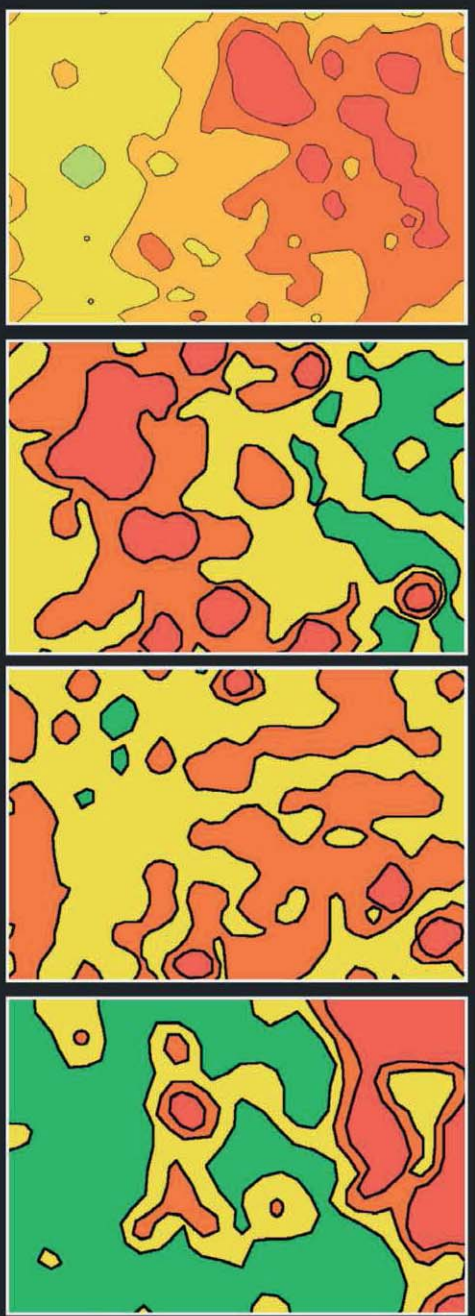
Plant Disease Epidemiology: Facing Challenges of the 21st Century

Edited by
S. Savary
B.M. Cooke



Springer

Plant disease epidemiology: facing challenges of the 21st Century

Plant disease epidemiology: facing challenges of the 21st Century

Under the aegis of an International Plant Disease Epidemiology Workshop held at Landernau, France, 10–15th April, 2005

Edited by
S. Savary and B.M. Cooke

Reprinted from *European Journal of Plant Pathology*, Volume 115 Issue 1, 2006

Springer

A C.I.P catalogue record for this book is available from the library of Congress

ISBN 1-4020-5019-4 (HB)
ISBN 1-4020-5020-8 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA, Dordrecht, The Netherlands

Printed on acid-free paper

Cover photos:
Patterns of change in multiple pathosystems over space: spatial distribution of four diseases in a groundnut plot, Côte d'Ivoire, France.
A- Groundnut rust, *Puccinia arachidis*;
B- Early leafspot, *Cercospora arachidicola*;
C- Late leafspot, *Cercosporidium personatum* (*Phaeoisariopsis personata*),
D- Web blight, *Rhizoctonia solani*.
Disease assessments were made at 90 days after sowing. Rust, early leaf spot, and late leaf spot: severity (% diseased leaf area) scales; web blight: incidence (% diseased plants) scale. From Lannou and Savary, 1991, modified.

Printed in the Netherlands

Contents

European Journal of Plant Pathology (2006) 115:1–2
DOI 10.1007/s10658-005-5946-6

Foreword

Plant Disease Epidemiology: Facing Challenges of the 21st Century

Plant disease epidemiology deals with diseases in plant populations. During the past century, it has become a vibrant field of science, achieving significant conceptual innovations with important impact on the management of plant diseases. Plant disease epidemiology mobilises concepts and methods from ecology, genetics, environmental physics, botany, and mathematics. It deals with cultivated and non-cultivated plants in environments where human activities have had large, or lesser, impact. As in many other fields of science, plant disease epidemiology faces important, sometimes new, questions. By and large, many of these questions emerge from changes in human societies and changes in the status of the planet on which we live.

Global climate is changing at a rapid rate: will it render plant diseases more, or less, harmful to man-made and spontaneous ecosystems? There is much debate on this issue, because global climate has varying, sometimes very large effects on the local environment of growing plant canopies, and because the physical micro-environment and its variation strongly influence plant diseases and their consequences on ecosystem functioning and performance; in addition, changes in global climate trigger many profound changes in the way ecosystems, cultivated or not, are managed. Interestingly, much of the early literature on botanical epidemiology dealt with climate-disease or climate-pathogen relationships – in fact these kinds of relationships have long been perceived as the bulk of epidemiological research by many. Plant disease epidemiologists thus have a strong scientific tradition in studying climate-pathogen-disease relationships. Can such an asset be mobilised by the epidemiological community to answer questions about the effect of climate change on plant diseases?

Global trade, and thus, trade of plant products, have increased at an unprecedented rate during the 20th century, and will continue to expand in the next century. Exchanges of plant materials at very different scales, local to global, have profound effects on plant diseases. Plant disease epidemiologists have become experts in assessing the risk of irruption of novel pathogens in plant communities, the consequences it may have on ecosystems, and ways to manage such perturbations. The concepts related to biological invasions or population displacements certainly are not new to plant pathologists: the epidemiological community in fact contributed to craft them in the past century. New threats may now also exist, whereby exotic or novel plant pathogens would intentionally be introduced: these threats must be dealt with. The consequences of plant pathogen transport are many: on local performances of spontaneous ecosystems and agricultural ecosystems; on farmers' livelihoods; on local, national, and regional economies; and perhaps more importantly, they can have adverse consequences on trade regulation. Will plant disease epidemiologists provide answers to such pressing questions?

Biodiversity, a buzzword of the past century, is also of global concern. The decline in global biodiversity that is currently taking place has been referred to as the sixth great extinction process our planet has experienced during its history, but this time, it is man-made. Generations of plant pathologists, and especially of plant disease epidemiologists, have been dealing with biodiversity. The huge diversity of life that resides in the rhizosphere and the phyllosphere are causes both of diseases in plants, and of their suppression. Much current research is addressing ways of harnessing such biodiversity not as enemies – of which pathogens are an inherent part – but rather as important biological allies to control disease epidemics. The diversity of plants is another facet of global biodiversity, and there are concerns about the decline in the genetic diversity of crop plants. It is from this diversity that possibly the most potent

instrument for disease management has been developed by plant pathologists: genetic host plant resistance. Will we run short of resistance genes against major plant pathogens? Host plant diversity, and the disease resistance genes it harbours, can be deployed over time and space, according to epidemiological principles. In-depth knowledge of the characteristics of individual pathogens causing specific diseases that must be controlled has been mobilised to develop appropriate strategies at the plant population, field, landscape, and sub-regional levels. Major successes have been achieved using such strategies, and the end of the past century has seen their recognition by the scientific community. Will epidemiologists succeed in the future in fully sharing these technologies with the farmer so that they are more fully utilised?

Food security was a central concern of the global agricultural research community in the middle of the 20th century, but apparently, not anymore. However, the world population still increases, and is expected to do so for several decades. One out of six human beings living on earth today suffers from lack of food. Many of today's poor live in cities, with no access to land and agriculture, and most of the projected increase in the world population will take place in the world's largest cities. Pests, including plant pathogens, cause losses in pre-harvest yield in the range of 20–40%; estimates of post-harvest losses are inadequate, but it is a fair assumption that they are often higher than 10 or 20%. Why are our estimates – the raison d'être of plant pathology – still so vague today? Seldom do economists currently address the issue of food security – why? Is it so that globalised exchanges, novel biological technologies, and the self-regulating mechanisms of trade, will be sufficient to fulfil the needs of future generations? Will these not have negative side-effects, and will they truly prevent the current over exploitation of natural resources, water and land in particular?

Sustainable production and crop protection systems need to be devised, which could exploit scarcer resources sparingly, and if possible enhance the resource base. Can these production and protection systems be designed so that they generate healthy, high-quality products that would find niche markets both locally and globally, and so provide farmers with the income they require, and offer consumers products that suit their needs and their incomes? Plant disease epidemiologists alone cannot provide answers to such questions, but certainly could significantly contribute to these new strategies.

The five-day International Plant Disease Epidemiology Workshop (held 10–15th April, 2005, in Landernau, France, the ninth of a series) reported in this special issue of the European Journal of Plant Pathology, obviously could not address all of these issues, and others, with all the depth good science demands. However it provided a unique opportunity for scientists interested in this field to meet and face challenging questions, contribute to animated debates, and reflect on the future development of the science of plant disease epidemiology.

Serge Savary
Mike Cooke

European Journal of Plant Pathology (2006) 115:3–23
DOI 10.1007/s10658-005-1229-5

Botanical epidemiology: some key advances and its continuing role in disease management

Laurence V. Madden
Department of Plant Pathology, The Ohio State University, Wooster, OH, 44691-4096, USA
(E-mail: madden.1@osu.edu)

Accepted 13 July 2005

Key words: basic reproduction number, Bayesian decision theory, epidemic, mixed models, model, population dynamics, receiver operating characteristic (ROC) curve

Abstract

Epidemiology involves the study of the temporal, spatial, and spatio-temporal dynamics of disease in populations, and the utilization of results of experiments and surveys to describe, understand, compare, and predict epidemics. Such understanding and description of epidemics can lead directly to the development and evaluation of efficient control strategies and tactics. Mathematical and statistical models are key tools of the epidemiologist. Recent advances in statistics, including linear and nonlinear mixed models, are allowing a more appropriate matching of data type and experimental (or survey) design to the statistical model used for analysis, in order to meet the objectives of the investigator. Coupled ordinary and partial differential equations, as well as simpler growth-curve equations, are especially useful deterministic models for representing plant disease development in fields in time and space over single seasons or many years, and their use can lead to appraisal of control strategies through metrics such as the basic reproduction number, a summary parameter that may be calculated for many general epidemic scenarios. Recently, compelling arguments have been made for the use of Bayesian decision theory in developing and evaluating real-time disease prediction rules, based on measured disease or weather conditions and either empirical or mechanistic models for disease or control intervention. Through some simple calculations of predictor accuracy and (prior) probability of an epidemic (or the need for control), the success of any predictor can be quantified in terms of the estimated probability of random observations being epidemics when predicted to be epidemics or not epidemics. Overall, despite the many contributions in epidemiology over the past four decades, more effort is still needed to convince those outside of epidemiology to more fully use epidemiological results and insights into the development and evaluation of disease controls.

Introduction

In 1963, van der Plank made a most compelling case for the importance of botanical epidemiology, both for understanding plant diseases at the population scale and for determining disease management strategies (van der Plank, 1963). He also made the bold statement at the time that 'epidemiology is here to stay.' Individual disciplines enjoy 'ups and downs' of popularity, of course, and epidemiology is no exception. The tremendous growth in the discipline within plant pathology during the 1960s, 1970s, and 1980s (e.g., Campbell and Madden, 1990; Kranz, 1990; Jones, 1998; Zadoks, 2001) has been eclipsed by growth in the larger field of molecular biology over the last two decades. Nevertheless, more than 40 years after van der Plank's book (1963), botanical epidemiology is still here, and still of utmost importance in giving a sound theoretical and practical basis for disease management. This view may not always be held outside of the discipline, however, and it

remains a challenge for epidemiologists to continue to make the compelling case that epidemiology matters.

Until molecular biology or more traditional breeding results in durable resistance to all plant pathogens on all crops, coupled with the acceptance of the new cultivars by growers and the public, there will be plant disease epidemics, and many of these will result in substantial reductions in yield. There is certainly increasing use of crop GMOs around the world (James, 2003), but cultivars with very broad-acting and durable resistance have yet to be developed. Moreover, the public opinion against their use remains strong in many regions; thus, it would be naïve to expect 'super resistant' cultivars in the foreseeable future. Use of fungicides and other chemicals in a protectant or curative manner is only practical for some crops and some diseases, and there is increasing societal pressure to (drastically) reduce the use of these chemicals in many regions. Thus, a scientific basis for applying or not applying chemicals is needed, and the decision clearly involves knowledge (or prediction) of the disease dynamics under different environmental conditions. The development of resistance to fungicides and antibiotics continues, and new cultivars have a finite lifetime.

No control tactics are known that will totally eliminate epidemics in crops and forests where the pathogen is present over large areas. Biological and cultural controls may be very beneficial, depending on the pathosystem (Maloy, 1993), but variability of control efficacy may be high with the former, and grower acceptance may be low with the latter (e.g., unwillingness to rotate crops).

The public and the scientific community have been definitely reminded of the importance of epidemiology, and the research tools that epidemiologists can bring to a problem, in recent years. A few examples are given. With increasing world trade of agricultural commodities as well as international travel, the risk of pathogen invasion of new countries or regions is well recognized (NRC, 2002), and predictions of the risk of invasion involve many epidemiological characteristics of pathogens, such as survival probabilities and reproductive potential (Madden and Wheelis, 2003). Moreover, the decision to attempt to eradicate or not also involves knowledge of disease epidemiology. The cases of citrus canker in Florida, karnal bunt in Arizona, and plum pox in Pennsylvania, U.S., are three examples of disease invasions (Gildow et al., 2004; Gottwald et al., 2001; Rush et al., 2005).

New pathogens (or pathogens new to a given crop) continue to be discovered, as well as strains, races, or biotypes of previously known pathogens. The new very aggressive biotype of African cassava mosaic virus in Africa is an example of a newly evolved isolate (Legg, 1999; Strange and Scott, 2005) that is proving very difficult to control. Sudden oak death, caused by *Phytophthora ramorum*, is a newly identified disease of oak and several other plant species, which is spreading naturally and (unfortunately) with the assistance of man, in the U.S. and elsewhere (Rizzo et al., 2002).

For diseases such as sudden oak death or Asian soybean rust (newly introduced into the U.S.), there is a great need to know the extent of spread from current locations (e.g., from the point of introduction) to other locations. For any disease that is locally concentrated (e.g., around the point of a new introduction), or does not yet exist in a country or region, ethically one cannot deliberately introduce the pathogen where it does not occur in order to study spore movement and resulting disease intensity. Thus, modelling is a key research tool for understanding risks based on key epidemiological characteristics or traits of a disease (Madden and van den Bosch, 2002; Madden and Wheelis, 2003). Epidemiology as a discipline depends heavily on the tools of mathematical and statistical modelling (Campbell and Madden, 1990), so epidemiologists are, in general, quite prepared to tackle the problem of disease spread through modelling. Model parameters for these types of situations can be obtained from observations where the disease of interest does occur naturally.

Most practicing epidemiologists would strongly support van der Plank's (1963) statement that epidemiology sets the strategy for disease control, and numerous examples can be given where epidemic knowledge leads to better control (Zadoks and Schein, 1979; Fry, 1982; Maloy, 1993). Furthermore, epidemiological principles and results can also lead to *specific* control recommendations, through the process of disease forecasting or risk prediction (Hardwick, 1998; Hughes et al., 1999), as demonstrated 45 years ago (Waggoner, 1960).

However, as pointed out recently by Jeger (2004), many controls are utilized and evaluated without explicit consideration of disease dynamics in fields. Although there is great danger in basing conclusions on disease intensity measured at one time in an epidemic (especially for polycyclic diseases; see Campbell and Madden, 1990), this unfortunately happens too often. Thus, epidemiologists still need to be pro-active in working with others in developing and evaluating disease control measures.

In the remainder of this article, I discuss a few developments that I consider to be very important in the development of plant disease epidemiology. Many more topics could have been covered. I have two major themes. One deals with the advancement in our theoretical understanding of the population-dynamic processes of disease spread in space and increase in time, coupled with the improvements in relating certain models (or their parameters) to empirical results (i.e., model fitting). The other theme deals with the prediction of plant disease on a real-time basis, or prediction of the need to impose a control measure, based on principles from probability theory. Citations are deliberately sparse, and are mainly to major reviews of topics rather than to all the (many) important original papers published over the last few decades. I assume throughout that modelling and statistical data analysis are methodological foundations for understanding epidemics and utilizing any gained knowledge in disease control.

Temporal and spatial dynamics of disease

Growth curve modelling and analysis

Van der Plank (1963) used the monomolecular and logistic equations as heuristic models of monocyclic (simple interest) and polycyclic (compound interest) disease epidemics. These models continue to be the benchmarks for quantification of epidemics, especially over single growing seasons. However, plant pathologists discovered in the 1960s and 1970s that these two models did not necessarily provide an adequate description (based on statistical principles of model fitting) for many disease progress curves (Campbell and Madden, 1990). Several alternative models were proposed or developed, some of them flexible in the sense that different degrees of skewness could be represented with the same model (depending on a realized value of a shape parameter). A feature of these models is that they are all based on a single response variable (disease intensity, y) in relation to continuous time, which can be obtained as a solution for the rate of change of y with time, dy/dt [e.g., $dy/dt = r_L y(1-y)$ for the logistic model]. In some cases, the solution can be expressed as a linear model, e.g., $\text{logit}(y) = a + r_L t$, where a is a transformation of disease intensity at time 0, r_L is the per capita rate parameter, and logit(y) is a linearizing transformation of y.

