

The first results of the present work are the input's modelling and the simple mathematical models using for simulation. Then the simulation results using the proposed algorithm are presented.

6.1 Approximation Functions

The acquired dataset was not enough to perform a long-time simulation, for this reason it becomes necessary to generate the corresponding signals by simulation from mathematical approximations in function of the time. The purpose of the approximation signals is not modelling the weather or crop environmental conditions, but to generate an average scenario for the simulations. Figure 6 shows the adjustment of the approximated functions on the real data obtained from experimental setup for the sunlight SIV in visible channel (400 nm–750 nm), soil temperature soilT in °c and soil matric potential SMP in kilo-Pascal Kpa. The sunlight and temperature were modelling based on sinusoidal functions multiplied with a synchronized pulse signal R_t, which has one-day period and changes the amplitude each day with an aleatory gain. The gain of R_t is saturated according to the minimum and maximum obtained for each experimental signal, then there is a positive pulse signal with a limited and aleatory gain which changes its value for each day, e.g. R_t is the factor that influences what would be a sunny day or a gray day in an aleatory manner. On the other hand, the soil matric potential was modeled as a two states function, for simulation ease, a first order function with a fast constant of time when irrigation is "on" and a first order function with a slow constant of time when irrigator is "off". The oscillation in SMP signal is a normal effect of temperature changes and is minor when the sensor is installed at greater depth. The obtained functions are given as:

$$\begin{aligned} SIV &= R_t\left(A\ sin(2\ pi\,f\ t+\alpha)^6\right) \\ soilT &= R_t(B\ sin(2\ pi\,f\ t+\beta)+b) \\ SMP_{on} &= K_1 t + a_1 \\ SMP_{off} &= -K_2 t + a_2 \end{aligned} \tag{4}$$

where

$A = 30, f = 3.4722e - 4, B = 5, b = 25\ °c, K_1 = -0.0008, a = -21.0056, K_2 = -0.034, a_2 = -1.6e2.$

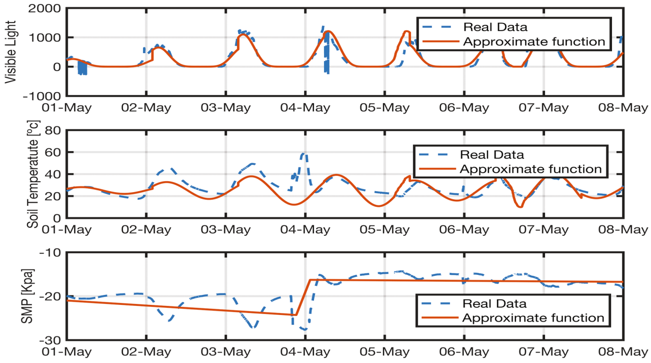


Fig. 6. Modelling results for the approximation functions.

6.2 Simulation Results

Given the fact that we are presenting a classification problem for an irrigation stated, we define a simulation scenario based on Table 1 with three crop cycles and considering as inputs the sunlight and soil matric potential signals. In this scenario, we aim to learn recursively the farmer irrigation rules for each cycle and to predict the irrigation actuator state in an automatic mode based on sensor inputs and the human experience. The behavior of the controlled variable (**SMP**) and restrictive variable (sunlight) during three

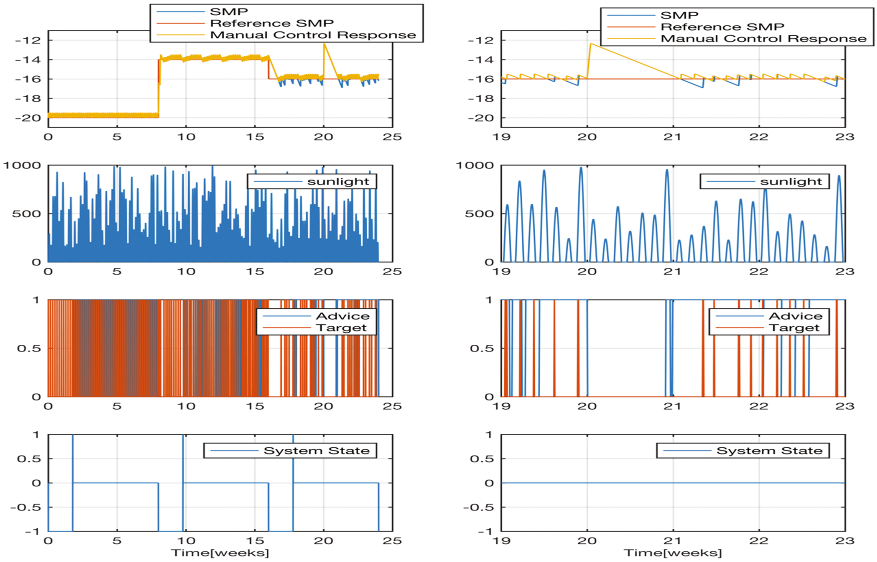


Fig. 7. Prediction of the actuator irrigation using sunlight, soil matric potential and agronomist actions information (cycle 1: week 1–8, cycle 2: week 9–16 and cycle 3: week 17–24). Left column: total simulation time, right column: week 19–23 in presence of a disturbance like a rainy day.

cycles of a crop are depicted by the two diagrams presented in Figs. 7 and 8, in order to test the dynamic performance, the Fig. 7 presents the simulation results during 24 weeks (8 week per cycle), the input signals soil moisture (top), sunlight (middle), the ideal state "Target" determined by the farmer and the advice prediction (middle) and the state indicator (bottom) of the algorithm according to Fig. 4 where "−1" represents the initial state, "+1" represents the learning state and "0" represents the automatic state.

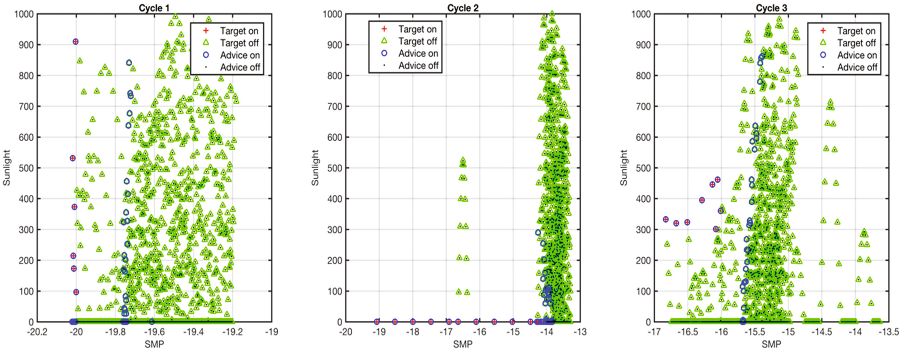


Fig. 8. Simulation results: discriminant analysis for irrigation actuator according to sensors data (sunlight and SMP). Left: cycle 1, middle: cycle 2 and right: cycle 3.

The results show the reference value for each cycle and both controlled **SMP** responses by using the learning control and the manual control, which represents the farmer and follow rigorously the irrigation rules (see Table 1). A disturbance rejection was tested by adding a pulse disturbance on the input of the soil model (simulation of a rainy day) at week 20, the **SMP** increases reached −12 kPa, then both controllers machine and "human" response turning off the irrigator until the soil water potential fell to the reference value. The middle figures show the sunlight signal and its restriction effects in terms of control law of advice and target, for cycle 2 the crop is irrigated at night and during the cycle 3 is irrigated in sunny hours of the day. Mean Squared Error MSE was calculated for automatic and manual controllers respect to the reference signal on the simulation conditions for a period of 168 days, the total error for the proposed system was 9.155 kPa^2/day (including training, learning and working time) and the compared ideal method was 8.599 kPa^2/day.

During the first few days, the system is recording data (system state = −1) to determine the normalization values (mean and variance), after first stage the algorithm jumps to the learning state (system state = +1) where AdaDelta is executed online until it converges to a solution, then, while the error doesn't overcome a fixed umbral nor does the user decide to change the cycle. The automatic mode (system state = 0) is working with the obtained coefficients. In the eighth week of cultivation, the irrigation rules change and the system responds changing the state from automatic to an initial state again by the user order on the system, recording new data to jump to the learning state by adapting its coefficients to the new conditions. Once learn again, return to the automatic state until the end of the second cycle and continues with the same sequence for the others crop cycles until the las cycle.

Figure 8 shows a discrimination cloud for all the simulation data between the target and the prediction, target is the ideal state for the actuator according to the **SMP** reference and the restrictions of sunlight, the prediction is the advice from the system for the actuator state according to the input signals and the trained coefficients. The true positives (characterized with "+") represent the activation of the actuator "on" when the actuator must be in activation, the true negatives (characterized with a triangle) represent the deactivation of the actuator "off" when the actuator must be deactivated, the prediction for the true positives (characterized with "o") represent an activation advice for the actuator and the prediction for the true negatives (characterized with ".") represent a deactivation advice for the actuator. The obtained cloud presents an obvious relation with the Table 1 as it was expected given the close loop effect; for the cycle 1 there is an evident activation **SMP** umbral in −20 kPa and the "on-off" data are not restricted by the light conditions, since it points is distributed in all the axis of ordinates. On the other hand, for the cycle 2 and 3, the region "on" is defined in both axes with limits in −14 kPa for cycle 2 with restriction of 10 for **SIV** and −16 kPa for cycle 3 with restriction of 300 for **SIV**. From this figure, it can be noted that although the learning data and the prediction data are not the same. The signal data is not far from the control reference for **SMP** and the output control signal obeys sunlight restriction, taking into account that it is an adaptive control and learning system.

Table 2 summarizes the results of Fig. 8 in four indicators, which show 97% accuracy in right decisions in a simulation of 4033 samples for 168 days. The directly participation from the user was of 22.37% of the total time and a 77.63% for the support decision system, compared to a traditional manual control which involucres the user at 100% for the making decisions about irrigation or the literature review where the simple systems are tuned by experts with a minor participation of the final user.

Table 2. Sensitivity and classification results from simulation test.

Cycle	True positive	True negative	False positive	False negative	Samples
1	13	1293	39	0	1345
2	21	1284	39	0	1344
3	8	1297	39	0	1344
Total	42	3874	117	0	4033

7 Conclusions

An automatic irrigation support system based on online learning for two inputs has been proposed. The system shows advantages such as simplicity, for low computational resources and adaptive conditions during changes of irrigation rules due to the growth cycles of crops. The dataset was emulated and extended from an experimental setup for sunlight sensor in channel of visible spectrum, soil temperature and water potential using a watermark soil matric potential sensor. Currently, a MATLAB simulation scenario test is applied to validate its functionality with change of target conditions for three crop cycles according to the assumed irrigation rules. The learning algorithm tested AdaDelta was simulated with an embedded function to be as accurate as possible, of implementation conditions for a microcontroller. The simulation results were compared with an ideal human controller over the reference method value and presented good static performance with 6.5% of MSE over the compared method, 97% accuracy in decision making, with 22.37% of direct participation from the farmer and good dynamic adaptation with short learning periods less than a minute; confirming that the proposed implementation is a possible approach for low-cost irrigation controllers and simple actuator irrigation systems with qualities to develop an experimental version which completely validate the method. For near future research, we aim to extend and evaluate the system in different scenarios with various irrigation rules and input variables.

Acknowledgements. This research is being developed with support of the Universidad de Ibagué project no. 15-367-INT. The results presented in this paper have been obtained with the assistance of students from the Research Hotbed on Instrumentation and Control (SI2C), Research Group D+TEC, Universidad de Ibagué, Ibagué-Colombia.

References

1. Food and Agriculture Organisation: How to feed the world in 2050. In: Proceedings of a Technical Meeting of Experts, Rome Italy, 24–26 June 2009 (2009)
2. Lobell, D.B., Schlenker, W., Costa-Roberts, J.: Climate trends and global crop production since 1980. Science **333**(6042), 616–620 (2011)
3. Peng, S., et al.: Rice yields decline with higher night temperature from global warming. Proc. Natl. Acad. Sci. U. S. A. **101**(27), 9971–9975 (2004)
4. UNESCO and T. United: Water in a Changing World, vol. 11, no. 4 (2009)
5. IDEAM: Estudio nacional del agua, Minist. Medio Ambient, p. 253 (2014)
6. ONU: Sustainable Development Goals. United Nations (2017). http://www.un.org/sustainabledevelopment/sustainable-development-goals/
7. Cai, X.L., Sharma, B.R.: Integrating remote sensing, census and weather data for an assessment of rice yield, water consumption and water productivity in the Indo-Gangetic river basin. Agric. Water Manag. **97**(2), 309–316 (2010)
8. Xu, J., Liu, X., Yang, S., Qi, Z., Wang, Y.: Modeling rice evapotranspiration under water-saving irrigation by calibrating canopy resistance model parameters in the Penman-Monteith equation. Agric. Water Manag. **182**, 55–66 (2017)
9. Birendra, K.: Irrigation Scheduling: a Soft Adaptor to Weather Uncertainties and Irrigation Efficiency Improvement Initiatives (2016)
10. Aladenola, O., Madramootoo, C.: XVII th World Congress of the International Commission of Agricultural and Biosystems Engineering (CIGR) Development of a Model for Estimation Current and Future Irrigation Water Demand in Canada (2010)
11. Braneon, C.V.: Agricultural water demand assessment in the Southeast U.S. under climate change (2014)
12. Ogata, K.: Modern Control Engineering. Prentice Hall, Upper Saddle River (2010)
13. Shalev-Shwartz, S.: Online learning and online convex optimization. Found. Trends® Mach. Learn. **4**(2), 107–194 (2012)