A good fit of an empirical model, or even a perfect fit, to data collected over time, is not proof of any mechanism for population growth (Campbell and Madden, 1990; Zadoks, 2001). But a good fit of a particular model for several disease progress curves could lead one to hypothesize about mechanisms, and then test the hypothesis with additional data or experiments. Moreover, using a model that provides a (reasonably) good fit to data is extremely important to accurately compare epidemics; among other things, using an inappropriate model will lead to biased estimates of the rate parameter and its standard error (Neter et al., 1983).

One clear trend in botanical epidemiology is the dramatically increasing complexity of statistical models and methods that have been applied to all epidemiological data over the last few decades (e.g., Gilligan, 2002; van Maanen and Xu, 2003). This is a natural development given the fact that epidemiology is a science of populations, and populations can only be adequately characterized and compared using the methodology of statistics. Although I am sure there are some who feel that the emphasis on mathematics and statistics obscures the understanding of the biology of epidemics, I would make the opposite claim, and declare emphatically that mathematical and statistical modelling are foundations for understanding epidemics. I further believe that, with some exceptions, the use of statistical analysis is actually still inadequate in most of epidemiological research, and certainly in most of plant pathology research! Many investigators still only: measure disease at a single time, do not match the chosen form of data analysis to the type of disease intensity variable (discrete for incidence, continuous but unequal variance for severity, ordinal for many disease rating scales); do not base their analysis on the

chosen experimental design; or perform inefficient (and sometimes uninformative) analyses. An example of the latter is the still common practice of performing a separate data analysis for each assessment time during an epidemic rather than simultaneously analyzing treatments (between-subject factors) and time (within-subject factors), and their interactions. Garrett et al. (2004) and citations therein can lead the reader to some of the important recent advances in statistical data analysis of relevance in plant pathology.

It has been known for many years (Madden, 1986) that disease values collected over time in the same experimental or sampling unit (e.g., plot) are serially correlated and that the variation in disease over time within plots is different from the variation between plots. This may be in part due to the cumulative nature of disease progress curves (see pp. 521–522 in Schabenberger and Pierce, 2002, for general discussion of cumulative processes over time). Serial correlations, sometimes called temporal autocorrelations, are especially troublesome in the comparison of treatments. My recent studies now show, however, that fitting of appropriate population-growth models to disease progress data often reduces the correlation of residuals to near zero for *individual* disease progress curves, reducing the need to directly utilize cumbersome adjustments to standard errors for calculated rates (unpublished). However, in the larger setting of multiple disease progress curves, corresponding to multiple treatment factors and blocks, there will always be non-zero correlations of observations *within* the plots by the nature of the experimental design (Schabenberger and Pierce, 2002). However, the structure of the correlations and variances may be quite complex, due to the cumulative process of disease development, but simple variance-covariance models can adjust for this property. For disease progress models that can be expressed in linear form through the use of a transformation of y [e.g., logit(y)], linear mixed models provide a tremendous (and still underutilized) tool for a thorough analysis of the epidemics (Garrett et al., 2004). Most plant pathologists (including epidemiologists) are not aware of the major advances made in mixed model analysis in statistics, a field that encompasses classical ANOVA and regression, and many other topics in a unified manner (Schabenberger and Pierce, 2002; Garrett et al., 2004). Instead of estimating disease progress model parameters for each epidemic, with a follow-up analysis of variance, through mixed models one can simultaneously estimate the disease progress parameters and their appropriate standard errors based on the explicit features of the design. The former approach (e.g., estimated slope for each plot, and then an ANOVA of these slopes), still common with researchers, is known to be the least powerful approach to detect differences in treatments (Wolfinger, 1996). Through these mixed-model methods, random effects (such as locations, blocks, and possibly genotypes), and their interactions with fixed effects (treatments) can be appropriately estimated and realistic inferences made.

Many population dynamic processes can be expressed only in nonlinear form (e.g., $y = f(t; a,b)$, where $f(\bullet)$ is a nonlinear function). The recent advances in nonlinear mixed models (Garrett et al., 2004) can be applied to these situations, but the range of experimental designs is much more limited (currently), and considerably larger data sets are required to estimate and compare parameters. Nevertheless, statistically savvy and motivated epidemiologists can make considerable progress here.

Mechanistic modelling (linked differential equations)

Van der Plank (1963) clearly realized that models such as the logistic were inadequate for a biologically meaningful characterization of disease progress in time. His approach was to use a so-called differential-delay equation in order to represent polycyclic disease development. This model relates dy/dt to the *infectious* disease intensity rather than to total disease intensity, with infectious disease estimated based on assumed fixed-duration latent and infectious periods. Although the use of differential-delay equations serve as a good foundation for developing computer simulation models with fixed time steps, such equations are extremely cumbersome for mathematical analysis, making it difficult to explore implications of different biological properties of hosts and pathogens, or of different control strategies, on long-term disease development. Eventually, plant pathologists discovered the mathematical elegance of linked or coupled differential equations for characterizing disease progress (Jeger, 1986a, b; van Maanen and

Xu, 2003). The approach – which was utilized as long ago as 1911 for representing malaria epidemics (Ross, 1911) – is to use two-to-several differential equations, with some variables of interest and parameters appearing in more than one of the equations. The beauty of this approach is that new terms can be easily added, as needed, to meet the objectives of the investigator and the details of the pathosystem, and asymptotic and steady state results (such as disease persistence) can be explored quantitatively. Furthermore, even though analytical solutions cannot generally be obtained (i.e., one cannot write out y as a function of parameters and time without the use of the integral symbol), numerical solutions are now easy to obtain with many mathematical programmes such as MATHCAD and MATHEMATICA.

Statistical software such as PROC MODEL of the SAS/EST system allows direct parameter estimation of one or more parameters for these types of models (Madden et al., 1987). The approach is iterative and computationally intensive, but readily accomplished by those who have a good understanding of nonlinear models and statistics. However, unlike the case for models with analytical solutions (linear or nonlinear; see previous sub-section above), one cannot easily incorporate the features of the experimental design (e.g., split plot, etc.) into the model fitting. Rather, one generally needs to estimate parameters for each individual epidemic (e.g., each field or plot) and then perform t-tests or analysis of variance on the estimated parameters (depending on the experimental design).

A relatively simple coupled differential equation model for a polycyclic disease with no plant mortality is given by:

$$\begin{aligned}\frac{dH}{dt} &= -\beta HI \\ \frac{dL}{dt} &= \beta HI - \omega L \\ \frac{dI}{dt} &= \omega L - \mu I \\ \frac{dR}{dt} &= \mu I\end{aligned} \qquad (1)$$

where H, L, I and R are the densities of disease-free (healthy), latently infected, infectious, and post-infectious (removed) individuals (e.g., plants, leaves, roots, or even sites on leaves), $1/\omega$ is the mean latent period, $1/\mu$ is the mean infectious period, and β is the per capita transmission rate (new diseased individuals per diseased individual per healthy individual per unit time). For fungal (or oomycetes) diseases, β is the product of spore production per time unit per infectious individual, the probability that a spore comes in contact with a healthy individual, and the probability that a spore in contact with a healthy host individual causes an infection. Total disease at any time is determined as $Y = L + I + R$, and disease intensity as a proportion is given by $y = Y/(H + L + I + R)$. If initial disease intensity is very low, then at $t = 0$, initial total host density is virtually the same as initial healthy host density, H_0. The product βH_0 is analogous to van der Plank's (1963) corrected basic infection rate (new diseased individuals per diseased individual per unit time).

A fundamental result with this model is that disease will increase (i.e., an epidemic will occur) only if $\beta H_0/\mu > 1$. The expression to the left of the inequality is known as the basic reproduction number, R_0 (Diekmann and Heesterbeek, 2000). This composite parameter also indicates the final intensity of disease (after a long time) and the initial exponential rate of increase (see Segarra et al., 2001, for details). An example realization of the model in equation 1 is shown in Figure 1 for the situation with $R_0 = 2.5$. Final disease is less than 100%, and is estimated by iteratively solving $y_\infty = 1\text{-}\exp(-R_0 y_\infty)$. Control strategies are developed or evaluated by finding combinations of β, ω, and μ that give $R_0 < 1$; specific control tactics (e.g., host resistance, protectant fungicide, curative fungicide) can then be directed at reducing β, etc.

An advantage of the equation 1 formulation is the easy expansion for other situations. For instance, a simple-interest disease component (infections from resident inoculum throughout the epidemic, rather than just at the start) can be incorporated by using the πxH term, where x is the density of inoculum and π is a simple-interest rate parameter. One can consider x to be constant or to change (typically, decline over time), so that $dx/dt = \mathrm{K}x$. When x does not change, then πx is equivalent to the monocyclic rate parameter (r_M) of the monomolecular model. The πxH term is subtracted from dH/dt and added to dL/dt in equation 1. A pure simple-interest epidemic results if $\beta = 0$; otherwise, a composite of polycyclic and

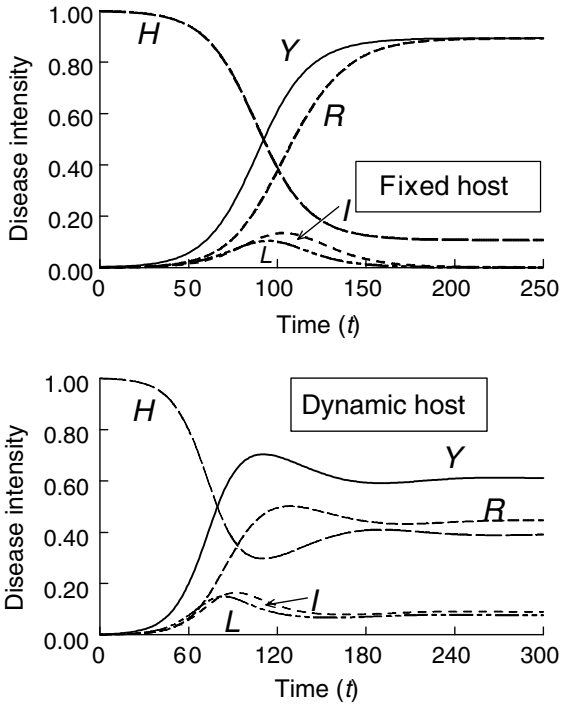


Figure 1. Density of healthy (H), latently infected (L), infectious (I), and post-infectious (R) individuals (on a proportion scale), together with total disease ($Y = L + I + R$), based on equation 1 (upper frame) and equation 2 (lower frame). Mean latent period ($1/\omega$) was 7, and mean infectious period ($1/\mu$) was 10 time units. Upper frame: $\beta H_0 = 0.25$ per time unit. Lower frame: $\beta H_0 = 0.35$ per time unit, $\eta = 0.02$, and $\pi = 0$ (no simple-interest component). Because of proportion scale, y and Y are the same here.

monocyclic processes occurs over time, very typical for root diseases (Gilligan, 2002). Host mortality can be incorporated by using a death-rate parameter η. Then ηH, ηL, ηI, and ηR are subtracted from the right hand sides of the equations for dH/dt, dL/dt, dI/dt, and dR/dt, respectively. Host growth can be incorporated in various ways. One approach is to consider just a single per capita growth rate (Ω) for disease-free individuals, and add the term Ω to the right hand side of the dH/dt equation. Suppose, further, that host size (e.g., number of citrus trees in a region) is fixed (say, at $H_{\max}$), and that new trees are only planted if others die. Then, the growth rate is also the mortality rate, and new host individuals can be expressed as $\Omega = \eta H_{\max}$; the combined growth/mortality for H can then be written as $\eta(H_{\max} - H)$.

A more general epidemic model can be written as

$$\begin{aligned}
\frac{dH}{dt} &= -\beta HI - \pi xH + \eta(H_{\max} - H) \\
\frac{dL}{dt} &= \beta HI + \pi xH - \omega L - \eta L \\
\frac{dI}{dt} &= \omega L - \mu I - \eta I \\
\frac{dR}{dt} &= \mu I - \eta R \\
\frac{dx}{dt} &= -\vartheta x
\end{aligned} \tag{2}$$

Note that in this example, total host size ($H + L + I + R$) does not change, even though there is continuous loss and addition of the host individuals (with a balance between the additions and losses). This can be seen by noting that $H_{\max} = H + L + I + R$ and adding the rates: $dH/dt + dL/dt + dI/dt + dR/dt = 0$. The model can be written in different ways to unlink the growth and mortality, to incorporate more complicated linkages, and to account for more than one disease or more than one host genotype at a time, but the example is useful to show one model formulation. When $\pi = 0$ (no simple interest component), an R_0 can be defined for many host-growth/mortality model situations. For instance, with $\pi = 0$ (no simple-interest component), $R_0 = [\beta H_{\max}/(\mu + \eta)] \cdot [\omega/(\omega + \eta)]$. An example realization of this model is shown in the lower frame of Figure 1. Note that Y ($= L + I + R$) and H oscillate a little before settling down to the steady states. The steady-state level of disease at a given R_0 is lower for the dynamic host than the fixed-host situation (equation 1); without the simple-interest component, the steady state Y is $1 - (1/R_0)$.

This approach of using a dynamic (but fixed total) host population size has been used in plant disease epidemiology (e.g., Madden et al., 2000), and even more so in medical epidemiology (Anderson and May, 1991) to determine whether or not an epidemic can occur (i.e., a disease invasion) as well as the persistence (or not) of disease long term. With primary infections occurring throughout the epidemic ($\pi > 0$), the concepts become a little more complicated, but there may still be a threshold (combination of parameters) that must be met for disease to persist (see review in Gilligan, 2002, and references cited therein).

Many other biological features can be incorporated in the model of equation 2. For instance,

since most plant viruses are transmitted by arthropod vectors, the rate of change in H and L does not directly depend on infectious plant individuals (I) but on infective vectors per plant (Z). Thus, the contact rate term, βHI in the first two equations of the model must be replaced by βHZ, where Z is the density of infective vectors per plant. Other components would be unchanged. There is also a need to add equations for the dynamics of the vector population, including virus-free and infective vectors. Details are given in Madden et al. (2000) and Jeger et al. (2004). Other expansions can incorporate disruptions caused by harvesting and/or planting for a multi-season time scale, as well as host responses to infection (e.g., Gilligan, 2002; Madden and van den Bosch, 2002).

The models shown so far are all deterministic. These can all be expressed in stochastic form, which is useful if one is specifically interested in heterogeneity of epidemics, small population sizes, or the epidemic outcome for individual plants or plant units. Gilligan (2002) and Gibson et al. (1999) provide more details. The mathematics definitely becomes more difficult with stochastic models.

Some spatial aspects of epidemics

There are two different threads to the characterization of the spatial component of plant disease epidemics. One thread deals with dispersal and resulting disease gradients, and the use of observed gradients to elucidate the form of the contact distribution (Campbell and Madden, 1990), the probability of a unit of inoculum at one location (ξ) coming in contact with a host individual at location s. This approach has been especially valuable for determining the rate of disease expansion from a focus, both within fields and higher spatial scales (e.g., continents) (van den Bosch et al., 1999). The contributions of van den Bosch and Zadoks (see Zadoks, 2001), Ferrandino (1993), and Aylor (1999) are especially noteworthy for aerial pathogens, and of Gilligan and colleagues (2002) for root diseases.

One of the advantages of the coupled differential equation approach of the previous section is that it can be directly expanded to account for disease at any location as well as any time. With two physical dimensions, it is now necessary to be explicit in notation about time t and location s. With two dimensions, we need to use partial derivatives rather than ordinary derivatives. Expanding equation 1, we can write the spatio-temporal model as:

$$\begin{aligned}
\frac{\partial H(t,s)}{\partial t} &= -\beta H(t,s) \int_{-\infty}^{\infty} I(t,\xi) D(s-\xi) d\xi \\
\frac{\partial L(t,s)}{\partial t} &= \beta H(t,s) \\
&\quad \times \int_{-\infty}^{\infty} I(t,\xi) D(s-\xi) d\xi - \omega L(t,s) \\
\frac{\partial I(t,s)}{\partial t} &= \omega L(t,s) - \mu I(t,s) \qquad (3) \\
\frac{\partial R(t,s)}{\partial t} &= \mu I(t,s)
\end{aligned}$$

where all parameters are as defined before, and $D(s-\xi)$ is the contact distribution, which is simply a scaled version of a disease gradient. Example contact distributions include the exponential, Pareto, Cauchy, and normal. Unlike with the simpler purely temporal model(s), the rate of decline in healthy host individuals at location s (and the rate of increase in latently infected host individuals at s) is explicitly based on the integration of the contributions of infectious individuals at all locations (all ξ values). The specific contribution at ξ to disease at s is the product of magnitude of infectious individuals at ξ multiplied by the probability that a unit of inoculum (say, spore) at ξ reaches location s (based on the contact distribution).