14. Kelleher, J.D., Mac Namee, B., D'Arcy, A.: Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies (2015)
15. Huang, Y., Li, C.: Real-time monitoring system for paddy environmental information based on DC powerline communication technology. Comput. Electron. Agric. **134**, 51–62 (2017)
16. Gutierrez, J., Villa-Medina, J.F., Nieto-Garibay, A., Porta-Gandara, M.A.: Automated irrigation system using a wireless sensor network and GPRS module. IEEE Trans. Instrum. Meas. **63**(1), 166–176 (2014)
17. Vellidis, G., Tucker, M., Perry, C., Kvien, C., Bednarz, C.: A real-time wireless smart sensor array for scheduling irrigation. Comput. Electron. Agric. **61**(1), 44–50 (2008)
18. Nikolidakis, S.A., Kandris, D., Vergados, D.D., Douligeris, C.: Energy efficient automated control of irrigation in agriculture by using wireless sensor networks. Comput. Electron. Agric. **113**, 154–163 (2015)
19. Intrigliolo, D.S., Castel, J.R.: Continuous measurement of plant and soil water status for irrigation scheduling in plum. Irrig. Sci. **23**(2), 93–102 (2004)
20. Nautiyal, M., Grabow, G.L., Huffman, R.L., Miller, G.L., Bowman, D.: Residential irrigation water use in the central piedmont of North Carolina. II: evaluation of smart irrigation technologies. J. Irrig. Drain. Eng. **141**(4), 4014062 (2015)
21. Romero, R., Muriel, J.L., García, I., Muñoz de la Peña, D.: Research on automatic irrigation control: state of the art and recent results. Agric. Water Manag. **114**, 59–66 (2012)
22. Navarro-Hellín, H., Martínez-del-Rincon, J., Domingo-Miguel, R., Soto-Valles, F., Torres-Sánchez, R.: A decision support system for managing irrigation in agriculture. Comput. Electron. Agric. **124**, 121–131 (2016)
23. Fisher, D.K., Kebede, H.: A low-cost microcontroller-based system to monitor crop temperature and water status. Comput. Electron. Agric. **74**, 168–173 (2010)
24. McCarthy, A.C., Hancock, N.H., Raine, S.R.: Advanced process control of irrigation: the current state and an analysis to aid future development. Irrig. Sci. **31**(3), 183–192 (2013)
25. Anaya-Isaza, A.J., Peluffo-Ordoñez, D.H., Ivan-Rios, J., Castro-Silva, J.A., Ruiz, D.A.C., Llanos, L.H.E.: Sistema de Riego Basado En La Internet De Las Cosas (IoT) Internet of Things for Irrigation System (IoT)
26. Shock, C.C., Wang, F.X., Flock, R., Feibert, E., Shock, C.A., Pereira, A.: Irrigation Monitoring Using Soil Water Tension
27. Rhoads, F.M., Yonts, C.D.: Irrigation Scheduling for Corn - Why and How, National Corn Handbook, vol. NCH-20, Electronic version (2000)
28. Maíz|Irritec – Sitemas de riego. http://www.irritec.com/es/soluciones/agricultura/maiz/. Accessed 08 Apr 2017
29. Villaú, J.M.: Manual Técnico del Manejo del Riego en el Cultivo de Maíz, Spain
30. Whalley, W.R., Ober, E.S., Jenkins, M.: Measurement of the matric potential of soil water in the rhizosphere. J. Exp. Bot. **64**(13), 3951–3963 (2013)
31. Toulis, P., Tran, D., Airoldi, E.: Stability and optimality in stochastic gradient descent. Harvard University, 12 May 2015
32. Irrometer: Basics. http://www.irrometer.com/basics.html. Accessed 15 May 2017
33. Ruder, S.: An overview of gradient descent optimization algorithms, September 2016
34. Wang, L., Yang, Y., Min, R., Chakradhar, S.: Accelerating deep neural network training with inconsistent stochastic gradient descent. Neural Netw. **93**, 219–229 (2017)
35. Zeiler, M.D.: ADADELTA: An Adaptive Learning Rate Method. Cornell University, December 2012

Interinstitutional Relational Capital of Support for Climate Change and Food Security, an Analysis from the Social Networks in Cauca, Colombia

José Raúl Canay Pazos[1], Wilfred Rivera Martínez[2](✉), and Carolina Quiñonez Zúñiga[3]

[1] Universidad Santiago de Compostela, Santiago de Compostela, Spain
raul.canay@usc.es

[2] Corporación Universitaria Autónoma del Cauca, Popayán, Colombia
wilfred.rivera.m@uniautonoma.edu.co

[3] Centro Regional de Productividad e Innovación del Cauca-CREPIC,
Popayán, Colombia
carolina.quinonez@crepic.org.co

Abstract. Network Analysis, as a technique for understanding social phenomena, brings valuable methodological and instrumental elements, and can be used in a wide range of applications. This research approach analyzes the usage of Interinstitutional social networks to support climate change adaptation and food security in the Department of Cauca, Colombia. The unit of analysis is the Interinstitutional Climate Change and Food Security Network of Colombia - RICCLISA, which comprises 60 institutional actors who have expressed their interest in working collaboratively. They have the necessary intellectual capital to generate positive effects in this area. Methodologically, data was gathered from a survey to calculate indexes of network analysis to understand the dynamics and reticular characteristics of these relationships. To carry out the reticular analysis, data were taken from surveys done by [1] with these sixty stakeholders in Cauca. It is observed that organizations with the ability to articulate stakeholders with local policymaker's interests become real catalysts of the region's relational capital that favorably impact collaborative work on climate change and food security challenges. These interface organizations are synergistic and operate across a range of scales, purposes, and intensities of information which flow between stakeholders and audiences.

Keywords: Networking · Relational capital · Climate change · Food security

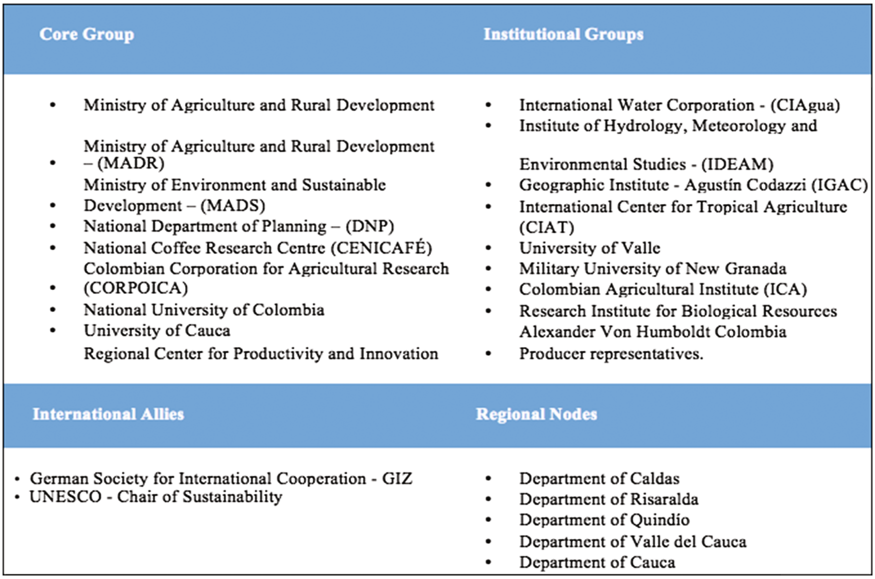
1 Introduction

Networks promote social capital by regulating coexistence through norms that guarantee a culture of trust [2]. However, a social network is not an isolated entity, but a variety of entities that have characteristics in common. They coincide in certain aspect of the social structure and facilitate individuals to realize certain actions [3]. Under this

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5_19

premise, the public and private stakeholders of Science, Technology and Innovation, related to Climate and Agriculture, created in 2013 the Interinstitutional Network of Climate Change and Food Security of Colombia, known by its acronym in Spanish – RICCLISA, to articulate and integrate agents to analyze and design strategies, procedures, and actions to develop and implement innovative alerting processes in the Colombian agricultural sector. This will adapt, conserve, and mitigate the effects of climate change in ecosystems, its biodiversity, and the productive capacity to sustain the competitiveness and food security of the Colombian population. For its operation, the RICLISSA Network has set up a structure that balances the different national and regional stakeholders. In that sense, the entity is led by a Core Group, composed of experts and supported by a technical committee and an articulating organization. The Institutional Group is formed by experts who belong to public and private organizations interested in participating in the process. It also has the participation of International Allies acting in Colombia and deployed in local areas. Five Regional Nodes conformed by institutions from each territorial entity jointly implement permanent actions to ensure the achievement of RICCLISA's mission in agreement with the initiatives posed by the Core Group (Table 1).

Table 1. Structure of RICLISSA Network. Source: Own elaboration



Core Group	Institutional Groups
• Ministry of Agriculture and Rural Development • Ministry of Agriculture and Rural Development – (MADR) • Ministry of Environment and Sustainable Development – (MADS) • National Department of Planning – (DNP) • National Coffee Research Centre (CENICAFÉ) • Colombian Corporation for Agricultural Research (CORPOICA) • National University of Colombia • University of Cauca Regional Center for Productivity and Innovation	• International Water Corporation - (CIAgua) • Institute of Hydrology, Meteorology and Environmental Studies - (IDEAM) • Geographic Institute - Agustín Codazzi (IGAC) • International Center for Tropical Agriculture (CIAT) • University of Valle • Military University of New Granada • Colombian Agricultural Institute (ICA) • Research Institute for Biological Resources Alexander Von Humboldt Colombia • Producer representatives.
International Allies	**Regional Nodes**
• German Society for International Cooperation - GIZ • UNESCO - Chair of Sustainability	• Department of Caldas • Department of Risaralda • Department of Quindio • Department of Valle del Cauca • Department of Cauca

2 Literature Review

2.1 Relational Capital as a Key Element of Intellectual Capital in Organizations

There are common elements in the different conceptions of relational capital. In the first place, it stands out the identification of knowledge as a key factor that is incorporated in the management of relationships [4] with a practical application in organizations. According to this perspective, relational capital is emphasized as a component of intellectual capital that must be demonstrated through the organization's goals. In this sense, the management of social networks tends to be a planned, organized, accomplished, and monitored process that lead the actors' efforts toward the achievement of these goals.

Another repetitive element when defining relational capital are networks, understood as internal and external, formal and informal interactions that facilitate the allocation and progressive distribution of resources to achieve a higher level of performance [5].

Relational capital is a useful resource; because it improves the organization's internal trust by connecting actors and external networks that provide resources [6, 7]. A fundamental element here lies in trust, often as a result of obligations, hierarchical authority, and interaction [8, 9]. This trust forms a kind of bond that holds the organizations together. Another aspect that makes up relational capital are ties that provide resources such as information [7]. The bonds that result from this relationship can occur at both, in the individual and organizational level, although they are often attributed mainly to individual agents. These links can be direct or indirect, their intensity can vary, and the results (in terms of linkage or connections of relational capital) depend on the type of network analyzed.

Granovetter [10] stresses on the importance of maintaining an extensive network of "weak ties" in obtaining resources (information on potential job opportunities). Weak ties are not very strong relationships between individuals, as opposed to the close ties as they would be found in a family. Weak ties are useful for obtaining information that would otherwise not be available or would be expensive to obtain. They extend their network by uniting individuals or organizations and providing an interface for exchanges.

The influence of the environment is another factor that characterizes the different conceptions of relational capital. In accordance to Ordóñez de Pablos [11], organizations cannot be considered as isolated systems, while Roos [12] emphasizes that relational capital is largely composed of the value generated by external interest groups. At least from the theoretical perspective there is a consensus that the sphere of action of relational capital is external and involves actors that can invigorate organizations as a result of interaction and joint work.

Stakeholders are agents whose interaction builds relational capital [13], in accordance to Larson and Starr [14], the greater the number of ties and their quality, the better the results for the organization. Relational capital bases its theoretical conception, according to the authors, in the interlinkages among stakeholders.

Relational capital often materializes through the identification of network relationships, sometimes defined by the strength of these ties, regular group activities, or the frequency of meetings and other formal interactions, as well as informal meetings and other social and family activities. From the organizational point of view, relational capital provides contacts that facilitate the discovery of opportunities, as well as the identification, collection, and allocation of scarce resources [15, 16]. Relational capital can also help with the collective growth process by providing and disseminating critical information and other essential resources.

Finally, we highlight the generation of positive externalities. Yang and Lin [17] consider the relational capital as the resulting value of actor's relationships. Correspondingly, Petrash [18] concluded that it is the society's perception of value when doing business, solving a problem, or diminishing a negative situation. In sum, the long-term purpose of the relational capital management should lead to the generation of favorable effects in society, which implies a complex exercise in appraising the intangible, also exhorts the academic community and society to think on this type of tools not only from the theoretical approach but also from its implications.

The Intellectus model, developed by the Foro del Conocimiento INTELLECTUS, has been chosen among the most widespread models. This model is based on the review and analysis of the different approaches to intellectual capital, especially the Intellect model in 1998. It is characterized by its open and flexible nature in relating intellectual capital to the organization's strategy; its structure is composed of blocks, elements and indicators. The blocks of intellectual capital are: human capital, structural capital (which in turn is formed by organizational capital and technological capital) and relational capital (constituted by business capital and social capital). The elements in turn represent the resources or intangible assets that integrate each component. The indicators show how to evaluate each of the previous elements [3].

3 Methodology

The analysis of the interinstitutional relational capital to support climate change mitigation and food security in Cauca originated from a mixed methodological design, in which the RICLISSA Network was the unit of analysis. This approach uses both quantitative and qualitative data collection and analysis tools that combined "provide a better understanding of the research problem" [18]. As a method it is widely used in social sciences [19] and very useful for research, since it enables a deep understanding of the phenomena under observation [20].

In this sense, the reticular analysis was carried out with the information collected from the sixty (60) members of the RICCLISA network in Cauca. These actors have built a collaborative work dynamics using information and communications technologies that support works on climate change mitigation and food security. Although these types of networks are perceived by the institutional fabric of the region and of the country, there are no previous exercises for measuring and quantifying this impact. Therefore, the techniques described in the methodology section were used to measure the influence of relational capital created among these actors that identifies strategic nodes for the management and possible replication of this dynamic.

In addition, a focus group was developed with 10 local experts and members of regional institutions. The purpose of these tools was to consult and discuss local dynamics, competencies, and key factors to mitigate climate change and contribute to food security.

With the information collected, the multidisciplinary research team processed the information to determine the most important metrics to measure the capacity of the Network [21]. The Centrality, Degree, Betweenness, and Closeness among the Nodal Network of Cauca stakeholders, who serve as a referent to define strategies for strengthening the relational capital of the actors involved in Cauca.

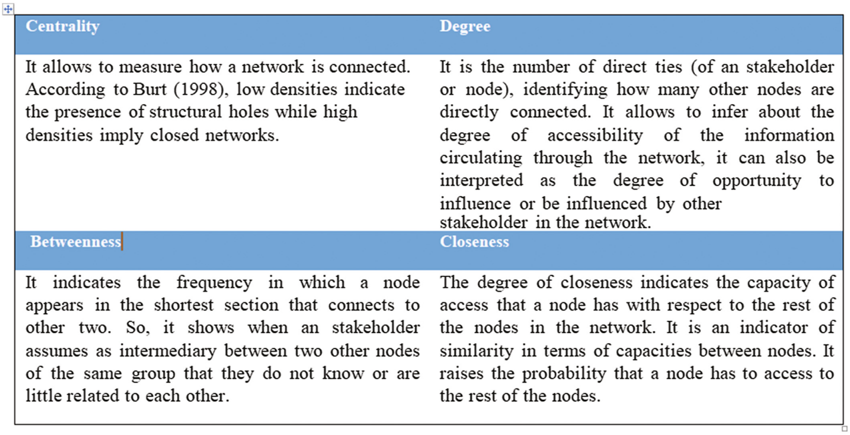
To analyze the dynamics of the activities, several instruments were applied: field diary to record the dynamics of the relationship; In-depth interviews to capture the impressions and experiences of the stakeholders, group sessions to share opinions and an individual form to qualify the proposed reticular variables.

The data were processed by the UCINET software v6.0 - quantitative data - and TextStat - qualitative data - and then synthesized in a comparative matrix, synoptic tables and conceptual maps according to the categories of analysis. The results were incorporated into the strategy design process in each of its stages: exploration, structuring, and evaluation-improvement.

In this research, social networks consisted of two elements: organizations (nodes) and the relationships (ties) between them. Once all the nodes and ties are known, it is possible to draw pictures of the network and discern every organization's position within it. Within a network, the "distance" between two organizations (also known as the "geodesic distance" or "degree of separation"), shows the shortest path in the network from one organization to another. The basic idea in social network analysis is that organizations are influenced by their location in a social network and by the happenings among organizations that are "nearby" them in the social network.

Once a full set of organizations and ties is observed, there is only one "network" *per se*. This network, however, can be analyzed or drawn according to the following metrics (Table 2).