Both so-called wave-like and non-wave-like disease expansion is documented, where the velocity of disease expansion into new areas is constant or increases with time, and supported by the theory summarized in equation 3. The velocity of expansion (or the acceleration of expansion) is generally proportional to $\ln(R_0)$, so that there is no spread if $R_0 \leq 1$. The form of the contact distribution makes the difference in type of expansion. An example realization is shown in Figure 2 for non-wave-like expansion. The linkage of temporal population dynamics of disease and focus expansion rates is of fundamental importance because it shows (qualitatively and quantitatively) how reproduction (infection) and contact probabilities (dispersal) fully determine spatio-temporal

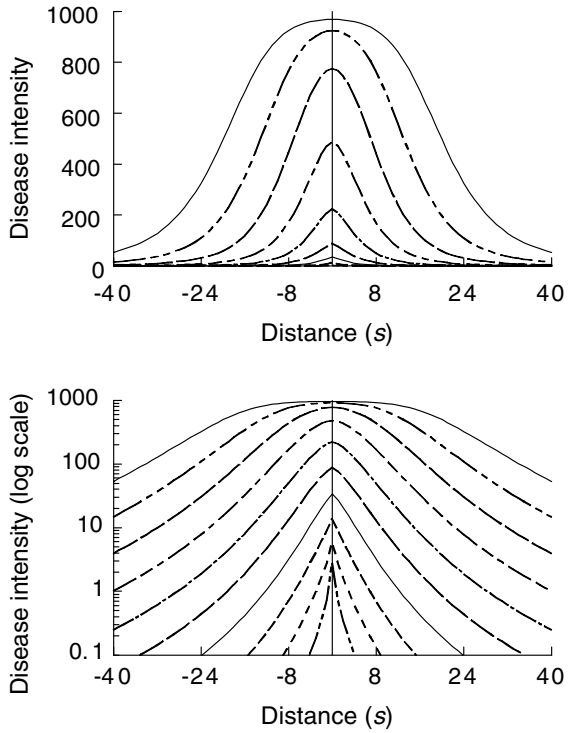


Figure 2. Density of diseased individuals ($Y = L + I + R$) vs. distance from a line source at 10-day time increments based on the numerical solution of equation 3. $H_0 = 1000$. Mean latent period ($1/\omega$) was 7, and mean infectious period ($1/\mu$) was 10 time units. $\beta H_0 = 0.4$ per time unit. A Pareto distribution was used for the contact distribution. The horizontal distance between pairs of successive curves at a single Y value (e.g., 0.1), divided by 10 gives the velocity of disease expansion.

outcomes, given a set of initial conditions. Control strategies are based, once again, on reducing R_0 to below 1, as well as reducing the scale of the contact distribution (spread parameter of the dispersal gradient) to a low value.

Equation 3 can be expanded for host growth, simple-interest dynamics, and so on, just as equation 1 was expanded to equation 2. It is (much) more difficult to work with partial differential equations than with ordinary ones, and finding numerical solutions can even be tedious. When the epidemic starts with a single focus (say, at the edge or centre of a region), then mathematical progress can be made, usually with additional assumptions (van den Bosch et al., 1990).

When there are several initial foci of infections, or unknown number and locations of initial inoculum, spatio-temporal differential-equation models, such as equation 3, are much less useful for studying epidemics because there is no single spatial starting point. With many original starting points (foci with disease at time 0), numerical solutions to equation 3 – or solutions to stochastic analogues of equation 3 (Xu and Ridout, 1998; van Maanen and Xu, 2003), – can be used to describe epidemics and explore implications of biological and physical features on disease progress, but it is more difficult to develop general principles or characterize expansion rates. Moreover, fitting a model such as equation 3 to data is generally impractical with standard statistical programmes. Thus, in epidemiology – as in ecology (Pielou, 1977) for that matter – more statistical (rather than mathematical) approaches have been generally followed to study spatial aspects of epidemics (Madden and Hughes, 1995, 2002; Hughes et al., 1997). This is the second thread of spatial characterization of epidemics. Concepts of clustering, aggregation, and regularity are utilized in terms of many different (but interrelated) statistical methods such as indices of dispersion, correlation, semivariograms, and distance statistics. This conceptual approach goes back to Cochran (1936) and Bald (1937) in plant pathology. A further advantage of the statistical approaches is that results (or concepts) are often directly useful for developing sampling plans, for either estimating disease intensity or making a decision regarding a control intervention (Madden and Hughes, 1999; Hughes et al., 2002).

The interrelationships between spatial aggregation of disease and temporal dynamics is gradually becoming more apparent. Using stochastic simulation, Xu and Ridout (1998) nicely showed how initial conditions, reproduction, and spatial contact distribution affect disease dynamics. A more theoretical approach has been to incorporate spatial properties of epidemics without explicitly using a spatial dimension (i.e., using models similar to equation 1). Models of this type are sometimes called spatially implicit, in contrast to the spatially explicit ones such as equation 3. The approach generally involves using a nonlinear function of I and/or H in the contact term, where the function depends on degree of aggregation (Zhang et al., 2000).

In recent years there has been considerable progress in bringing the two threads together (Gibson, 1997; Keeling et al., 2004), through the ingenious use of stochastic models and parameter

estimation. The results are primarily for the situation where individual plants are spatially referenced, and disease intensity is measured as a binary variable (diseased or healthy). There is still more work to do in this area, both in terms of testing the new approaches and for expanding approaches for other spatial situations (e.g., spatial referencing of just sampling units, not individual plants) and other measures of disease intensity (e.g., severity).

General thoughts on spatio-temporal disease dynamics

There is no doubt that through the expansion of models such as equations 1–3, as much detail as desired can be incorporated into models of plant disease epidemics (van Maanen and Xu, 2003). Such expansions require both knowledge of the pathosystem and knowledge of mathematics to realistically link model structure and parameterization to meaningful population-dynamic properties of the disease. Even with a model with just a few parameters, such as equation 3, mathematical insight may no longer be feasible unless starting conditions are restricted (e.g., one initial focus). Once other complicating factors are introduced, such as incorporation of environmental effects on the parameters (i.e., turning parameters into new variables that are functions of environment and new parameters), results will be limited to interpretation of numerical simulations with the model(s).

Although a model can be made indefinitely complex to represent an indefinitely complex biological system (such as plant pathosystem), such a model would violate the important principle of parsimony – keeping the model as simple as possible for the objectives of the investigator. Models, by definition, are simplifications of reality, which are useful for many purposes, including descriptions, comparisons, statistical inference, prediction, and developing understanding. Constructing models that are more complex than needed to meet the needs of the investigator – whether or not the basic biological knowledge is available for the construction of the model – is inefficient and can lead to faulty conclusions because of unrecognized (possibly erroneous) properties of the complicated model. The conclusions of Jeger (1986a) regarding the value of models with (relatively) small numbers of parameters and variables compared with large multi-variable and multi-parameter systems models is relevant here.

Thus, for many objectives, relatively simple models – such as the logistic, exponential, and monomolecular disease progress models, and empirical regression equations – will continue to be indispensable tools for the epidemiologist (Jeger, 2004). Although these models clearly are approximations, so are more complicated models. As stated by Bertrand Russell, 'Although this may seem a paradox, all exact science is dominated by the idea of approximation' (Auden and Kronenberger, 1966). Whether or not a model is too much of an approximation will always depend on the needs of the investigator.

Decision making in epidemiology

The case for disease forecasting

As stated by Gilligan (1985), 'Of the potential benefits of mathematical modelling to improving the efficiency of control of crop disease, prediction stands foremost.' Sometimes predictions or forecasts can be based explicitly on the rate parameter of a model such as the logistic or exponential (or even more complicated mechanistic model), as done with EPIPRE (Hardwick, 1998). That is, one can either use r_L to predict disease intensity some time period into the future based on either: (1) calculated r_L from previous estimates of disease in the current epidemic; or (2) predicted r_L based on environment (etc.), where the equation was developed in other studies. However, predictions need not necessarily be tied to population growth models in an explicit manner. A wide range of empirical models (often derived from regression or discriminant analysis) are utilized to simply identify conditions leading to a disease outbreak or a large reduction in yield (Madden and Ellis, 1988; Campbell and Madden, 1990). In fact, the prediction model (risk algorithm) may actually be derived without any formal statistical analysis; a good example of this is the collection of early prediction models for late blight of potato (Hardwick, 1998).

Epidemiologists continue to develop new prediction systems for plant diseases, usually used for scheduling fungicide applications, that is, for

decision-making in real time regarding the need for a control intervention (e.g., spray or not spray) (Madden and Ellis, 1988). A major development in this area over the last decade has been the application of formal Bayesian decision theory to either the construction or evaluation of the prediction systems. Growers and others (e.g., crop consultants) make numerous decisions before, during, and after growing seasons, such as when and where to plant, which cultivar to grow, whether to treat seeds with fungicide, and whether or not to apply a pesticide at any given time. Each decision can be correct or incorrect, and Bayesian decision theory provides a framework for making decisions objectively (e.g., spraying a crop) and for evaluating decisions that have been made.

The key contributions in plant pathology have been by J. Yuen, G. Hughes, and some of their colleagues (Yuen et al., 1996; Hughes et al., 1999; Yuen and Hughes, 2002; Yuen, 2003). A very recent and thorough example is Turechek and Wilcox (2005). The approach outlined below is explicitly used in the disease predictive system for Sclerotina stem rot of oilseed rape (Twengström et al., 1998), and qualitative aspects of the approach are used *implicitly* by most investigators developing and using predictive systems. I would argue, however, that a fuller utilization of the quantitative aspects of the approach will lead to better predictive systems and more efficient use of the ones already developed. The Bayesian-decision method centres on the determination of the probability of a disease outbreak (or need for a control intervention) before and after using the predictor. This approach has much in common with medical diagnostic research, where the prediction of a disease epidemic here is analogous to the diagnosis of an individual for a given disease condition. Both areas involve decisions (predictions of disease in a field or region, or the prediction of a disease status of an individual) that can only be made with some error. Plant disease prediction for crops has an additional level of uncertainty compared with medical diagnosis, since the decision is made for an entire population (e.g., a field of a given crop, or even a region where the crop is grown) rather than just for the individual.

Decision theory for disease prediction in plant pathology can be explained best with a detailed example. De Wolf et al. (2003) developed a model (their Model B) to predict major epidemics of *Fusarium* head blight of wheat in north America. The model, which is really a prediction rule in this scenario, was developed based on an analysis of 50 location-years for the disease in several parts of the U.S. The predictive system has evolved in several ways since the 2003 publication (L.V. Madden, unpublished), with many more observations analyzed as well as the development of new models, but I restrict the discussion here to the data and results of the published paper. Eighteen of the location-years were considered to be major epidemics (i.e., requiring control, if available), simply called epidemics for convenience. One can thus consider the so-called *prior* probability of an epidemic (E+) to be estimated or predicted as Prob(E+) = 18/50 = 0.36. Of course, the data set for analysis here is not necessarily representative of all locations for an indefinite period of time, but we use this calculated prior probability for now since other information was not available. Yuen (2003) discusses the use of location-dependent prior probabilities. With *Fusarium* head blight of wheat, there is a little more than a one in three chance overall that an epidemic will occur in a given location and year in the U.S. With no other information, such as measured weather variables or inoculum levels in the atmosphere or on crop debris, one would predict no epidemic – that is, one would bet against an epidemic at a given location and year, (even though one would sometimes lose the bet). This idea could be applied not just at a yearly time scale. For apple scab, one could determine the proportion of days (or weeks, for instance) where an infection period occurs in the spring. This can be considered the estimated prior probability of the need to apply a fungicide, independent of any other information (e.g., ignoring weather data).

Returning to *Fusarium* head blight, the probability of no epidemic (E−) in the De Wolf et al. data set is given by Prob(E−) = 32/50 = 0.64 = 1–0.36. For ease of calculations, it is convenient to determine the *odds* from the probability. In general, if A is some event, then odds(A) = Prob(A)/[1-Prob(A)]. If the odds are known, the probability is obtained from Prob(A) = odds(A)/[1 + odds(A)]. With the example, the odds are: odds(E+) = 0.36/[1–0.36] = 0.563, and odds(E−) = 0.64/[1–0.64] = 1.778. Note that the odds(A) = 1 when Prob(A) = 0.5. Probabilities above ½ give odds above 1, and probabilities less than ½ give odds below 1. The main symbols used in this part of the article are summarized in Table 1, for convenience.

Table 1. Some of the notation used regarding decision theory for disease prediction

Symbol	Description
Prob(E+)	Prior probability of an epidemic (or major epidemic, or for the need for a control intervention).
Prob(E−)	Prior probability of no epidemic (or for the lack of need for a control intervention)
odds(E+)	Prior odds of an epidemic: Prob(E+)/[1−Prob(E+)]
odds(E−)	Prior odds of no epidemic: Prob(E−)/[1−Prob(E−)]
Prob(P+|E+)	Probability of an actual epidemic being correctly predicted using some specified disease forecasting or predictive system; the conditional probability of a prediction of an epidemic (i.e., Z > threshold) given that an epidemic has occurred.
Prob(P−|E−)	Probability of an actual non-epidemic being correctly predicted using some specified disease forecasting or predictive system; the conditional probability of a prediction of a non-epidemic (i.e., Z < threshold) given that an epidemic has not occurred.
Prob(P+|E−)	Probability of an actual non-epidemic being incorrectly predicted to be an epidemic; [= 1−Prob(P−|E−)]; the conditional probability of a prediction of an epidemic given that an epidemic has not occurred.
Prob(P−|E+)	Probability of an actual epidemic being incorrectly predicted to be a non-epidemic [= 1−Prob(P+|E+)]; the conditional probability of a prediction of a non-epidemic given that an epidemic has occurred.
Z	Indicator of the risk of an epidemic (or the need for a control intervention). Can be derived with statistical or non-statistical methods.
TPP	True positive proportion (*sensitivity*); proportion of epidemics correctly predicted. An estimate of Prob(P+|E+)
TNP	True negative proportion (*specificity*); proportion of non-epidemics correctly predicted. An estimate of Prob(P−|E−)
FPP	False positive proportion; proportion of non-epidemics incorrectly predicted to be epidemics [= 1−TNP]. An estimate of Prob(P+|E−).
FNP	False negative proportion; proportion of epidemics incorrectly predicted to be non-epidemics [= 1−TPP]. An estimate of Prob(P−|E+)
J	A measure of accuracy [= TPP + TNP − 1 = TPP − FPP], known as Youden's index.
Prob(E+|P+)	Posterior probability of an epidemic given that one is predicted. Also known as the positive predictive value, PV(+).
Prob(E−|P−)	Posterior probability of no epidemic given that one is not predicted. Also known as negative predictive value, PV(−).
Prob(E-|P+)	Posterior probability of no epidemic given that one is predicted [= 1−Prob(E+|P+)].
Prob(E+|P−)	Posterior probability of an epidemic given that one is not predicted [1−Prob(E−|P−)].
LR(+)	Likelihood ratio of a positive prediction (i.e., prediction of an epidemic), TPP/FPP.
LR(−)	Likelihood ratio of a negative prediction (i.e., prediction of a non-epidemic), FNP/TNP.
ROC	Receiver operating characteristic curve, a plot of TPP vs. FPP. Can be written mathematically as TPP = *f*(FPP).
odds(E+|P+)	Posterior odds of an epidemic given that one is predicted (see equation 4).
odds(E−|P−)	Posterior odds of a non-epidemic given that one is not predicted (see equation 5).
CR	Cost ratio, approximately equal to the cost of a false positive (C_{FP}) divided by the cost of a false negative (C_{FN})
LR*(+)	Instantaneous likelihood ratio; the slope of the tangent to the ROC curve at any (FPP, TPP) point. Also given as first derivative *f*′(FPP), of the model for the ROC curve (see equation 8 for example).

The question that arises in the context of disease forecasting or predictive systems is: can one substantially change the predicted probability of an epidemic (or the odds) based on other information? De Wolf et al. (2003) used logistic regression to develop a risk algorithm (an equation in this case) for predicting an epidemic. The following predictor was obtained:

$$Z = -3.725 + 10.5(X_1 X_2)$$

where X_1 is number of hours that temperature is between 15 and 30 °C for the 7 days prior to wheat flowering, and X_2 is the number of hours that temperature is between 15 and 30 °C and relative humidity is at least 90% for the 10 days starting at flowering, and Z is the predicted logit of the probability of an epidemic given the two weather variables. For other pathosystems, Z could represent a direct measurement of, for instance, hours of wetness, rather than being a function derived from other variables. Z could also represent an estimate of disease intensity (e.g., measured disease early in a growing season) that is used to predict final disease or crop yield (see Turechek and Wilcox, 2005; Yuen and Hughes, 2002).