Table 2. Indicators used for reticular analysis in RICLISSA. Source: [22].



Centrality	Degree
It allows to measure how a network is connected. According to Burt (1998), low densities indicate the presence of structural holes while high densities imply closed networks.	It is the number of direct ties (of an stakeholder or node), identifying how many other nodes are directly connected. It allows to infer about the degree of accessibility of the information circulating through the network, it can also be interpreted as the degree of opportunity to influence or be influenced by other stakeholder in the network.
Betweenness	**Closeness**
It indicates the frequency in which a node appears in the shortest section that connects to other two. So, it shows when an stakeholder assumes as intermediary between two other nodes of the same group that they do not know or are little related to each other.	The degree of closeness indicates the capacity of access that a node has with respect to the rest of the nodes in the network. It is an indicator of similarity in terms of capacities between nodes. It raises the probability that a node has to access to the rest of the nodes.

4 Findings

The last years have witnessed the development of a new way of studying social structure: the so-called network analysis, which has reached high levels of methodological and technical sophistication, showing its high value in a wide range of applications. In this case, it analyzes networks of relations and their effects on the institutional framework that may influence the mitigation of climate change and the variation of food security.

From the above, an important set of statistical indicators summarize both the characteristics of the network and the stakeholders, as these indicators refer to Centrality, as to Closeness, Betweenness, and Degree in order to identify the network's central stakeholders by identifying positions of relevance and power [22]. The sixty (60) institutional stakeholders of the RICLISSA Network and the label with which each was nominated for the purposes of processing and analysis are listed below (Table 3).

4.1 Relational Capital Strengthening

Degree

The analysis of the degree (Table 4 and Fig. 1) shows the preponderance of the Regional Center for Productivity and Innovation of Cauca - CREPIC – that has a direct relationship to 33 stakeholders (56% of total), followed by the Río Piedras Foundation with 29 links, the Secretary of Planning of the Government of Cauca – GCAUSPLD - with 26 links, SUPRACAFÉ COLOMBIA SA with 22 links. The centrality of the Network is 47.63%, which suggests a high level of dispersion, but also shows three strong relationship nodes around CREPIC, Río Piedras Foundation and Secretary of Planning of the Government of Cauca – GCAUSPLD.

For the purposes of the Network, these three organizations have the capacity to connect other stakeholders and coordinate efforts around climate change and food security. In the case of CREPIC and Rio Piedras Foundation, these organizations have an extensive experience on rural development projects. They have the recognition of institutions and rural/urban communities, which is a key element that generates confidence when developing collaborative work.

The Regional Center for Productivity and Innovation of Cauca (CREPIC) was created in 2000 by private/public institutions to improve the competitiveness of regional organizations in Colombia through the articulation and development of the innovative and productivity capacities of its stakeholders. According to its mission statement and the perception of its stakeholders, the organization articulates regional capacities to increase quality of life. On the same issue, the Rio Piedras Foundation is responsible for ensuring the sustainability of water supply and conservation of the main water supply basins in the city of Popayan (Cauca). It was created by 11 public/private organizations, including: The Mayor of Popayán, Fish Farm El Diviso, University of Cauca, The Governor of Cauca, Smurfit Cardboard of Colombia, Special Administrative Unit of Natural Parks of Colombia, Departmental Coffee Growers Committee of Cauca, FUP University, Cedelca Power Plants, CRC, Water and Sewerage Company of

Table 3. List of stakeholders of the RICLISSA network. Source. Own elaboration based on data from [1]

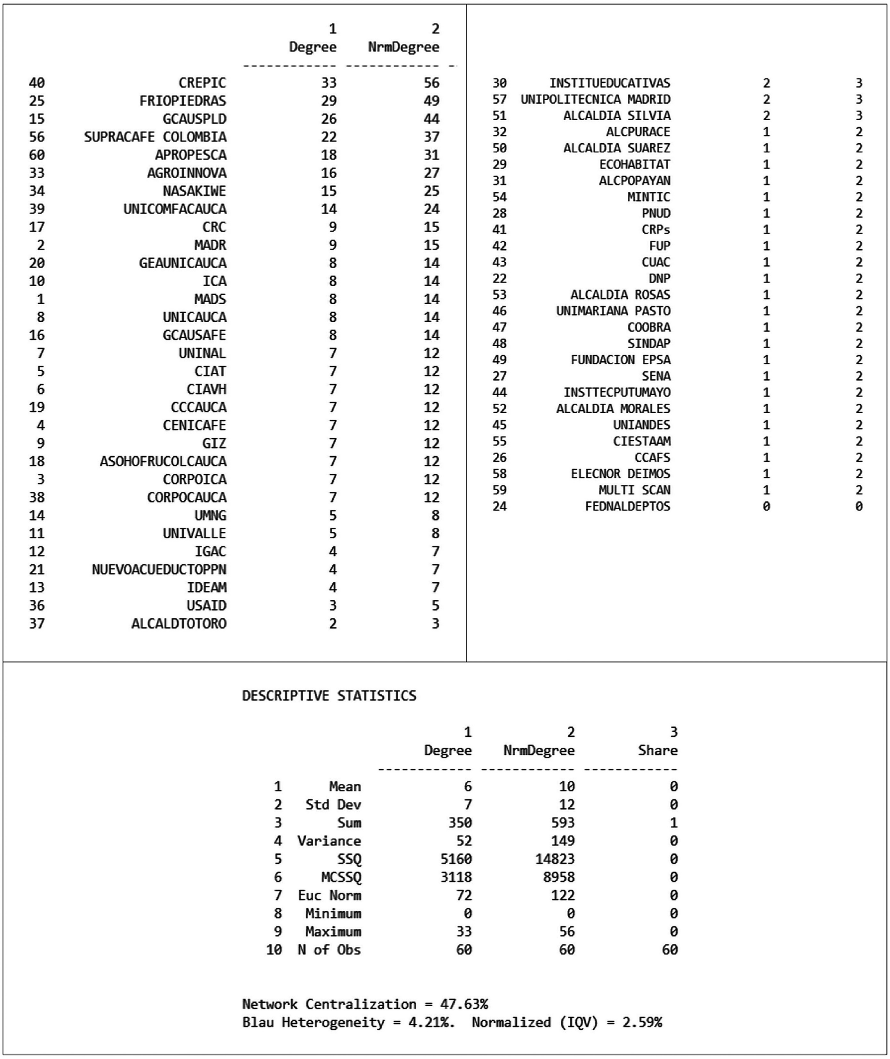
ACTOR´S NAME	LABEL
Ministry of Environment and Sustainable Development	MADS
Ministry of Agriculture and Rural Development	MADR
Colombian Corporation of Agricultural Research	CORPOICA
International Center for Tropical Agriculture	CIAT
National University of Colombia	UNINAL
University of Cauca	UNICAUCA
German Society for International Cooperation - GIZ	GIZ
Colombian Agricultural Institute	ICA
University of Valle	UNIVALLE
Geographic Institute Agustín Codazzi	IGAC
Institute of Hydrology, Meteorology and Environmental Studies	IDEAM
Military University of New Granada	UMNG
Secretary of Planning of the Government of Cauca	GCAUSPLD
Secretary of Agricultural Development and Economic Development	GCAUSAFE
Autonomous Regional Corporation of Cauca	CRC
Horticultural Association of Colombia - Cauca	ASOHOFRUCOLCAUCA
Chamber of Commerce of Cauca	CCCAUCA
Research group in environmental studies of the University of Cauca	GEAUNICAUCA
Aqueduct of Popayán	NUEVOACUEDUCTOPPN
National Department of Planning	DNP
Administrative Department of Science, Technology and Innovation	COLCIENCIAS
National Federation of Departments	FEDNALDEPTOS
Río Piedras Foundation	FRIOPIEDRAS
Climate Change, Agriculture and Food Security	CCAFS
National Learning Service	SENA
United Nations Development Program	PNUD
Eco-Habitat Foundation	ECOHABITAT
Representative of educational institutions of Cauca	INSTITUEDUCATIVAS
Mayor of Popayán. Cauca	ALCPOPAYAN
Mayor of Puracé. Cauca	ALCPURACE
Incubator of agribusinesses	AGROINNOVA
Nasa Kiwe Corporation	NASAKIWE
Institution for education Enrique Vallejo de Belacazar	NORMALNALBELAC

United States Agency for International Development,	USAID
Mayor of Totoró. Cauca	ALCALDTOTORO
Corporation for the development of Cauca	CORPOCAUCA
Higher Education Institution UNICOMFACAUCA	UNICOMFACAUCA
Regional Center for Productivity and Innovation of Cauca	CREPIC
Representative Regional Productivity Centers	CRP's
Higher Education Institution, FUP	FUP
Higher Education Institution, Uniautonoma	CUAC
Technological Institute of Putumayo	INSTTECPUTUMAYO
University of Los Andes	UNIANDES
Mariana University	UNIMARIANA PASTO
Cooperative of beneficiaries of the Cauca agrarian reform	COOBRA
Corporation for research, agricultural and environmental development	SINDAP
Foundation of the Pacific energy company S.A.	FUNDACION EPSA
Mayor of Suarez. Cauca	ALCALDÍA SUAREZ
Mayor of Silvia. Cauca	ALCALDÍA SILVIA
Mayor of Morales. Cauca	ALCALDÍA MORALES
Mayor of Rosas. Cauca	ALCALDÍA ROSAS
Ministry of Information and Communication Technologies.	MINTIC
Center for Economic, Social and Technological Research of Agribusiness and World Agriculture. University of Chapingo	CIESTAAM
Supracafe Colombia	SUPRACAFE COLOMBIA
Polytechnic University of Madrid	UNIPOLITECTICA MADRID
Elecnor Group Technology Area	ELECNOR DEIMOS
MULTISCAN Technologies	MULTI SCAN
Association for the production and commercialization of aquaculture and agricultural products of Silvia, Cauca	APROPESCA

Popayán S.A. As can be appreciated, articulation is in the DNA of CREPIC and Rio Piedras Foundation.

The Secretary of Planning of the Government of Cauca – GCAUSPLD, has the capacity of managing and financing initiatives allowing a greater level of effectiveness in achieving interinstitutional approaches. So, according to these results and the characteristics of these stakeholders, the recognition for their community work, the ties of trust established through collaborative work, and the possibility of financing are, in short, the key factors that explain the high number of links that these three organizations have with other stakeholders in the analysis of this reticular variable.

Table 4. Variable analysis range of the RICLISSA network. Source. Own elaboration based on data from [1]



```
                                 1            2
                            Degree    NrmDegree
                      ------------ ------------ -
40               CREPIC         33           56     30       INSTITUEDUCATIVAS          2           3
25          FRIOPIEDRAS         29           49     57  UNIPOLITECNICA MADRID          2           3
15             GCAUSPLD         26           44     51        ALCALDIA SILVIA          2           3
56   SUPRACAFE COLOMBIA         22           37     32              ALCPURACE          1           2
60            APROPESCA         18           31     50        ALCALDIA SUAREZ          1           2
33           AGROINNOVA         16           27     29             ECOHABITAT          1           2
34             NASAKIWE         15           25     31             ALCPOPAYAN          1           2
39        UNICOMFACAUCA         14           24     54                MINTIC           1           2
17                  CRC          9           15     28                  PNUD           1           2
 2                 MADR          9           15     41                  CRPs           1           2
20          GEAUNICAUCA          8           14     42                   FUP           1           2
10                  ICA          8           14     43                  CUAC           1           2
 1                 MADS          8           14     22                   DNP           1           2
 8             UNICAUCA          8           14     53        ALCALDIA ROSAS           1           2
16             GCAUSAFE          8           14     46       UNIMARIANA PASTO          1           2
 7               UNINAL          7           12     47                COOBRA           1           2
 5                 CIAT          7           12     48                SINDAP           1           2
 6                CIAVH          7           12     49        FUNDACION EPSA           1           2
19              CCCAUCA          7           12     27                  SENA           1           2
 4             CENICAFE          7           12     44        INSTTECPUTUMAYO          1           2
 9                  GIZ          7           12     52      ALCALDIA MORALES           1           2
18     ASOHOFRUCOLCAUCA          7           12     45              UNIANDES           1           2
 3             CORPOICA          7           12     55              CIESTAAM           1           2
38           CORPOCAUCA          7           12     26                 CCAFS           1           2
14                 UMNG          5            8     58         ELECNOR DEIMOS          1           2
11             UNIVALLE          5            8     59            MULTI SCAN           1           2
12                 IGAC          4            7     24           FEDNALDEPTOS          0           0
21    NUEVOACUEDUCTOPPN          4            7
13                IDEAM          4            7
36                USAID          3            5
37          ALCALDTOTORO         2            3
```

```
DESCRIPTIVE STATISTICS

                            1            2            3
                       Degree    NrmDegree        Share
                 ------------ ------------ ------------
 1       Mean               6           10            0
 2    Std Dev               7           12            0
 3        Sum             350          593            1
 4   Variance              52          149            0
 5        SSQ            5160        14823            0
 6      MCSSQ            3118         8958            0
 7   Euc Norm              72          122            0
 8    Minimum               0            0            0
 9    Maximum              33           56            0
10   N of Obs              60           60           60


Network Centralization = 47.63%
Blau Heterogeneity = 4.21%.  Normalized (IQV) = 2.59%
```

Information and communication technologies – ICT - are the channel that has facilitated the establishment of links and communication between these actors. Once trust is created between organizations, the use of ICT allows to monitor and to maintain permanent communication links.

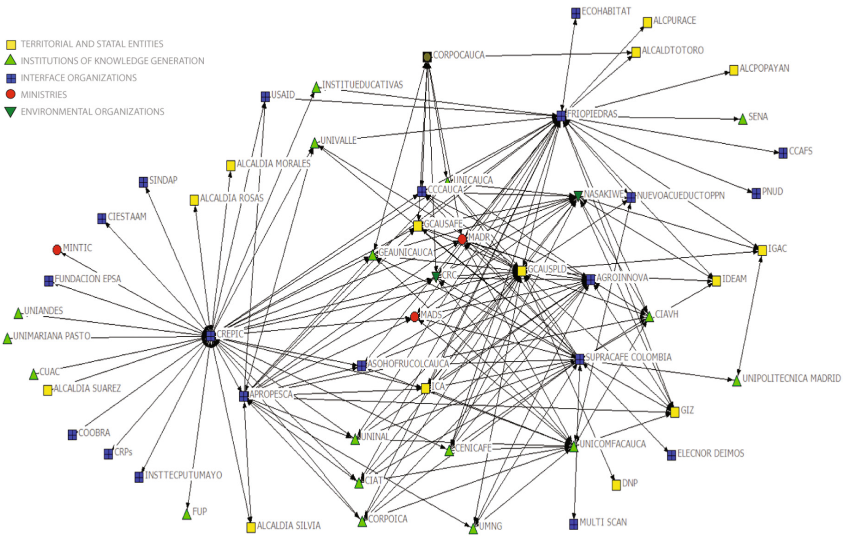


Fig. 1. Relational capital of the RICLISSA Network. Source: Own elaboration based on data from [1]

Betweenness

The information in Table 5 shows that CREPIC is the most active stakeholder between those involved in the process, followed by the Río Piedras Foundation, and Secretary of Planning of the Government of Cauca and SUPRACAFÉ COLOMBIA SA.