It turns out that the chosen threshold to use for predicting an epidemic with this data set is $Z = -0.40$ (recall that Z is a logit); for a given location-year, if Z is above the threshold, then predict an epidemic (and label this P+), otherwise predict no epidemic (and label this P−). Using this rule, 15 of the 18 known epidemics were correctly predicted, giving a true positive proportion of TPP = 15/18 = 0.833. Also, 27 out of the 32 known non-epidemics were correctly predicted, giving a true negative proportion of TNP = 27/32 = 0.844. One can also calculate the proportion of known non-epidemics predicted to be epidemics, which is the false positive proportion, FPP = 5/32 = 0.156. Finally, the proportion of known epidemics predicted to be non-epidemics is the false negative proportion, FNP = 3/18 = 0.167. It can be shown that FPP = 1-TNP, and that FNP = 1-TPP. All of these calculations are based on the known (or assumed) status of each observation in the data set. Overall accuracy could be reported as (15 + 27)/50 = 0.840. However, this metric depends on the TPP and TNP values as well as the fraction of observations in each category, and can thus be a misleading indicator of model (predictor) success if the fraction of epidemics is fairly far from ½. A better overall measure of accuracy is given by Youden's index, J = TPP + TNP − 1 = TPP − FPP; J equals 1 for a perfect predictor. For the example, $J = 0.677$.

The TPP is often called the *sensitivity* of a predictor or sensitivity of a model, and is an estimate of the probability of an actual epidemic being correctly predicted, Prob(P+|E+). Likewise, TNP is often called the *specificity* of a predictor (or of a model), and is an estimate of the probability that a non-epidemic is correctly predicted, Prob(P−|E−). FPP is the estimate of the probability that an actual non-epidemic is incorrectly predicted to be an epidemic, Prob(P+|E−); FNP is the estimate of the probability that an actual epidemic is incorrectly predicted to be an non-epidemic, Prob(P−|E+). The following table summarizes the metrics and the estimates for the example.

Predicted → Actual↓	P+	P-
E+	TPP Prob(P+|E+) 0.833	FNP (= 1−TPP) Prob(P−|E+) 0.167
E−	FPP (1−TNP) Prob(P+|E−) 0.156	TNP Prob(P−|E−) 0.844

Although TPP and TNP are very similar here, this is not necessarily the case.

The effectiveness of a predictor can be expressed in another way, which is extremely useful for some calculations below. The likelihood ratio of a positive prediction (i.e., prediction of an epidemic) is estimated by: LR(+) = TPP/(1-TNP) = TPP/FPP. Furthermore, the likelihood ratio of a negative prediction (i.e., prediction of a non-epidemic) is estimated by: LR(−) = (1-TPP)/TNP = FNP/TNP. For the example, one obtains LR(+) = 5.34 and LR(−) = 0.20. An accurate predictor has, in general, large LR(+) (above 1) and small LR(−) (close to 0).

The use of a threshold of −0.4 for Z gives an overall high accuracy (high J), but this is not the only possibility. This can be seen by the TPP and TNP values over the full range of possible Z

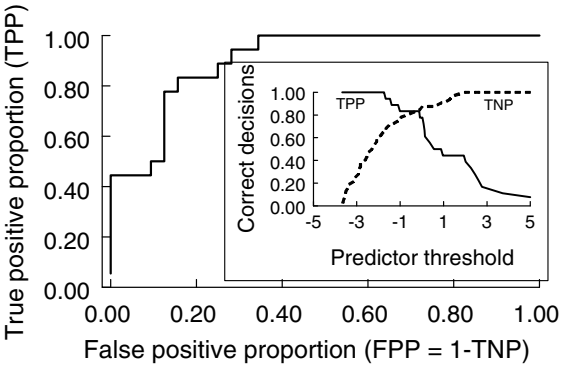


Figure 3. Main graph: An ROC curve, that is, the true positive proportion (TPP) vs. the false positive proportion (FPP), for the predictor model in De Wolf et al. (2003). Inset graph: TPP and true negative proportion (TNP = 1−FPP) vs. a full range of decision thresholds.

values, as shown by the insert of Figure 3. If one chose a low threshold (say, −3.5), then all the actual epidemics would be correctly predicted (TPP = 1), since, essentially, all observations are then predicted to be epidemics (all observations have Zs larger than −3.5). There is a major consequence of a low threshold, however, in that the actual non-epidemics are also predicted to be epidemics (TNP = 0 or FPP = 1). The J value would then be 0, a completely undesirable result. As the threshold increases above −3.5 for a predicted epidemic, TPP goes down, since it is now harder to predict an epidemic, and increasing numbers of the actual epidemics are predicted to be non-epidemics (i.e., some actual epidemics have Z values less than the threshold). However, TNP goes up with the increasing threshold, as higher numbers of non-epidemics are being correctly predicted (i.e., many of the actual non-epidemics have Z values below the threshold, as desired). Ultimately, with a very high threshold, all actual non-epidemics are correctly predicted (TNP = 1 or FPP = 0), since the threshold is higher than all the observations. However, this means that all the actual epidemics are also predicted to be non-epidemics (TPP = 0, FNP = 1), since it is then impossibly hard to predict an epidemic (i.e., all actual epidemics have Z values less than the high threshold). In between these extremes, there are thresholds around −1 to 0 where both TPP and TNP are high.

The overall performance of any predictor can be summarized with a receiver operating characteristic (ROC) curve (Metz, 1978; Linnet, 1988; Hughes et al., 1999), which is a plot of TPP vs. FPP (see Figure 3), that is, a plot of sensitivity vs. 1-specificity. The curve goes from (0,0) at the lower left corner to (1,1) at the upper right corner. The upper corner represents the *lowest* threshold tested (i.e., smallest Z value), corresponding to maximum sensitivity (high TPP) but minimum specificity (low TNP). The lower left corner represents the *highest* threshold tested (i.e., largest Z value), corresponding to minimum sensitivity and maximum specificity. If the predictor is of no value, the ROC curve will give a straight line through these two extremes, with a slope of 1. An ideal predictor will give a curve that goes vary rapidly from (0,0) up to a TPP value of 1 at an FPP barely above 0 (i.e., 0^+). The maximum J over all possible thresholds of Z for accuracy occurs at the point on the ROC curve that is closest to the upper left corner. The area under the ROC curve is an overall measure of the prediction accuracy, with a maximum of 1; for the example, the area is 0.9.

One can think of the ROC curve as representing the model TPP = f(FPP). It can be shown that the first derivative of this model [f'(FPP)] is the *instantaneous* value of LR(+) at any point FPP, that is, the tangent to the TPP:FPP curve at any FPP (Hughes and Madden, 2003). I call this likelihood ratio LR*(+). The more common calculation of LR(+), and the only one possible when the ROC curve is not available, is the straight-line slope over the interval from the point (0,0) to the point (FPP,TPP), which equals LR(+) = TPP/FPP (as indicated above).

Predictors in practice

All of the statistics shown so far deal with the success of the predictor for *known* epidemics and non-epidemics (i.e., for known status of the observations). To assess the predictor in practice, one must determine the probability that a random observation of unknown status (a particular location-year in the example) is an epidemic, given that the predictor score is positive (Z > the threshold), written as Prob(E+|P+), or the probability that a random observation of unknown status is not an epidemic, given that the predictor score is negative, written as Prob(E−|P−). Note that the conditional probabilities have been turned around from that used in developing the predictor, where the epidemic status was known; now, the prediction

status is known and not the actual status of an observation. To determine these and related probabilities for the population of interest (i.e., all location-years where the predictor is being used), one invokes Bayes' Theorem (Yuen and Hughes, 2002; Hughes and Madden, 2003), which can be most easily written (and interpreted) in terms of odds rather than probabilities.

The odds of an epidemic, given one is predicted [odds(E+|P+)] depends on the accuracy of the predictor, expressed as LR(+), and the prior odds that an observation is an epidemic [odds(E+)]. This new odds value can be written as:

$$\text{odds}(E+|P+) = LR(+)(\text{odds}(E+) \quad (4)$$

which is a simple direct product of the two terms. The left-hand side of this equation is known as the posterior odds of an epidemic (or a disease outbreak, or the need for a control intervention, etc.), given that one is predicted. The posterior odds will be high (relatively speaking) if the prior odds is high (e.g., a relatively high proportion of location-years for *Fusarium* head blight are epidemics) or if LR(+) is high (i.e., high accuracy). For the example, the posterior odds is given by: odds(E+|P+) = 5.34·0.563 = 3.0. The posterior probability can be determined by the transformation of the odds (see above), which produces Prob(E+|P+) = 3.0/(1 + 3.0) = 0.75. The posterior probability can be calculated directly, without use of prior and posterior odds (Yuen and Hughes, 2002), but the formula is cumbersome and less intuitive.

In typical usage, we would only predict an epidemic if the posterior odds is above 1 [or Prob(E+|P+) > ½]. In the example, the predicted odds of an epidemic occurring when the model predicts one is more than five times the average (or overall) odds of an epidemic when the predictor-variable information is not known (or not used). The predicted posterior probability of an epidemic when one is predicted (0.75) is a little more than double the prior probability (0.36) when no information is known (or used). Note that a predictor is only valuable if LR(+) is larger than 1 in this simple situation. When LR(+) = 1, the prior and posterior odds are the same, as well as the prior and posterior probabilities, which means that using the predictor is not giving the user any additional information about the chance of an epidemic – one is no more certain of the epidemic status of a random observation when the predictor is used compared to when it is not used.

One can also determine the posterior probability that an observation is not an epidemic when an epidemic is actually predicted, Prob(E−|P+), by first calculating the posterior odds: odds(E−|P+) = odds(E−)/LR(+). Note that Prob(E−|P+) is also given by 1-Prob(E+|P+). For the example, Prob(E−|P+) = 0.25. Thus, there is still a reasonable probability (¼) that an observation is not an epidemic even when one is predicted, with the given accuracy of the model. Another valuable term is the posterior probability that an observation is not an epidemic, given that a non-epidemic is predicted, Prob(E−|P−). This can be readily determined from:

$$\text{odds}(E-|P-) = \text{odds}(E-)/LR(-) \quad (5)$$

With the example, the posterior odds is estimated by: odds(E−|P−) = 1.778/0.20 = 8.98. The posterior probability is thus: Prob(E−|P−) = 8.98/(1 + 8.98) = 0.90. In other words, the probability that there will not be an epidemic when an epidemic is not expected (= 0.9) is increased compared to the prior probability of a non-epidemic (0.64) when no other information is available. Finally, the posterior probability of an epidemic given that a non-epidemic is predicted, Prob(E+|P−), can be determined from 1-Prob(E−|P−), or by first calculating the posterior odds as: odds(E+|P−) = odds(E+)LR(−). For the example, one obtains Prob(E+|P−) = 0.10, meaning that there is only a small probability that a random unknown observation is actually an epidemic when a non-epidemic is predicted. Because of the properties of the predictor model and the prior probability of an epidemic, in the example one would have somewhat more confidence in predictions of non-epidemics than in predictions of epidemics.

The predictor is clearly of value based on the achieved likelihood ratios and the prior odds (or prior probabilities). The point of interest here is how the success of the predictor depends on the prior odds of an epidemic (which comes directly from the prevalence of epidemics). For instance, consider what would happen if epidemics were much less common than assumed originally, but one still wanted to use the developed predictor.

I call this scenario B (and the original scenario above as A). Assume the prior probability of an epidemic is Prob(E+) = 0.05, which gives: odds(E+) = 0.053, Prob(E−) = 0.95, odds(E−) = 19.0. For Fusarium head blight this is unrealistically low, but the value is used for demonstration purposes. The likelihood ratio is unchanged since this is a property of the predictor, not the prior probability. The posterior odds of an epidemic then is calculated from equation 4 as odds(E+|P+) = 5.34·0.053 = 0.28, resulting in a posterior probability of an epidemic when one is predicted of Prob(E+|P+) = 0.28/(1 + 0.28) = 0.22 (Table 2). The posterior probability of an observation not being an epidemic when one is predicted is very high, equal to Prob(E−|P+) = 1–0.22 = 0.78. The other posterior probabilities are: Prob(E−|P−) = 0.99, and Prob(E+|P−) = 0.01 (i.e., there is a very low probability that a random observation is an epidemic when one is not predicted). Using environmental data in the form of the predictor (Z), the probability that a random observation (a location-year) is an epidemic clearly increases when one is predicted (from the prior probability of 0.05 to the posterior probability of 0.22). However, there is still considerably less than a 50% chance that an observation is an epidemic when one is predicted, and there is nearly an 80% chance (posterior probability of 0.78) that an observation is a non-epidemic when one is predicted. In other words, use of the current predictor (with the selected threshold of Z for a positive prediction) would be of little value in disease management when epidemics are rare – most control interventions would be wasted since most of the predicted epidemics would turn out to be non-epidemics. This shows in general that disease forecasting may not be of direct value for rare diseases, unless one has a predictor with an extremely high overall accuracy (very large LR(+)).

With an imperfect predictor (i.e., TPP < 1, TNP < 1), there is uncertainty in any predictions of epidemics. Given that epidemics do not occur that often (hypothetically, when Prob(E+) = 0.05), the evidence must be stronger than that obtainable from the use of the predictor to conclude (at least with more than a 0.50 probability, or more than an odds of 1) that an epidemic will occur when predicted. However, if LR(+) was 20 (i.e., a much more accurate predictor), then the posterior odds would be 1.06 (when prior odds of an epidemic was 0.053), and the posterior probability would be 0.51. Under these circumstances, the use of the predictor would be of greater value in management. However, finding such accurate predictors in plant pathology may be very difficult.

There is an alternative to improving the overall prediction accuracy for rare diseases. One can use a different threshold of Z for an epidemic, which can be demonstrated with the *Fusarium* head blight results. As shown in Figure 3, TPP declines, and TNP increases, as the threshold increases. If one used a threshold (on the logit scale) of +2 (instead of −0.4), one would obtain TPP = 0.39, TNP = 0.99, FPP = 1–0.99 = 0.01, and FNP = 1–0.39 = 0.61; the likelihood ratios would be LR(+) = 39.0 and LR(−) = 0.62. I call this scenario C (see Table 2). Using a higher threshold means moving down the ROC curve (Figure 3) towards the lower left corner. By using a high threshold, almost all the known non-epidemics are correctly predicted (more specifically, almost all the *known* non-epidemics have Z values less than the new higher threshold; TNP ≈ 1), but only 40% of the *known* epidemics are correctly predicted

Table 2. Evaluation of disease predictor for *Fusarium* head blight of wheat (see De Wolf et al., 2003) under various scenarios of prior probability of an epidemic and threshold used for predicting an epidemic[a]

Scenario	Prob(E+)	Threshold	TPP	TNP	LR(+)	LR(−)	Prob(E+|P+)	Prob(E−|P−)
A[b]	0.36	−0.4	0.833	0.844	5.34	0.20	0.75	0.90
B	0.05	−0.4	0.833	0.844	5.34	0.20	0.22	0.99
C	0.05	+2.0	0.39	0.99	39.0	0.62	0.67	0.97
D	0.85	−0.4	0.833	0.844	5.34	0.20	0.97	0.47
E	0.85	−1.7	0.944	0.656	2.74	0.085	0.94	0.67

[a]See text and Table 1 for explanation of symbols and notation, as well as for terms not given in table.
[b]Scenario A is the nominal (or standard) use of the predictor as described in the article.

(TPP $\approx$ 0.4) because many of these cases have Z values below the new threshold. In other words, it is now more difficult to correctly predict a known epidemic. Although this may seem to be undesirable, the low prior probability of an epidemic (0.05) means that the posterior probability of an epidemic when one is predicted to occur is actually improved (i.e., predicted epidemics are more likely to actually be epidemics). Using the numbers in this paragraph (including a prior probability of an epidemic of 0.05), Prob(E+|P+) = 0.67 (substantially higher than the 0.22 posterior probability for scenario B), which means that in only about one third of the time, on average, would a predicted epidemic actually correspond to a non-epidemic [i.e., Prob(E−|P+) = 1–0.67 = 0.33]. When one requires stronger evidence for an epidemic [a higher threshold, giving a larger LR(+)], there is less of a chance that the prediction of an epidemic is wrong. There is a slight cost here to using the higher threshold – the posterior probability of an observation being a non-epidemic when predicted to not be an epidemic is reduced to Prob(E−|P−) = 0.97, compared with 0.99 with the nominal predictor threshold (scenario B). This is due to the increase in LR(−) compared to the original choice of threshold. Here, very little was lost in identifying non-epidemics by changing the threshold for a positive prediction (since it is, relatively speaking, easy to predict non-epidemics when epidemics are rare).

One can also consider the implication of much more common occurrence of epidemics. For instance, if the prior probability of an epidemic is Prob(E+) = 0.85 (much higher than realistic for Fusarium head blight), one obtains: odds(E+) = 5.67, Prob(E−) = 0.15, odds(E−) = 0.176. Consider the predictor used at the nominal threshold (−0.4), which gives, once again, LR(+) = 5.34 and LR(−) = 0.2. I call this scenario D (Table 2). One obtains the following posterior probabilities: Prob(E+|P+) = 0.97, Prob(E−|P+) = 1–0.97 = 0.03, Prob(E−|P−) = 0.47, and Prob(E+|P−) = 0.53. Here, the predictor works very well for predicting epidemics (there is only a 3% chance that an observation is a non-epidemic when one is predicted), but works less well for predicting the non-epidemics. Based on Prob(E+|P−), about half the observations predicted to be non-epidemics are, on average, actually epidemics. In the absence of a more accurate predictor model [that would give a combined higher LR(+) and lower LR(−)], one could move the threshold to a lower value (see Figure 3), which corresponds to a higher TPP and lower TNP (the opposite direction than used when epidemics were rare). A lower threshold means moving up the ROC curve towards the upper right corner.

With a threshold of −1.7, one obtains: TPP = 0.944, TNP = 0.656, FPP = 1–0.656 = 0.344, and FNP = 1–0.944 = 0.056; the likelihood ratios would be LR(+) = 2.74 and LR(−) = 0.085. This is scenario E (Table 2). By making it easier to predict known epidemics (i.e., lowering the threshold of Z for deciding in favour of an epidemic) when epidemics are common, one does not change the predictions of epidemics very much; that is, because of high prevalence of epidemics, Prob(E+|P+) = 0.94, only slightly less than under scenario D]. However, it is much more likely that an observation predicted to be a non-epidemic is, in fact, a non-epidemic. That is, Prob(E−|P−) = 0.67, compared to 0.47 for scenario D. But there is still a fairly high probability that a random observation is an epidemic when a non-epidemic is predicted [Prob(E+|P−) = 0.33].

The above evaluation was totally presented in terms of commonness of epidemics (or of the need to control, in general), measured by estimated prior odds, and the accuracy of the predictor for known cases, measured by likelihood ratios. The entire evaluation can be recast in terms of costs for each of the four possible decisions (true positive, true negative, false positive, and false negative), or more simply, the costs of the two incorrect decisions (false positives and false negatives) (Linnet, 1988). Hughes and Madden (2003) give a detailed account of this for regulatory problems (invasive organism risk analysis) rather than for disease forecasting. In brief, if C_{FP} and C_{FN} are the costs of a false positive and a false negative prediction, respectively, then define CR as the ratio of these: CR $\approx C_{FP}/C_{FN}$ (see Table 1). CR actually depends also on the costs of true positives and true negatives, but relative to the costs of the errors, it is quite practical to consider these other costs as nil. Then, the decision rule that minimizes the average cost of using the predictor can be shown to be:

odds(E + |P+) > CR, then predict an epidemic;

odds(E + |P+) < CR, then predict a non-epidemic.

The posterior odds is calculated from equation 4, based on the prior odds and accuracy of the predictor (the likelihood ratio). The lower the CR, the lower the posterior odds needed to predict an epidemic. A low CR would occur when the costs of false positives (such as the cost of spraying a crop where the fungicide application is not needed) is relatively low compared to the costs of false negatives (such as the yield loss due to the disease that occurs because a needed fungicide spray was not used). As shown by Hughes and Madden (2003), the optimum threshold of Z to use for minimizing costs of using a predictor can be determined based on CR. In particular, one can write:

$$\mathrm{LR}^{*}(+)_{\mathrm{opt}} = \mathrm{CR}/\mathrm{odds}(\mathrm{E}+) \tag{6}$$

where the left-hand side is the optimum LR*(+) in which to make epidemic predictions (a function of TPP and FPP, which are properties of the predictor). Equation 6 is easy to calculate. The right hand side is based on commonness of epidemics and the relative costs of predictor errors (simply as a ratio, so that absolute values of the costs are not needed), but does not involve the accuracy of the predictor. It should be re-emphasized that the predicted LR*(+) here is the 'instantaneous' change in TPP with change in FPP (slope of the tangent to the TPP:FPP curve at FPP), given as f' (FPP). To translate equation 6 into an exact predicted combination of TPP and TNP (or FPP = 1-TNP), one must first have a specific model for the ROC, TPP = f(FPP), in order to obtain f' (FPP) at the optimum point (Hughes and Madden, 2003). An example is equation 4 in Hughes and Madden (2003). The fit of this model to the ROC curve in Figure 3 using nonlinear least squares results in the following equation:

$$\mathrm{TPP} = \left(1 + e^{-4.42}\left(\mathrm{FPP}^{-2.37} - 1\right)\right)^{-1/2.37} \tag{7}$$

As shown in Figure 4, the model provides a good fit to the TPP values. The first derivative of equation 7 is given as:

$$f'(\mathrm{FPP}) = \frac{\left(1 + e^{-4.42}\left(\mathrm{FPP}^{-2.37} - 1\right)\right)^{-1/2.37} e^{-4.42}\mathrm{FPP}^{-2.37}}{\mathrm{FPP}\left(1 + e^{-4.42}\left(\mathrm{FPP}^{-2.37} - 1\right)\right)} \tag{8}$$

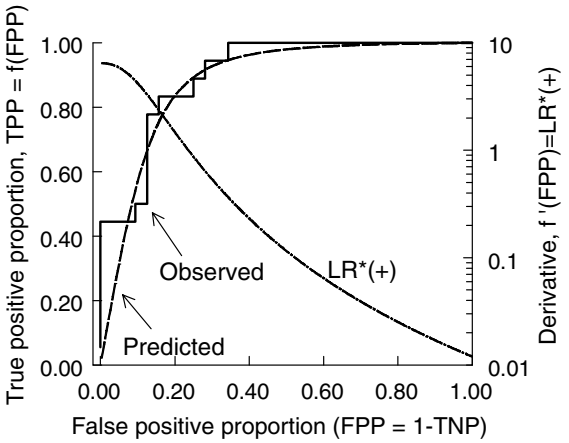


Figure 4. Left-hand axis: An ROC curve (true positive proportion, TPP, vs. the false positive proportion, FPP) for the predictor model in De Wolf et al. (2003), together with predicted TPP from equation. Right-hand axis: derivative of equations 7 and 8 vs. FPP, the instantaneous likelihood ratio [LR*(+)].

which is based on equation 5 in Hughes and Madden (2003). As required because of the shape of the ROC curve, f' (FPP) = LR*(+) declines with increasing FPP (Figure 4; right-hand axis). Mathematically, one can solve the f' (FPP) equation for FPP, and then obtain TPP based on f(FPP) (equation 7). From this combination of TPP and FPP (which gives TNP and FNP), one can determine the standard LR(+) (= TPP/FPP) and use equation 4 for risk assessment.

The use of equations 6 and 7 can be demonstrated with the *Fusarium* head blight data. Previously, an epidemic was predicted in practice if odds(E+|P+) was greater than 1. This is equivalent to specifying that the cost ratio, CR, equals 1, where both types of errors are equally costly. Using odds(E+) = 0.563, as stated previously, one finds from equation 6 that $\mathrm{LR}^{*}(+)_{\mathrm{opt}} = 1.78$ when CR = 1. Graphically, one finds this value of f' (FPP) in Figure 4, and then determines the corresponding FPP and TPP at this value. One can see that FPP $\approx$ 0.18 and TPP $\approx$ 0.81 at f' (FPP) = 1.78 in the graph, which are similar to the

values used in the nominal situation described in Table 1 (with a Z threshold of −0.4; scenario A). (There will be some discrepancy because equation 7 is not a perfect fit to the ROC curve.) With the TPP and FPP values here, LR(+) ≈ 4.5, and the posterior odds of an epidemic when one is predicted is 4.5·0.563 = 2.5, giving Prob(E+|P+) = 0.71 (close to the value found at slightly different TPP and TNP values in Table 1). It is important to emphasize that the posterior odds of an epidemic (or non-epidemic) are actually calculated with LR(+), *not* with the instantaneous rate LR*(+).

Now consider the situation in which a false negative decision is four times as costly as a false positive decision ($CR = C_{FP}/C_{FN} = 1/4 = 0.25$). With the nominal prior odds of an epidemic, one finds that $LR^*(+)_{opt} = 0.25/0.563 = 0.444$. From Figure 4, one can see that this derivative occurs at FPP ≈ 0.32 (or TNP ≈ 0.68) and TPP ≈ 0.94. An increased TPP and decreased TNP compared to the nominal situation (with CR = 1) is higher up the ROC curve (towards the right-hand corner), which means that an epidemic is predicted to occur at a lower Z threshold (Figure 3). That is, there is a less stringent criterion to predict an epidemic. Using the listed sensitivities and specificities here, one obtains LR(+) = 0.94/0.32 = 2.94, and LR(−) = (1–0.94)/0.68 = 0.088. The posterior odds of an epidemic when one is predicted then is odds(E+|P+) = 2.94·0.563 = 1.66, giving a posterior probability of Prob(E+|P+) = 1.66/(1 + 1.66) = 0.62. It can be shown that the posterior probability of a non-epidemic when a non-epidemic is predicted is Prob(E−|P−) = 0.95. One is more certain about the true epidemic status of a random observation when non-epidemics are predicted compared to when epidemics are predicted. Also, Prob(E+|P+) is lower here than when CR = 1, but this reduction in certainty of epidemics is required to minimize the average cost of making predictions.

In general, as demonstrated in the previous paragraph, as CR declines at a given prior probability of an epidemic, one moves up the ROC curve towards the right-hand corner (higher TPP and FPP; lower TNP and FNP), which means that a lower Z threshold is used for predictions of epidemics. Loosely speaking, with high cost of false negative decisions, one would not want to make too many of these errors (i.e., one would want a low FNP). Conversely, as CR increases (e.g., false positives are more costly than false negatives), one moves down the ROC curve towards the left-hand corner (lower TPP and FPP; higher TNP and FNP), which means that a higher Z threshold is used for predictions of epidemics. Loosely speaking, with high cost of false positives, one would not want to make too many of these errors (i.e., one would want a low FPP). The approach outlined here can also be coupled with consideration of different prior probabilities of epidemics, as presented previously in this section. It is quite possible for a given pathosystem that there are combinations of prior probabilities and costs of false predictions that one would always assume that an epidemic will occur or always assume that an epidemic will not occur. The decision-theory approach provides the formal mechanism for evaluating these scenarios.

The analyses discussed here are just the beginnings of the possibilities for applying risk assessment to disease prediction (Yuen and Hughes, 2002; Yuen, 2003). In fact, only the initial aspects of this approach have been formally applied to *Fusarium* head blight forecasting at this stage; costs of decisions and consideration of prior probabilities of epidemics for this pathosystem will be addressed more formally after a more accurate prediction system is developed for known epidemics and non-epidemics. Other areas requiring research for plant diseases in general include: having more than a dichotomy of decisions (such as spray, do not spray, and wait-and-see what happens); dealing with more than a dichotomy of predictions (such as predicting the degree of expected damage from a disease, predicting spraying once or weekly for the rest of the season); dealing with more than a dichotomy of measured outcomes (such as intensity of disease on a continuum for different predictions, for either binary predictions or a continuum of predictions); and dealing with multiple diseases or pests.

Discussion

Botanical epidemiology has advanced on many fronts in the years since van der Plank's first book was published in 1963, and the discipline continues to be a foundation for understanding and predicting diseases at the population scale. I have

chosen to outline just two out of many possible broad topics where substantial advances have been, and continue to be, made. Many of the speakers and poster presentations at the 9th International Workshop on Plant Disease Epidemiology reported in this special issue of the *European Journal of Plant Pathology* have dealt with other valuable topics.

The use of growth-curve and mechanistic population dynamic models, especially the coupled (ordinary and partial) differential equations outlined in this article, provide a flexible and powerful methodology for representing the temporal, spatial, and spatio-temporal dynamics of diseases, and provides the framework to elucidate general thresholds for epidemic occurrences (disease invasion), long-term persistence of disease, velocity at which disease expands from foci, and the initial rates of disease increase over time. Many of these qualitative and quantitative properties of epidemics can be summarized by the basic reproduction number (R_0).

The coupled differential equations (or other model formulations, such as stochastic difference equations) can be made extremely complicated, and care should be taken to keep the principle of model parsimony in mind when modelling epidemics! Strategies for disease control can be readily explored by finding the combination of disease properties (e.g., latent and infectious periods, transmission rate) that result in a R_0 less than 1. This approach is less useful for real-time prediction of epidemics, or for determining the need to intervene with a control measure, because precise estimates of parameters under specific environmental conditions are often not available. Usually, simpler prediction equations (such as regression equations, discriminant functions, or *ad hoc* rules) are used to actually make predictions. This use of more descriptive equations can be further justified by the fact that even very complicated mechanistic models often result in simple exponential-type population increase when disease intensity is not high (see Segarra et al., 2001).

The incorporation of probabilistic decision theory into disease predictions (whether these come from empirical rules or equations, or even population-dynamic models) is the second topic I covered where key advances have been made recently (Yuen and Hughes, 2002). Although most plant pathologists, including epidemiologists, clearly are not yet thinking formally or explicitly in terms of prior and posterior odds, and likelihood ratios, the concepts follow directly from intuitive understandings of how prediction rules are applied when there is some inaccuracy in the predictors and when epidemics and non-epidemics (or the need to intervene or not with a control method) are not equally common in a given area. The next generation of advances in this area will deal more with the costs of decisions (correct or incorrect), and in addressing some of the biological and environmental interactions in more complicated pathosystems, possibly involving multiple diseases (and crops simultaneously) (McRoberts et al., 2003).

Whether one is working with elaborate population-dynamic models or testing prediction rules for decision making, I accept as an axiom that appropriate statistical methods be used to fit models to data, compare results within and between studies, and test hypotheses. Developments in linear and nonlinear mixed models, for instance, are drastically improving the matching of the statistical methods to the data and experimental design, and intended inferences of the investigator (Schabenberger and Pierce, 2002). There is also growing evidence that Bayesian methods are also useful for developing prediction equations (Mila and Carriquiry, 2004), and not just in evaluating the performance of predictors (as demonstrated in this article). Except for the most quantitative of the botanical epidemiologists, more effort is still needed to encourage plant pathologists to keep abreast of developments in statistics and utilize the appropriate old and new techniques (Garrett et al., 2004).

Acknowledgements

Salaries and research support were provided by state and federal funds to the Ohio Agricultural Research and Development Center, Ohio State University, USA.

References

Anderson RM and May RM (1991) Infectious Diseases of Humans; Dynamics and Control, Oxford Science Publication, Oxford.

Auden WH and Kronenberger L (1966) The Viking Book of Aphorisms, Viking, New York.

Aylor DE (1999) Biophysical scaling and the passive dispersal of fungus spores: relationship to integrated pest management strategies. Agricultural and Forest Meteorology 97: 275–292.

Bald JG (1937) Investigations on 'spotted wilt' of tomatoes III: infection in field plots. Bulletin 106. Melbourne, Australia: Council for Scientific and Industrial Research.

Campbell CL and Madden LV (1990) Introduction to Plant Disease Epidemiology, John Wiley & Sons, New York

Cochran WG (1936) The statistical analysis of field counts of diseased plants. Supplement to Journal of Royal Statistical Society 3: 49–67.