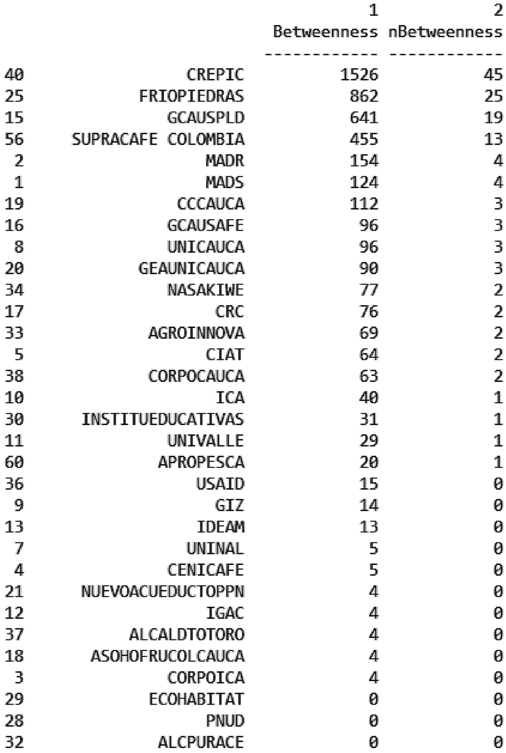
It is worth noting that this index measures the possibilities that stakeholders have in connecting with those who they do not know or have not work together. Both CREPIC and the Rio Piedras Foundation consider articulation as a key element in their strategic platform which allows them to effectively achieve their objectives, hence its ability to interact and connect with different stakeholders. On the other hand, the Secretary of Planning of the Government of Cauca, for its role as coordinator of the implementation of the Development Plan of the region and Adviser in the investment of regional, national, and international cooperation agencies, it has a holistic view of the regional dynamics and therefore of the different institutional stakeholders who can cooperate in the implementation of different initiatives.

Their role as articulating organizations and the knowledge of the different territorial dynamics related to climate change and food security are essentially the characteristics that explain the high levels of intermediation that these three organizations have.

Closeness

The degree of closeness indicates the access capacity that a stakeholder has with respect to other actors in the network. The indicator of access between intervening actors favors the Secretary of Planning of the Government of Cauca, followed by the Río Piedras Foundation, the Ministry of the Environment and Rural Development, the Local Agency for the Management of Resources for Environment, known by its

Table 5. Variable analysis of betweenness of RICLISSA. Source. Own elaboration based on data from [1].



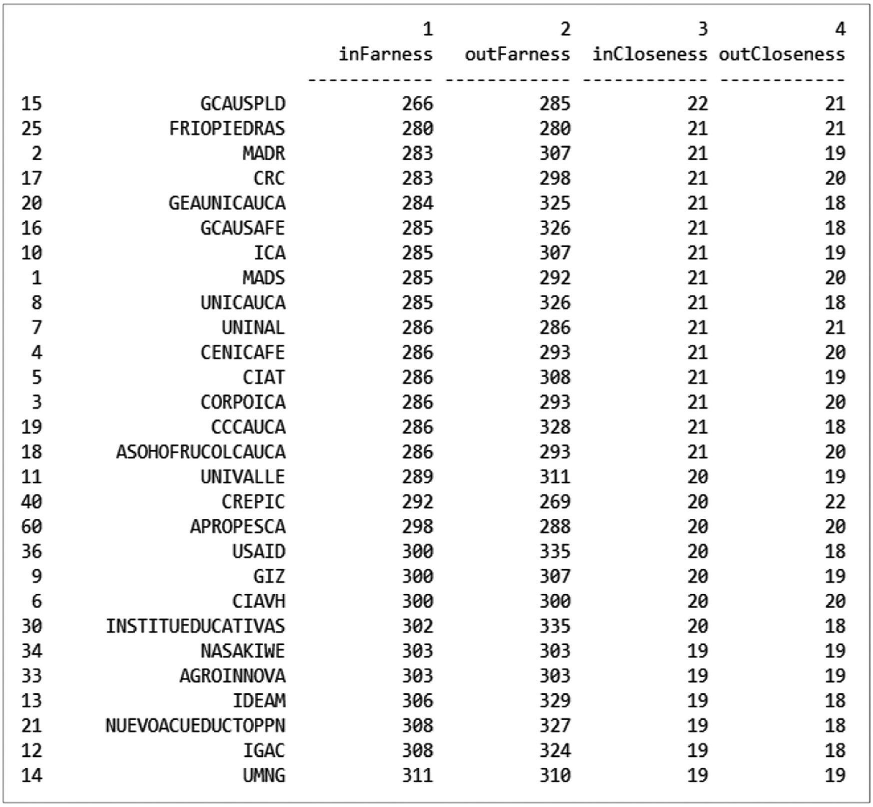
```
                                   1            2
                         Betweenness nBetweenness
                        ------------ ------------
40              CREPIC          1526           45
25         FRIOPIEDRAS           862           25
15            GCAUSPLD           641           19
56   SUPRACAFE COLOMBIA          455           13
 2                MADR           154            4
 1                MADS           124            4
19             CCCAUCA           112            3
16             GCAUSAFE           96            3
 8             UNICAUCA           96            3
20          GEAUNICAUCA           90            3
34             NASAKIWE           77            2
17                  CRC           76            2
33           AGROINNOVA           69            2
 5                 CIAT           64            2
38           CORPOCAUCA           63            2
10                  ICA           40            1
30    INSTITUEDUCATIVAS           31            1
11             UNIVALLE           29            1
60            APROPESCA           20            1
36                USAID           15            0
 9                  GIZ           14            0
13                IDEAM           13            0
 7              UNINAL             5            0
 4             CENICAFE            5            0
21   NUEVOACUEDUCTOPPN             4            0
12                 IGAC            4            0
37         ALCALDTOTORO            4            0
18      ASOHOFRUCOLCAUCA           4            0
 3             CORPOICA            4            0
29            ECOHABITAT           0            0
28                 PNUD            0            0
32            ALCPURACE            0            0
```

acronym in Spanish as – CRC - and the Environmental Studies Research Group of the University of Cauca – GEAUNICAUCA (Table 6).

The Ministry of the Environment is the public entity in charge of defining the national environmental policy and ensuring sustainable development, and the Secretary of Planning of the Government of Cauca. It sets the policies at a regional level, both institutions influence directly over all the Network, orienting the actions of actors to the fulfillment of the public policy on environmental issues posed by the National and Regional government.

Statistical analysis reflects the capacity of these organizations to reach all the network nodes. In the case of the Ministry of the Environment and Sustainable Development – MADS - and the Secretary of Planning of the Government of Cauca, they act as a policymaker with a strong influence over other actors. On the other hand, Río Piedras Foundation, and its public/private composition, holds a strong relational capital that is useful at different levels.