DeWolf ED, Madden LV and Lipps PE (2003) Risk assessment models for wheat Fusarium head blight epidemics based on within-season weather data. Phytopathology 93: 428–435

Diekmann O and Heesterbeek JAP (2000) Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation, John Wiley and Son, Ltd., Chichester, U.K, 303

Ferrandino FJ (1993) epidemic waves: I. Focus expansion within a linear planting. Phytopathology 83: 795–802

Fry WE (1982) Principles of Plant Disease Management, Academic Press, New York

Garrett KA, Madden LV, Hughes G and Pfender WF (2004) New applications of statistical tools in plant pathology. Phytopathology 94: 999–1003

Gibson GJ (1997) Investigating mechanisms of spatiotemporal epidemic spread using stochastic models. Phytopathology 87: 139–146

Gibson GJ, Gilligan CA and Kleczkowski A. (1999) Predicting variability in biological control of a plant-pathogen system using stochastic models. Proceedings Royal Society London B 266: 1743–1753

Gildow F, Damsteegt B, Stone A, Schneider W, Luster D and Levy L (2004) Plum pox in North America: identification of aphid vectors and a potential role for fruit in virus spread. Phytopathology 94: 868–874

Gilligan CA (1985). Introduction. In: Gilligan CA) Advances in Plant Pathology, Vol. 3; Mathematical Modelling of Crop Disease (pp 1–10) Academic Press, London

Gilligan CA (2002) An epidemiological framework for disease management. Advances in Botanical Research 38: 1–64

Gottwald TR, Hughes G, Graham JH, Sun X and Riley T (2001) The citrus canker epidemic in Florida: the scientific basis of regulatory eradication policy for an invasive species. Phytopathology 91: 30–34

Hardwick NV (1998). Disease forecasting. In: Jones DG (eds) The Epidemiology of Plant Diseases (pp 207–230) Kluwer Academic Publishers, Dordrecht

Hughes G and Madden LV (2003) Evaluating predictive models with application in regulatory policy for invasive weeds. Agricultural Systems 76: 755–774

Hughes G, McRoberts N, Madden LV and Nelson SC (1997) Validating mathematical models of plant disease progress in space and time. IMA Journal of Mathematics Applied in Medicine and Biology 14: 85–112

Hughes G, McRoberts N and Burnett FJ (1999) Decision-making and diagnosis in disease management. Plant Pathology 48: 147–153

Hughes G, Gottwald TR and Yamamura K (2002) Survey methods for assessment of citrus tristeza virus incidence in urban citrus populations. Plant Disease 86: 367–372

James C (2003) Global Status of Commercialized Transgenic Crops: 2003, International Service for the Acquisition of Agri-Biotech Applications (ISAAA), Manila

Jeger MJ (1986a). The potential of analytic compared with simulation approaches to modeling in plant disease epidemiology. In: Leonard KJ and Fry WE (eds) Plant Disease Epidemiology, Vol. 1: Population Dynamics and Management (pp 255–281) Macmillan Publishing Company, NY

Jeger MJ (1986b) Asymptotic behaviour and threshold criteria in model plant disease epidemics. Plant Pathology 35: 355–361

Jeger MJ (2004) Analysis of disease progress as a basis for evaluating disease management practices. Annual Review of Phytopathology 42: 61–82

Jeger MJ, Holt J, vanden Bosch F and Madden LV (2004) Epidemiology of insect-transmitted plant viruses: modeling disease dynamics and control interventions. Physiological Entomology 29: 1–14

Jones DG(.) (1998) The Epidemiology of Plant Diseases, Kluwer Academic Publishers, Dordrecht, 460

Keeling MJ, Brooks SP and Gilligan CA (2004) Using conservation of pattern to estimate spatial parameters from a single snapshot. Proceedings National Academy of Science USA 101: 9155–9160

Kranz J(.) (1990) Epidemics of Plant Diseases: Mathematical Analysis and Modeling, 2nd edn. Springer-Verlag, Berlin

Legg JP (1999) Emergence, spread and strategies for controlling the pandemic of cassava mosaic virus disease in east and central Africa. Crop Protection 18: 627–637

Linnet K (1988) A review of the methodology for assessing diagnostic tests. Clinical Chemistry 34: 1379–1386

Madden LV (1986). Statistical analysis and comparison of disease progress curves. Pages 55–84. In: Leonard K and Fry WE (eds) Plant Disease Epidemiology (pp 372) Macmillan Publishing Co, New York

Madden LV and Hughes G (1995) Plant disease incidence: distributions, heterogeneity, and temporal analysis. Annual Review of Phytopathology 33: 529–564

Madden LV and Hughes G (1999) Sampling for plant disease incidence. Phytopathology 89: 1088–1103

Madden LV and Hughes G (2002). Plant epidemics, models, and analysis. In: El-Shaarawi A and Piegorsch W (eds) Encyclopedia of Environmetrics (pp 1557–15623) John Wiley & Sons

Madden LV and Ellis MA (1988). How to develop disease forecasters. In: Kranz J and Rotem J (eds) Experimental Techniques in Plant Disease Epidemiology (pp 191–208) Springer-Verlag Publishers

Madden LV and vanden Bosch F (2002) A population-dynamics approach to assess the threat of plant pathogens as biological weapons against annual crops. BioScience 52: 65–74

Madden LV and Wheelis M (2003) The threat of plant pathogens as weapons against U.S. crops. Annual Review of Phytopathology 41: 155–176

Madden LV, Pirone TP and Raccah B (1987) Temporal analysis of two viruses increasing in the same tobacco fields. Phytopathology 77: 974–980

Madden LV, Jeger MJ and vanden Bosch F (2000) A theoretical assessment of the effects of vector-virus transmission mechanism on plant virus disease epidemics. Phytopathology 90: 576–594

Maloy OC (1993) Plant Disease Control, John Wiley & Sons, New York

McRoberts N, Hughes G and Savary S (2003) Integrated approaches to understanding and control of diseases and pests in field crops. Australasian Plant Pathology 32: 167–180

Metz CE (1978) Basic principles of ROC analysis. Seminars in Nuclear Medicine 8: 283–298

Mila AL and Carriquiry AL (2004) Bayesan analysis in plant pathology. Phytopathology 94: 1027–1030

Neter J, Wasserman W and Kutner MH (1983) Applied Linear Regression Models, Richard, D. Irwin, Inc, Homewood, IL, 547

NRC (2002). Predicting Invasions of Non-indigenous Plants and Plant Pests. National Research Council, National Academy of Sciences, National Academy Press, Washington, DC

Pielou EC (1977) Mathematical Ecology, John Wiley & Sons, New York

Rizzo DM, Garbelotto M, Davidson JM, Slaughter GW and Koike ST (2002) *Phytophthora ramorum* as the cause of extensive mortality of *Quercus* spp. and *Lithocarpus densiflorus* in California. Plant Disease 86: 205–214

Ross R (1911) The Prevention of Malaria, 2nd edn. Murray, London

Rush CM, Stein JM, Bowden RL, Riemenschneider R, Boratynski T and Royer MH (2005) Status of Karnal bunt of wheat in the United States 1996 to 2004. Plant Disease 89: 212–223

Schabenberger O and Pierce FJ (2002) Contemporary Statistical Models for the Plant and Soil Sciences, CRC Press, Boca Raton, Florida

Segarra J, Jeger MJ and vanden Bosch F (2001) Epidemic dynamics and patterns of plant diseases. Phytopathology 91: 1001–1010

Strange RN and Scott PR (2005) Plant disease: a threat to global food security. Annual Review of Phytopathology 43: (in press)

Turechek WW and Wilcox WW (2005) Evaluating predictors of apple scab with receiver operating characteristic (ROC) curve analysis. Phytopathology 95: 679–691

Twengström E, Sigvald R, Svensson C and Yuen J (1998) Forecasting sclerotinia stem rot in spring sown oilseed rape. Crop Protection 17: 405–411

vanden Bosch F, Metz JAJ and Diekmann O (1990) The velocity of population expansion. Journal of Mathematical Biology 28: 529–565

vanden Bosch F, Metz JAJ and Zadoks JC (1999) Pandemics of focal plant disease, a model. Phytopathology 89: 495–505

vanMaanen A and Xu X-M (2003) Modelling plant disease epidemics. European Journal of Plant Pathology 109: 669–682

vander Plank JE (1963) Plant Diseases: Epidemics and Control, Academic Press, NY, 349

Waggoner PE (1960). Forecasting epidemics. In: Horsfall JG and Dimond AE (eds) Plant Pathology, AN Advanced Treatise (pp 291–313III) Academic Press, New York

Wolfinger RD (1996) Heterogeneous variance–covariance structures for repeated measures. Journal of Agricultural, Biological, and Environmental Statistics 1: 205–230

Xu X-M and Ridout MS (1998) Effects of initial epidemic conditions, sporulation rate, and spore dispersal gradient on the spatio-temporal dynamics of plant disease epidemics. Phytopathology 88: 1000–1012

Yuen J (2003) Bayesian approaches to plant disease forecasting. Online. Plant Health Progress doi:10.1094/PHP-2003-1113-06-RV. [http://www.plantmanagementnetwork.org/php]

Yuen JE and Hughes G (2002) Bayesian analysis of plant disease prediction. Plant Pathology 51: 407–412

Yuen JE, Twengström E and Sigvald R (1996) Calibration and verification of risk algorithms using logistic regression. European Journal of Plant Pathology 102: 847–854

Zadoks JC (2001) Plant disease epidemiology in the twentieth century: a picture by means of selected controversies. Plant Disease 85: 808–816

Zadoks JC and Schein RD (1979) Epidemiology and Plant Disease Management, Oxford University Press, Oxford, 427

Zhang X-S, Holt J and Colvin J (2000) A general model of plant-virus disease infection incorporating vector aggregation. Plant Pathology 49: 435–444

European Journal of Plant Pathology (2006) 115:25–34
DOI 10.1007/s10658-005-4513-5

Framework development in plant disease risk assessment and its application

X.B. Yang
Department of Plant Pathology, Iowa State University, Ames, Iowa, 50011, USA (E-mail: xbyang@iastate.edu)

Accepted 20 October 2005

Abstract

This article reviews recent developments in plant disease risk assessment. The role of risk assessment as an application area in macrophytopathology and its contribution to the development of macroscale disease study are discussed. This article also discusses the concepts and components of risk assessment for different end points and the assessment framework of different potential ranges of a new pathogen: establishment range, suitability range, damage range, and dispersal range. Different end points generate risk information suitable for decision makers at different levels. New insights gained from selected major diseases, especially from risk assessment due to the recent global movement of soybean rust, are presented. The role of pathologists in presenting risk information has extended beyond the professional research domain and has become critical in influencing decision-making, evident by soybean rust in both South and North America. The bias components of risk communication are defined, and different levels of receivers for risk information are identified based on their interpretation capability of risk information, bias potential, and utilization of risk information. Lack of predictability of dispersal potential contributes to uncertainty of risk assessment for airborne diseases. Potential research areas in disease risk assessment are discussed.

Introduction

Risk assessment is part of botanical epidemiology. Some principles of disease risk assessment can be found in disease epidemic forecasting in the early development stage of epidemiology. In the 1970s, threat analysis was used in regulatory plant pathology to determine quarantine subjects. Now, different terms such as threat analysis, risk analysis, or risk assessment are used for studies to determine the epidemic potential of exotic, new, and emerging diseases. Over the last ten years, advances in computer computation have made quantitative analysis with large sets of climatic data possible, along with risk assessment of exotic diseases (Yang et al., 1991). The most systematic study is on soybean rust, *Phykoposora pachyrhizi*. Since the introduction of Asian soybean rust into the New World, tremendous resources and effort have gone toward assessing the risk of soybean rust. This article reviews recent developments in risk assessment of exotic and emerging new diseases with special reference to soybean rust.

Definition of risk assessment

In a narrow sense, risk assessment involves determining the potential epidemiological and economic impact of emerging or new diseases (domestic or foreign). The information is critical to decision makers at higher levels, for example, in deciding policies to deal with risk mitigation and risk preparation at regional or national levels. In a sense, risk assessment is the application of epidemiology to regulatory plant pathology or to disease management. A study of risk assessment is a macroscale, long-term disease prediction that encompasses assessment of establishment potential, entry potential (when the range of a disease expands beyond a political border), and epidemic potential (epidemic frequency and epidemic severity or extent of the disease), and the potential losses in a region or country once an epidemic occurs. Risk assessment is an epidemiological study to predict future occurrence of a disease by using non-experimental approaches, often involving computer modelling. The modelling

Table 1. Comparison of purpose and scales of risk information among different users of risk information for Sclerotinia stem rot of soybean, caused by *Sclerotinia sclerotiorum*, in the North Central region of the United States (from Yang, 2004)

User	Purpose	Temporal scale	Spatial scale
Farmers	Chemical control	In-season	Fields/farm
	Variety selection	Coming season	Fields/farm
	Tillage	Coming season	Fields/farm
Extension		In-season	Fields/area
Agronomists	Advice	Coming season	Fields/area
Seed/chemical	Marketing strategies	Next year	Regional
Companies	Breeding decisions	Next few years	Regional
	Product development	Years/decade(s)	Regional
Government	Funding decision	Years	Regional

results are not repeatable with extrapolation from limited field studies or published data.

In a broad sense, risk prediction or risk assessment has been extended to predicting the seasonal occurrence of endemic diseases, especially in horticultural and high-value crops (Luo et al., 2001). Research in this area involves predicting the upcoming season's disease occurrence in a defined production area. Use of the term 'disease risk' for disease forecasting reflects advances in disease risk communication at the farm-level. Yang (Yang, 2003) recently outlined a conceptual example of risk prediction according to a temporal and spatial scale that uses risk information (Table 1). In this article, I exclude the seasonal prediction of the outbreak risk of an endemic disease in a region for disease management practices.

Need for risk assessment

The rapid development of risk assessment reflects the adaptation of plant pathology to new global agriculture trends, the consequence of which leads to an increased risk of new diseases. The first trend is frequent seed and plant material movement by international companies. For example, some soybean and corn seeds planted in the US Midwest are now produced in South America. The movement of large amounts of germplasm is thus unavoidable and could facilitate the movement of plant pathogens and consequently the introduction of new diseases. Assessments are made for regulatory decision-making, and several studies are underway in this area. The second trend is climate change (IPCC, 2001). Climate change has been shown to be a driving force in the long-term dynamics of plant diseases (Yang and Scherm, 1997); new diseases and re-emerged diseases have been attributed to climate change (Rosenzweig et al., 2001). Such changes can result in the emergence of new threats from minor diseases or to range expansion of a disease to production regions where the disease previously was not a concern. Expansion of the damage range of soybean bean pod mottle virus into the US North Central Region and sudden death syndrome (SDS) into the northern United States are recent examples. The increase of SDS prevalence in US north central region has been associated with increased planting of soybean in early spring, a production measure associated with warmer springs. The third trend is the change in farming practices, exemplified by the expansion of no-tillage systems.

As an applied research area, disease risk assessment has recently gained importance in the political arena of biosecurity (Madden and Van den Bosch, 2002;Madden and Wheelis, 2003). Because biosecurity-related risk assessment is new and is highly relevant to political issues, risk assessment in this area will likely embed bias potential, and contributions of research from this area to risk assessment are needed.

Risk assessment and macrophytopathology

Macrophytopathology is the study of disease occurrence patterns and disease management at the macroscale (Zeng, 2003). A narrower definition is the study of statistical patterns of disease distribution, disease range, and epidemic frequencies on large spatial and temporal scales. Ecologically, macrophytopathology addresses the questions when, where, and why a new disease emerges and becomes a major production threat. Such a study can be used to predict the damage

potential of a disease (Yang, 2003). A theoretical framework of macrophytopathology has not yet been developed. The concept of risk assessment predates macrophytopathology. Risk assessment is the application of epidemiological methodology to predict the long-term risk of new or emerging diseases. With such information, disease risk can be mitigated by managing the movement and distribution of a new disease on a macroscale (Magarey, personal communication). Regulatory measures are effective approaches for risk mitigation. Strategically, a sound assessment helps make decisions in developing resistance programmes, such as screening for resistance germplasm. Development of resistant varieties is an expensive, long-term investment and would not be initiated until the entry of the disease.

It has been suggested that macrophytopathology is basically the same as geophytopathology as initially proposed by Weltzien (Weltzien, 1972). Conceptually, macrophytopathology is different from geophytopathology in two respects. In his review, Weltzien used the idea of geophytopathology in which 'documentation, analysis, and prognosis of plant epidemics seem to be an appropriate theme for maps as basic contributions.' For geophytopathology, the study began in advanced stages of epidemiology and uses quantitative epidemiological approaches to study disease occurrence patterns on a large scale. At the time geophytopathology was proposed, botanical epidemiology as a discipline or field was in its cinfancy. In macrophytopathology, new information technology is integrated with sophisticated modelling techniques to handle vast climatological data. It also includes disease management on a macroscale. The core area of macroscale study, risk assessment, is the application of theories and methods of epidemiology. In the context of macrophytopathology, risk assessment is equivalent to disease forecasting in the conventional scale of plant disease study. Weltzien (1972) did foresee the potential of geophytopathology for disease management, but it has yet to be demonstrated in macrophytopathology.

Disease range concepts and risk assessment

The concept of disease range was proposed by Yang and Feng (Yang and Feng, 2001) to describe the two-dimensional distribution of the occurrence of a disease over a geographic area. When an exotic disease is introduced into a new geographic region or a new disease emerges, assessment of the potential range of the disease is to determine the geopathosystem range, which is directly associated with impact assessment. Epidemiologically, a disease range is related to the following four other ranges important to the ultimate impact of a disease.

1) *Establishment range*, or year-round survival range, in which a pathogen can sustain itself from one growing season to the next by completing disease cycles. For a soil-borne disease, the establishment range is the same as the disease range. For an airborne disease, the establishment range is smaller than the disease range, and the establishment region serves as the source area of inoculum for other regions during a season. For example, with wheat rust in the United States, the establishment range is the overwintering range of wheat rust fungi in southern Texas and Louisiana.
2) *Suitability range* defines a geographic area where the conditions are suitable for disease to occur, which is important for airborne disease risk assessment. A range suitable for a disease to occur does not necessarily mean the disease will occur in all areas defined by the range because of the uncertainty of inoculum availability. The airborne inoculum may never reach certain areas defined in the region. Assessment based on this range represents the maximum risk.
3) *Dispersal range* defines the geographic area into which airborne diseases can spread seasonally from overwintering areas defined by the establishment range. Dispersal range of a disease does not mean range of inoculum spread, which is far larger than seasonal disease dispersal range. Recent studies show that air currents can, within a season, carry the spores of soybean rust from Brazil to almost anywhere in the Western Hemisphere.
4) *Damage range* is a region where the disease can cause significant economic yield loss in a frequency that warrants implementation of production measures. Sometimes, it can be referred to as the endemic region of a disease. Geographically, the physical sizes of

these ranges for an airborne disease have the following order: establishment range < damage range < dispersal range < suitability range. For soil-borne diseases, the establishment range should be equal to the suitability range and no smaller or larger than the damage range (establishment range = suitability range > damage range).