Table 6. Variable analysis degree of closeness of the RICLISSA Network. Source. Own elaboration based on data from [1]



```
                                       1            2            3            4
                               inFarness   outFarness  inCloseness outCloseness
                            ------------ ------------ ------------ ------------
15             GCAUSPLD            266          285           22           21
25          FRIOPIEDRAS            280          280           21           21
 2                 MADR            283          307           21           19
17                  CRC            283          298           21           20
20          GEAUNICAUCA            284          325           21           18
16             GCAUSAFE            285          326           21           18
10                  ICA            285          307           21           19
 1                 MADS            285          292           21           20
 8             UNICAUCA            285          326           21           18
 7               UNINAL            286          286           21           21
 4             CENICAFE            286          293           21           20
 5                 CIAT            286          308           21           19
 3             CORPOICA            286          293           21           20
19              CCCAUCA            286          328           21           18
18     ASOHOFRUCOLCAUCA            286          293           21           20
11             UNIVALLE            289          311           20           19
40               CREPIC            292          269           20           22
60            APROPESCA            298          288           20           20
36                USAID            300          335           20           18
 9                  GIZ            300          307           20           19
 6                CIAVH            300          300           20           20
30    INSTITUEDUCATIVAS            302          335           20           18
34             NASAKIWE            303          303           19           19
33           AGROINNOVA            303          303           19           19
13                IDEAM            306          329           19           18
21   NUEVOACUEDUCTOPPN             308          327           19           18
12                 IGAC            308          324           19           18
14                 UMNG            311          310           19           19
```

A Qualitative Analysis of the Contributions of the RICLISSA Network

The two (2) focus groups with ten institutions addressed the following topics: (i) the characteristics of the networks formed in RICCLISA, (ii) the Network's contributions to climate change and food security, and (iii) the network's challenges.

Among the common denominators were: the critical mass of actors constituted in Cauca by the 60 participants, and the necessity to give continuity to the advanced activities, an element that in the institutional sphere is difficult to manage due to changes in the administration. As isolated factors were presented: the management of resources, the training of human talent, and the publicity and dissemination of results, elements that although important, are not shared by all attendees.

The analysis carried out from the focus groups show the potential of Cauca node due to the sixty (60) stakeholders that make it up offering possibilities to collaborative and interdisciplinary work.

The actors opinion emphasize in the necessity of strengthening these institutional capacities in two ways. In the first place, it is necessary to improve the monitoring and alerting of the climate behavior and the externalities generated in terms of food security. In this way, the research work and efforts made by the Colombian state to plan and control these effects are important to avoid disasters either on society or on the productive sector of the region.

Furthermore, it is fundamental to develop appropriation exercises that include awareness, training, and management so communities and productive sectors increase their commitment and have the necessary preparation. In this way, the collaborative work and technology transfer promote jointly efforts to mitigate climatic change.

5 Conclusions

The analysis of the variables Centrality, Degree, Closeness, and Betweenness, shows high dispersion among actors that generates difficulties in the management of the Network. Although from the reticular perspective, this condition expands the relational capital and allows the structuring of ambitious action plans that use the contacts of a network and each organization capacities. It is necessary to establish flexible and practical mechanisms that allow the Network to fulfill its objectives without falling into the inefficiencies of large networks. Undoubtedly, the challenges of mitigating climate change effects and ensuring food security are high-level purposes for the State and Society, which will demand the usage of appropriate management technologies to keep the actors of the Network integrated and avoid the natural desertions in this type of processes.

After analyzing the reticular variables of the RICLISSA Network in Cauca, one of the strengths is the ability of certain actors to articulate the work of different agents of climate change and food security. It is evident that the binomial Government/Entities interface enhances the efforts of the actors. Specifically, the joint work between a government actor - the Secretary of Planning of Cauca with interface organizations - CREPIC and the Río Piedras Foundation - allows the effects of the triple propeller - University, Enterprise, State - proposed by Etzkowitz and Leydesdorff [23] to deploy their potential in the region.

Both, the analyzed statistical data and the interviews demonstrated that academic research, linked to the efficiency of the productive sector and accompanied with public policy formulation capacity of the State, can generate positive long-term impacts on society in environmental terms. The achievements and advances of this network in the future open new lines for further research.

The collected quantitative and qualitative data coincide with the wide capacity that Cauca has on facing RICLISSA challenges. In order to catalyze this dynamics, there are organizations called to lead the process due to its attractive reticular properties, such is the case of the Regional Center for Productivity and Innovation of Cauca, the Río Piedras Foundation, and the Secretary of Planning of Cauca. From the society's perspective, it has been detected the importance of developing exercises of appropriation that permit the intensive usage of these mechanisms on citizens daily life. The meeting point of these aspects will certainly result in higher possibilities for public policies in climate change mitigation and food security to benefit the population and productive sector.

References

1. Amaya, M.: Fortalecimiento de la Red Interistitucional de Cambio Climático y Seguridad Alimentaria (RICCLISA) y su articulación en el contexto de la asociatividad para la Gestión y Transferencia de conocimiento, la mitigación del cambio climático, la seguridad y la soberanía alimentarias. Documento de trabajo. Programa RICCLISA (2015)
2. Putnam, R.: Making Democracy Work. University Press, Princeton (1994)
3. Coleman, J.: Foundations of Social Theory. Belknap Press, Cambridge (1990)
4. Capital Intelectual: Modelo Intellectus: Medición y gestión del capital intelectual. [Documento 5]. Centro de Investigación sobre la Sociedad del Conocimiento (CIC), Universidad Autónoma de Madrid. Documentos Intellectus (2003)
5. Stevenson, H.H., Jarillo, J.C.A.: Paradigm of entrepreneurship: entrepreneurial management. Strateg. Manag. J. **11**, 17–27 (1990)
6. Adler, P., Kwon, S.: Social capital: prospects for a new concept. Acad. Manag. Rev. **27**(1), 17–40 (2002)
7. Putnam, R.D.: Bowling alone: America's declining social capital. In: Culture and Politics, pp. 223–234. Palgrave Macmillan US, New York (2000)
8. Coleman, J.S.: Social capital in the creation of human capital. Am. J. Sociol. **94**, S95–S120 (1988)
9. Granovetter, M.: Economic action and social structure: the problem of embeddedness. Am. J. Sociol. **91**(3), 481–510 (1985)
10. Granovetter, M.S.: The strength of weak ties. Am. J. Sociol. **78**(6), 1360–1380 (1973)
11. Ordóñez de Pablos, P.: Intellectual capital reporting in Spain: a comparative view. J. Intellect. Capital **4**(1), 61–81 (2003)
12. Roos, G.: Intellectual capital analysis as a strategic tool. Strateg. Leadership **29**(4), 21–26 (2001)
13. Inkinen, H.: Review of empirical research on intellectual capital and firm performance. J. Intellect. Capital **16**(3), 518–565 (2015)
14. Larson, A., Starr, J.A.: A network model of organization formation. Entrepreneurship Theory Pract. **17**(2), 5–16 (1993)
15. Birley, S.: The role of networks in the entrepreneurial process. J. Bus. Ventur. **1**(1), 107–117 (1986)
16. Uzzi, B.: The role of networks in the entrepreneurial process. J. Bus. Ventur. **1**, 107–117 (1999)
17. Yang, C., Lin, C.: Does intellectual capital mediate the relationship between HRM and organizational performance? Perspective of a healthcare industry in Taiwan. Int. J. Hum. Res. Manag. **20**(9), 1965–1984 (2009)
18. Petrash, G.: Dow's journey to a knowledge value management culture. Eur. Manag. J. **14**(4), 365–373 (1996)
19. Creswell, J.W., Shope, R., Plano Clark, V.L., Green, D.O.: How interpretive qualitative research extends mixed methods research. Res. Schools **13**(1), 1–11 (2006)
20. Pereira Pérez, Z.: Los diseños de método mixto en la investigación en educación: Una experiencia concreta. Revista Electrónica Educare **15**(1), 15–16 (2011)
21. Pole, K.: Diseño de metodologías mixtas. Una revisión de las estrategias para combinar metodologías cuantitativas y cualitativas (2009)
22. Velázquez, A.: El valor agregado de las redes sociales: Propuesta metodológica para análisis del capital social. Redes. Revista hispana para el análisis de redes sociales
23. Etzkowitz, H., Leydesdorff, L.: The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. Res. Policy **29**(2), 109–123 (2000)

Author Index

P. Angelov et al. (eds.), *Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change*, Advances in Intelligent Systems and Computing 687, https://doi.org/10.1007/978-3-319-70187-5