Components in risk assessment

The fundamental parameters of potential risk of a new or exotic disease to crop production in a geographical region are disease range, frequency of potential epidemics, and intensity of epidemics in terms of economic losses. Epidemiologically, risk assessment for a new disease consists of the following exercises, each of which can be independent. Negative results of each assessment indicate the non-threatening status of a new disease. These components are establishment potential, suitability of the environment to disease occurrence, dispersal potential, and yield-loss potential. In risk assessment, almost all assessment starts with a suitability assessment, and the order of assessment procedures is determined by the availability of techniques and data at the time when the assessment is made. Because effects of temperature and dew are known key factors in determining the infectivity of a plant pathogen, these factors are investigated first in epidemiological studies and therefore are available for the environment suitability assessment of risk assessment. For soybean rust, there are three critical components of uncertainty: 1) suitability of climatic conditions for rust epidemics in soybean production areas, 2) likelihood of establishment of the fungus in North America, and 3) the seasonal dispersal potential of the pathogen from overwintering regions to major soybean production regions.

Establishment assessment

This assessment addresses the question: once an exotic pathogen is introduced, can it survive from season to season in a country or geographic region, and if so, where? If the disease cannot be established in the region or in an area of reachable distance during the growing season, the pathogen should not be considered a threat. For this assessment, the key aspects are to determine the availability of alternate hosts of the pathogen and its overwintering potential in a non-host growing season. Quantification of the source strength early in spring may be important to mid-term season disease forecasting for airborne diseases.

Suitability assessment

This assessment determines the suitability of the climate in the studied geographic region or country. Suitability assessment is almost always the first assessment to address in risk analysis. If the climate is unsuitable for the occurrence of the disease, no further assessment is needed. The normal approach for this assessment is to use a disease model together with climatic data to assess the epidemic potential in the region. Sometimes, the epidemic potential is further fed into a yield-loss model to determine the maximum yield loss potential of a disease. When long-term climatic data are available, determination of potential epidemic frequencies and severity, a higher level of assessment, are useful to policy makers. If severe epidemics of a disease have a frequency one-in-eight or ten years, the disease may not be a major production concern. However, suitability assessment is almost always made with an assumption that initial inoculum is available early in a growing season. Therefore, the estimated risk would be greater than the real loss if the disease is airborne because initial inoculum of an airborne disease is not always available. Risk estimated from such an assessment represents maximum risk.

Dispersal potential assessment

This assessment applies to a disease caused by an airborne pathogen that establishes regionally in a country but poses a threat to the rest of the production area. The damage level of the disease depends on the yearly reintroduction potential of inoculum. It is not a concern for soilborne diseases, however, once the establishment assessment is completed. For soybean rust, the pathogen overwinters in southern Florida and Texas, which is far from the major US soybean production region. For risk assessment, this is the last component to study because epidemiology does not

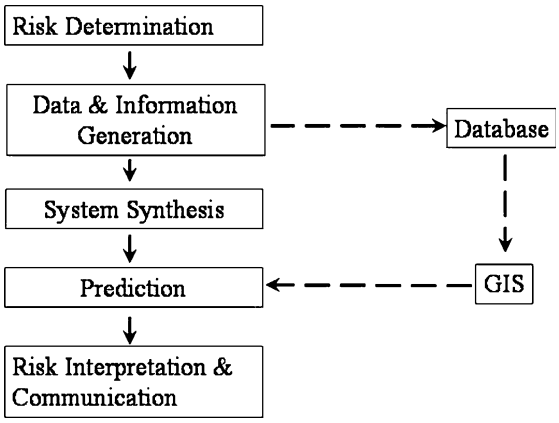


Figure 1. Flow chart showing the process of a risk assessment study. To have a completed risk study, the last step of risk interpretation and communication by the modellers is essential.

provide methodology for such an assessment. For soybean rust, this uncertainty has been a factor in decision-making in Argentina where the disease can overwinter only in the northern production regions. Figure 1

Conceptually, risk assessment for exotic, new, or emerging diseases could be generalized according to the types of assessment (Figure 2). The outcome of each assessment and its usefulness depend on the type of assessment. Information from the establishment assessment is significant for quarantine purposes. The risk from the suitability assessment is the maximum risk useful for policy decision-making in which accountability is a concern. The maximum risk could be far from the real losses because of a lack of inoculum in a season or missing a dispersal component in the study. For epidemiologists, the challenging part is to determine the most likely losses, which is information useful for industry, whose concern is on the impact of a new disease on its profitability. The arrow in Figure. 2 indicates the most likely losses assessed with comprehensive epidemiological information. Depending on the availability of information/data and skills provided to an assessment, the assessed risk may be higher or lower than the most likely risk. Over time, the assessed value from a risk assessment should approach the most likely risk.

Examples of assessment for a soilborne disease

Risk assessment of soybean sudden death syndrome (SDS) caused by *Fusarium solani* f. sp. *glycines* is an example of an emerging disease. This soilborne disease was first reported in Arkansas in the early 1970s and caused endemic production

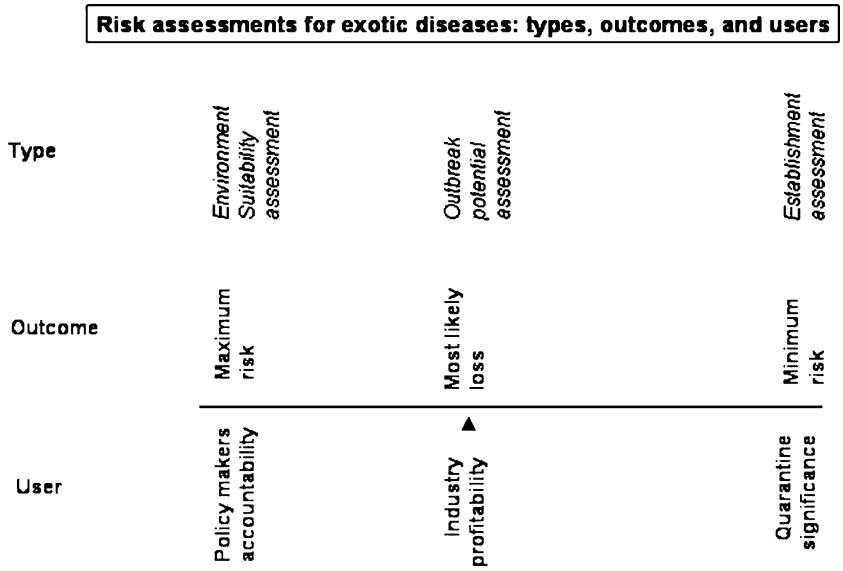


Figure 2. Conceptual framework of risk assessment for exotic, new, or emerging diseases, showing the types of assessment, outcome of each type, and users of each outcome. The arrow indicates the most likely losses assessed with comprehensive epidemiological information.

problems in the southern states (Roy et al., 1997). It was initially domestic and no establishment assessment was needed. Because it is soilborne, the dispersal assessment was also not applicable. The need for a potential impact assessment was not realized until 1993 when the disease was found in Iowa, a leading soybean production state. The soybean industry needed to know the level of the threat from this disease to the North Central Region, which produces 78% of US soybean, so that the funding agencies could prioritize investment of research funds. A risk assessment for SDS was conducted using Climex, computer software developed by CSIRO (Sutherst and Maywald, 1991), with disease parameters generated from experiments conducted under controlled conditions (Scherm and Yang, 1999). The assessment predicted that the disease would cause more losses in the north central region than in the southern United States where the disease originated. The assessment has proved correct. By 2002, SDS had spread into Canada and Minnesota and was ranked the number one damaging fungal disease in the north central US soybean production. When data were presented in 1995 at a regional soybean conference, SDS immediately gained the attention of the soybean industry. Breeding for resistance to SDS had started before the disease became a production problem in the US North Central Region. Now, many seed companies have resistant varieties available to growers. Without this risk assessment, which promoted resistance breeding and management research, current disease prevalence levels and the frequency of epidemics would probably be higher.

Example of assessment for an airborne disease

Suitability assessment

Risk assessment of soybean rust, caused by the fungus *Phakopsora pachyrhizi*, to US soybean production is the best example for an airborne disease. The assessment started in the early 1980s and was one of the earliest risk assessment programmes. Risk assessment of this disease has contributed to the development of general concepts and quantitative methodology. For suitability analysis, research efforts were divided into two phases: 1) understanding infection components based on research in a containment facility at Fort Detrick in Federick, MD, and in fields in Asia, which provided baseline information for disease modelling; and 2) development of computer modelling to quantitatively assess the potential effects of rust on soybean yield in the United States.

Research under controlled conditions focused on determining the importance of each epidemic component or subcomponent in the soybean rust disease cycle and on quantifying its response to host and environmental variation. The components—spore germination, infection, latent period, lesion expansion, sporulation, and senescence of uredia—were studied by several researchers at Ft. Detrick and elsewhere (Keogh, 1974; Marchetti et al., 1976; Bromfield et al., 1980; Meching et al., 1989); Patil et al., 1997; Hundekar and Hiremath, 2001). The effects of dew duration and temperature on infection have been quantified as a two-dimensional relationship from which an infection model was developed to estimate infection (Marchetti et al., 1976). These studies provided critical background information for building epidemiological models for risk assessment. The field experiments were conducted at the Asian Vegetable Research and Development Centre in southern Taiwan. Soybean can grow year-round in this area, and disease occurs most of the year, except during the winter. These data allowed analysis of the seasonal variation in rust epidemics.

From the data compiled from field and greenhouse studies, a computer simulation model, SOYRUST was developed. This simple disease model includes most weather variables that influence disease epidemics. The model was validated with data from Taiwan, and predictions matched observations. SOYRUST was integrated, as a subroutine, into the soybean growth model SOYGRO (Wilkerson et al., 1985), developed at the University of Florida to simulate disease progress during the production season and to predict yield. With the assumption that spores are available early in the growing season, the simulation results showed that considerable yield loss could occur in some areas of the United States. In recent years, USDA–APHIS generated a US risk map by using continuous moisture measurements and dew days data. The general consensus is that the environmental conditions in the US soybean production regions are suitable for the occurrence of this disease (Bromfield, 1984).

Establishment assessment

Predicting the year-round survival of the soybean rust fungus is important for determining availability of spores in the spring and for determining potential dissemination into major soybean production regions during a growing season from an overwintering area. Models were used to predict where climatological conditions are suitable for the year-round persistence of *P. pachyrhizi* worldwide. Long-term meteorological averages were used to assess stress by using the CLIMEX software developed by Sutherst and Maywald(1985). Integration of stresses was used to predict the likelihood of survival of *P. pachyrhizi* within a location (Pivonia and Yang, 2004). The assessment shows that areas presumed suitable for year-round survival of *P. pachyrhizi* in the Western Hemisphere extend from southern Brazil to southern Texas and Florida. In the United States, the fungus is likely to overwinter in areas where climatic conditions in winter are similar to those in southern China. During mild winters, the coastal region of the Gulf of Mexico is also in the *P. pachyrhizi* year-round survival zone (Figure. 3). The reported occurrence of soybean rust in kudzu plants near Tampa, Florida, in February of 2005 validated the assessment.

Dispersal assessment

Several means existed for long distance spread of airborne diseases, such as the tobacco blue model to assess spore movement from the Caribbean to the southeastern US. For soybean rust, there are several means for northward movement. Climatological models have been integrated with epidemiological models for prediction of soybean rust

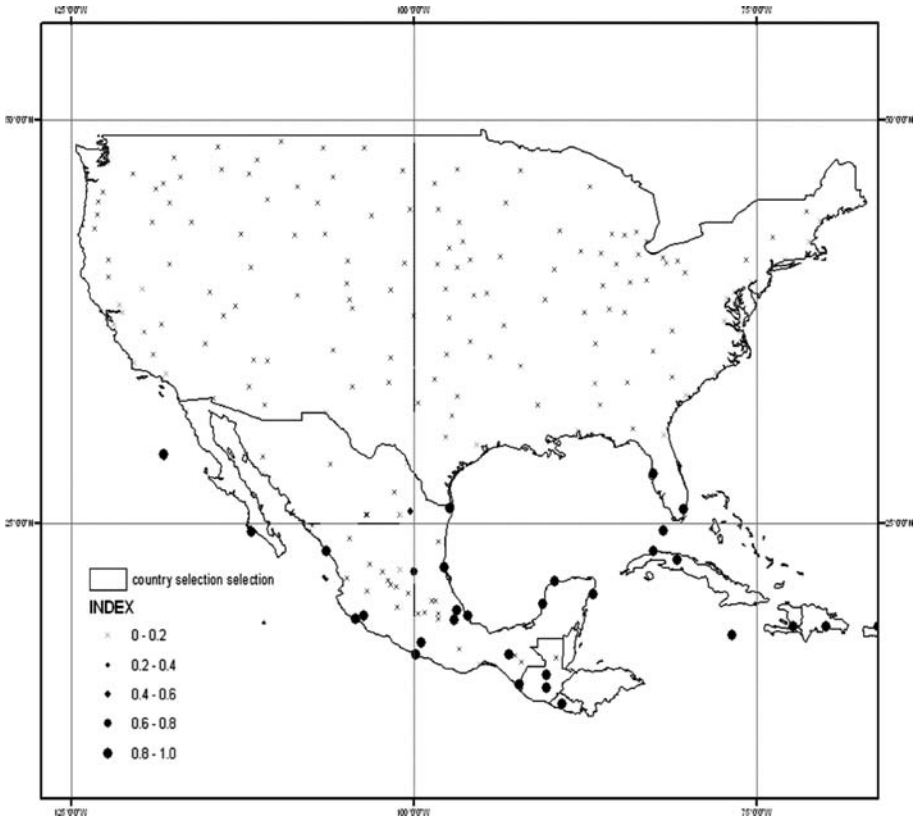


Figure 3. Illustration of the potential range of soybean rust, *P. pachyrhizi*, in North America with a focus on the United States (Yang, 2003). The establishment range is indicated by the probability of overwintering, which is the US coastal area. The suitability range is estimated to extend to from Gulf Coast to northern border of the US in east of Rocky Mountain. The dispersal range and the damage range are yet to determined.

movement. Such integration could be a potential research area to generate new directions for epidemiological research. The MM 5 model is a global circulation model for air particle movement and high split model for rust prediction. The model has been used to correctly predict the 2004 season rust spore movement to Argentina and Colombia. In August 2004, it also predicted the spore movement from Colombia to the southern US before the disease was found in Louisiana.

Risk communication

To plant pathologists, risk communication is relatively new in the framework of risk study compared with plant disease risk analysis. Elements of risk communication consist of receivers of risk information, interpretation of the uncertainty, and delivery of the information with an approach or method according to information receivers and the capability to digest risk information of receiver aspects. To the risk communicators who are often plant pathologists, it is crucial to understand the epidemiological principals used in the assessments. Most models are built with certain assumptions that are critical to risk interpretation. Without stating the underlining assumption while discussing the risk, the risk level can be overestimated, which often occurs when risk information is disseminated by non-pathologists. For soybean rust, when citations are made for the study of USDA economic assessment or yield loss (Yang et al., 1991), the writers often reported figures of potential losses from a risk assessment without providing the assumptions addressed by authors in their studies.

Receivers of risk information

After the risk assessment is made, the information is disseminated to the public. There are several levels of receivers: policy makers, scientists in chemical or seed companies for product development, managers of funding agencies, producers for day-to-day farm operation, and crop advisers in private and public domains. Based on the decision levels, the risk information is generated differently in terms of temporal and spatial scales (Table 1) (Yang, 2003). For decision-making, the temporal scale ranges from decades to weeks, and the spatial scale ranges from the entire country to individual farms. The capability of digesting risk information varies, and the purpose of taking risk information differs. Therefore, interpretations of the risk on the receiver sides are different. Receivers of risk information can be grouped by three levels based on their knowledge and interpretation of disease risk.

Level 1. This level consists of pathologists or groups of professionals who have in-depth knowledge of plant pathology. Detailed outcomes on risk assumption can be explained and the uncertainty of the information is fully understood by receivers. This level includes industry experts in chemical and seed companies whose annual profitability relies on decision with a measured risk. The setting for risk information delivery includes professional meetings through presentations by authorities on specific topics, and this level has no bias in interpretation of disease risk.

Level 2. This includes decision makers at the policy-making level at the public domain, such as government officials or society leaders who have limited understanding of assumptions or who overlook the assumptions made for assessment for political liability and address the maximum risk. Bias interpretations of the risk assessment are not uncommon. To make a better argument for making policies, the maximum risk is used most frequently for accountability reasons. The example is the frequent use of soybean rust economic assessment of US $ 7.2 billion (Kuchler et al., 1984) in decision-making, although a later much smaller figure (~US $ 1–2 billion) has been made in a new risk assessment.

For high-level policy makers, decision-making is based on political rational and accountability, which has a tendency towards self-protection and therefore naturally embeds bias of select use of risk information. For soybean rust, decisions at higher levels were likely made using the worse-case scenario. Economic calculation is less essential compared with the producers or profit-driven businesses. Finally, risk communication varies from culture to culture. Some cultures are more sensitive to disease risk than others. For soybean rust, the response of industry to soybean rust in the United States has been much greater than in South American countries.

Level 3. This includes laypersons or producers who have a limited knowledge of plant pathology. Some highly competitive producers, however, have

good plant pathology knowledge and should be considered in level 1. For level 3, communication is made through indirect approaches. Information is delivered mostly through media, agricultural magazines, or radio, where normally the maximum risk of a disease is presented without further explanation. Risk is often selectively stated by media to achieve sensational effects. In the dissemination of soybean rust information, only maximum yield losses of assessments (loss of 80%) were used, without providing the dispersal assumptions used in the assessment. Unfortunately, producers are handlers of risk in production, and risk information indirectly delivered to them decreases the effectiveness and quality of risk management. Overreactions to soybean rust were common among the US soybean producers in the first two years.

It is important to disseminate risk information differentiated by simplicity or complexity to different receivers to avoid confusion or panic and to maintain the credibility of the research community. One example was the early release of spore dispersal assessment. In 2003 winter, a USDA trajectory analysis for the Western Hemisphere was prematurely released. The results showed that air parcels carrying fungal spores from lower elevations in Brazil soybean production regions could reach the United States. Unfortunately, this statement was interpreted by receivers as indicating the possibility of soybean rust occurrence in the coming season in the United States, which caused unnecessary panic in some US growers who consequently purchased fungicides for the predicted arrival of spores during the next season.

Future research for risk assessment

Our knowledge in epidemiology has enabled us to assess environmental suitability, establishment potential, and survival potential after the introduction of a plant disease. By adding yield loss models, yield loss potential can be determined. The outcome of risk assessment from the above-mentioned components represents the worst-case scenario. For soilborne diseases, the establishment range of a disease is equivalent to the distribution range or the range suitable to disease occurrence, with the damage range being smaller than the pathogen distribution range. Assessment with the worst-case scenario may not approach the realistic maximum damage of a soilborne disease. However, for airborne diseases, the establishment range of a disease would be no larger than damage range and smaller than the suitability range, depending on the dispersal potential of the disease. To have a more realistic assessment, the long-distance continental dispersal potential of the airborne disease needs to be predicted.

Dispersal potential is a key uncertainty in predicting the risk of an exotic airborne disease, which is important in determining the entry risk and damage risk after establishment. For the example of soybean rust, uncertainty of the disease risk depends on our understanding of dispersal potential of the disease. Before entry of this disease, dispersal information was critical to determine when the disease would arrive in the United States. Such information is critical for chemical companies to determine when to stockpile fungicide for disease control. For commodity groups, entry time is used to embargo the importation of soybean from occurrence countries, a temporary strategy to raise the local market price.

Once a disease establishes in the new country, its damage potential to the production region depends on the seasonal dispersal of airborne spores from overwintering regions into major production regions. Models for the dispersal potential are still in their infancy, and the prediction needs to determine the source area and its relationship to inoculum density in the receiving region, after long-distance dispersal. Such a relationship should exist in a pathosystem as demonstrated by Zeng (1988) in wheat stripe rust.

Latent period after entry. There are two critical times after the entry of an exotic plant pathogen into a new geographic region: time of first detection and time of first outbreak. For risk mitigation, when the pathogen can be detected after entry and when the first outbreak would occur are key questions. Practically no introduced diseases have produced severe, region-wide epidemics in the first season of detection in the United States. Similarly, soybean rust in South America did not cause economically significant losses in the 2000–2001 growing season after it was first found in southern Brazil and Paraguay. It caused losses $>$\$125 million in the second year of detection (Yorinori et al., 2003). However, predicting when an introduced pathogen would become prevalent after subsequent entry and reach outbreak levels is

critical to disease prevention or risk management. The available response period depends on detection efficiency—and the earlier the better.

References

Bromfield KR, Melching JS and Kingsolver CH (1980) Virulence and aggressiveness of *Phakopsora pachyrhizi* isolates causing soybean rust. Phytopathology 70: 17–21.

Bromfield KR (1984) Soybean rust. Monograph No. 11. Am. Phytopathol. Soc., St. Paul, MN.

Hundekar AR and Hiremath PC (2001) Development, prevalence and seasonal severity of soybean rust caused by *Phakopsora pachyrhizi* Syd. Karnataka Journal of Agricultural Science 14: 71–75.

IPCC (2001) Agriculture in a changing climate: Impacts and adaptation. In: Watson RT, Zinyowera MC and Moss RH (eds) Climate Change 1995. Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses (pp. 427–467) Cambridge University Press, Cambridge.

Keogh RC (1974) Studies on *Phakopsora pachyrhizi* Syd.: The Causal Agent of Soybean Rust. M.S. Thesis, University of Sydney, Sydney, Australia.

Kochman JK (1979) The effect of temperature on development of soybean rust. Australian Journal of Agricultural Research 30: 273–277.

Kuchler F, Duffy M, Shrum RD and Dowler WM (1984) Potential economic consequences of the entry of an exotic fungal pest: The case of soybean rust. Phytopathology 74: 916–920.

Luo Y, Morgan DP and Michailides TJ (2001) Risk analysis of brown rot blossom blight of prune caused by *Monilinia fructicola*. Phytopathology 91: 759–768.

Madden LV and Van den Bosch F (2002) A population-dynamics approach to assess the threat of plant pathogens as biological weapons against annual crops. Bioscience 52: 65–74.

Madden LV and Wheelis M (2003) The threat of plant pathogens as weapons against US crops. Annual Review of Phytopathology 41: 155–176.

Marchetti MA, Markle GM and Bromfield KR (1976) effects of temperature and dew period on germination and infection by uredospores of *Phakopsora pachyrhizi*. Phytopathology 66: 461–463.

Meching JS, Dowler WM, Koogle DL and Royer MH (1989) Effects of duration, frequency, and temperature of leaf wetness periods on soybean rust. Plant Disease 73: 117–122.

Patil VS, Wuike RV, Thakare CS and Chirame BB (1997) Viability of uredospores of *Phakopsora pachyrhizi* syd. at different storage conditions. Journal of Maharashtra Agricultural University 22: 260–261.

Pivonia S and Yang XB (2004) Assessment of potential year-round establishment of soybean rust throughout the world. Plant Disease 88: 523–529.

Rosenzweig C, Iglesias A, Yang XB, Epstein PR and Chivian E (2001) Climate changes and US agriculture: The impacts of warming and extreme weather events on productivity, plant diseases, and pests. Global Change and Human Health 2: 90–103.

Roy KW, Rupe JC, Hershman DE and Abney TS (1997) Sudden death syndrome of soybean. Plant Disease 81: 1100–1111.

Scherm H and Yang XB (1995) Interannual variations in wheat rust development in China and the United States in relation to the El nino/southern oscillation. Phytopathology 85: 970–976.

Scherm H and Yang XB (1999) Risk assessment for sudden death syndrome of soybean in the north-central United States. Agricultural Systems 59: 301–310.

Sutherst RW and Maywald GF (1985) A computerized system for matching climates in ecology. Agricultural Ecosystems and Environment 13: 281–299.

Sutherst RW and Maywald GF (1991) Climate modeling and pest establishment. Plant Protection Q. 6: 3–7.

Weltzien HC (1972) Geophytopathology. Annuual Review of Phytopathology 10: 277–298.

Wilkerson GG, Jones JW, Boote KJ and Mishoe JW (1985) SOYGRO, University of Florida, Gainesville.

Yang XB, Dowler WM and Tschanz AT (1991) A simulation model for assessing soybean rust epidemics. Journal of Phytopathology 133: 187–200.

Yang XB and Feng F (2001) Ranges and diversity of soybean fungal diseases in North America. Phytopathology 91: 769–775.

Yang XB and Scherm H (1997) El Nino and infectious disease. Science 275: 739.

Yang XB (2003) Risk assessment: Concepts, development and future opportunities. Plant Health Progress, http://www.plantmanagementnetwork.org/sub/php/review/2003/concepts/.

Yorinori JT, Pavia WM, Frederick RD, Costamilan LM, Bertagnolli PF, Hartman G, Godoy C and Nunes J Jr (2003) Epidemics of soybean rust (Phakopsora pachyrhizi) in Brazil and Paraguay from 2001 to 2003. Phytopathology 93:S103.

Zeng S (1988) Inter-regional spread of wheat yellow rust in China. Phytopathologica Sinica 18: 219–223.

Zeng (2003) Discussion on the macro-phytopathology. Review of China Agricultural Science and Technology 5: 3–7.

European Journal of Plant Pathology (2006) 115:35–51
DOI 10.1007/s10658-005-4050-2

Ecological genomics and epidemiology

K.A. Garrett[1,*], S.H. Hulbert[1], J.E. Leach[2] and S.E. Travers[1]
[1]*Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA;* [2]*Department of Bioagricultural Sciences and Pest Management, Colorado State University, Ft. Collins, CO, 80523, USA;* **Author for Correspondence (Fax: +1-785-532-5692; E-mail: kgarrett@ksu.edu)*

Accepted 13 October 2005

Abstract

The huge amount of genomic data now becoming available offers both opportunities and challenges for epidemiologists. In this "preview" of likely developments as the field of ecological genomics evolves and merges with epidemiology, we discuss how epidemiology can use new information about genetic sequences and gene expression to form predictions about epidemic features and outcomes and for understanding host resistance and pathogen evolution. DNA sequencing is now complete for some hosts and several pathogens. Microarrays make it possible to measure gene expression simultaneously for thousands of genes. These tools will contribute to plant disease epidemiology by providing information about which resistance or pathogenicity genes are present in individuals and populations, what genes other than those directly involved in resistance and virulence are important in epidemics, the role of the phenotypic status of hosts and pathogens, and the role of the status of the environmental metagenome. Conversely, models of group dynamics supplied by population biology and ecology may be used to interpret gene expression within individual organisms and in populations of organisms. Genomic tools have great potential for improving understanding of resistance gene evolution and the durability of resistance. For example, DNA sequence analysis can be used to evaluate whether an arms race model of co-evolution is supported. Finally, new genomic tools will make it possible to consider the landscape ecology of epidemics in terms of host resistance both as determined by genotype and as expressed in host phenotypes in response to the biotic and abiotic environment. Host phenotype mixtures can be modeled and evaluated, with epidemiological predictions based on phenotypic characteristics such as physiological age and status in terms of induced systemic resistance or systemic acquired resistance.

Introduction

The field of plant disease epidemiology has incorporated new technologies and perspectives on biology as they have become available, from computer simulation modeling to automated environmental sensing. Over the past decade, the study of DNA within all areas of biology has gone through a revolution, providing new types and new quantities of genomic data for epidemiological analyses. Given the advent of new technologies associated with rapid analysis and miniaturization, informatics, and molecular biology, it is now possible to expand the scale of studies of both agricultural and wild species to include entire genomes. The high-throughput advances associated with genomics and other "-omics" (e.g. proteomics, metabolomics) have allowed an unprecedented collaboration among scientists working at different biological scales and have fostered a new science, ecological genomics. In this "preview", we discuss how these new approaches may dovetail with plant disease epidemiology.

Epidemiology has already benefited from information about the population genetics of pathogens, as reviewed by Milgroom and Peever (2003). By simultaneously studying how pathogen gene frequencies change within and among populations as a result of both natural selection and gene flow, and how pathogen populations grow and spread, it has been possible to track disease outbreaks (e.g., Zwankhuizen et al., 1998),

develop predictions about sources of inoculum and pathogen life cycles (e.g., Cortesi et al., 2000; Cortesi and Milgroom, 2001), understand the evolution of virulence (Escriu et al., 2000a, b, 2003), and make predictions about the durability of resistance in crop genotypes (Escriu et al., 2000a, b). Ultimately, modeling plant disease epidemics and pathogen evolution depends on a complete understanding of both plant and pathogen traits that influence the dynamics between a pathogen and its host. To completely understand any trait and its significance in a dynamic interplay between species requires the simultaneous use of molecular, cellular, organismal, population and ecological approaches. Past efforts to combine epidemiology and population genetics have come up against an upper limit on the number of ecologically important genes that could be surveyed or lack of information on gene function and significance. Yet, host plants, as well as pathogens, exist in a matrix of hundreds or thousands of other taxa and their genes. Population changes in pathogens, reproduction and dispersal will all depend on the interactions among these organisms that can influence the dynamics of resistance evolution and direct effects on pathogen populations (Antonovics, 2003).

The developing synthesis of a functional genomics approach combined with a population and ecological perspective promises to lead to new avenues of research and understanding of plant/pathogen interactions. Evolutionary and ecological functional genomics or EEFG (Feder and Mitchell-Olds, 2003) has as a goal to understand ecological and evolutionary processes that maintain genotypes and phenotypes. The emphasis so far has been on wild species, but agricultural systems offer both an important application and relatively well-characterized systems for experimentation. The field of ecological genomics will address new types of questions beyond applications based on molecular markers. Microarrays allow synoptic measurements of gene expression in tens of thousands of genes. Real-time PCR allows highly accurate quantitative evaluation of gene expression at many time steps. It will also be possible to identify hundreds or even thousands of organisms simultaneously from individual samples as microarrays are developed with sequences representative of desired sets of species, potentially including non-culturable species. Advances in sequencing allow analysis of great numbers of "markers" with added information about their likely role through reference to databases such as GenBank (Black et al., 2001), thus revealing the gene content of particular organisms.

Functional genomics, or the use of genomic technologies (e.g. microarrays) to find genes and polymorphisms that affect traits of interest and to characterize the mechanisms underlying those effects, has been applied effectively in agricultural contexts and has potential in natural systems. Functional genomics moves beyond simple sequence analysis to evaluate the function of particular DNA sequences through, for example, gene knockout mutants or gene activation mutations. These techniques have natural applications for the study of resistance and virulence, but might also be usefully applied in the study of other epidemiological features. By simultaneously scanning thousands of plant genes for changes in expression in response to variables of interest (e.g. stress, infection) it has been possible to identify candidate loci or suites of genes and molecular mechanisms involved in the phenotypic expression of key traits of economically important crop species (Frick and Schaller, 2002; Jones et al., 2002; Mysore et al., 2003). A great deal has been learned about plant defense against disease through the use of functional genomics and model plant systems such as *Arabidopsis* (Wan et al., 2002; Schenk et al., 2003; Whitham et al., 2003a).

An intriguing area of epidemiology that will develop with the availability of new tools for studying gene expression is the study of phenotypic resistance and its responses to the biotic and abiotic environment. Infection with an incompatible pathogen, or a virulent pathogen that causes cell death, can make a plant more resistant to subsequent infection by the same or different pathogens, a phenomenon designated systemic acquired resistance (SAR; Durrant and Dong, 2004). The SAR response in *Arabidopsis* confers resistance to several diseases (Ryals et al., 1996). Resistance to pathogens can also be influenced by non-pathogenic organisms; systemic changes in disease resistance in response to colonization by rhizosphere-colonizing *Pseudomonas* bacteria have been well-documented and are commonly referred to as induced systemic resistance (ISR; Iavicoli et al., 2003; Cui et al., 2005). Dissection of the SAR and ISR signaling systems in *Arabidopsis*

indicate they are controlled by different pathways and signaling molecules with some common components. Understanding which genes are expressed during specific defense responses can provide indications of what pathways are activated in different biotic environments (Pieterse and van Loon, 1999). Tools are now available to begin studying these phenomena more widely in epidemiology.

This paper will address the following topics in ecological genomics. (1) Population genetics and population genomics can inform epidemiology to further our understanding of epidemics and to provide insights for disease management. We will also consider how studies of gene expression can potentially add predictive power at finer spatial and temporal scales than was possible in the past. (2) Models of populations and communities may apply to analogous systems of gene expression within organisms and in populations of organisms to inform a "population biology" of gene expression. (3) Genomics can contribute to understanding of resistance gene evolution and durability of resistance. (4) The landscape ecology of host populations and communities, in terms of both genotypic and phenotypic resistance, can now be studied more thoroughly as it affects epidemics. In addressing these topics, we will emphasize genes that influence the relationship between plant host and pathogen, but the same general concepts would apply to interactions between plant species, between plants and insect herbivores, etc.

How population genetics and population genomics can inform epidemiology

Epidemiology has traditionally used information about host species, pathogen and vector species, and environmental variables such as temperature and precipitation to predict epidemic progress. These models can be adapted to incorporate much more detailed information about the genomic status of the host and pathogen communities considered in the context of a broadly defined environment, i.e., defined to include abiotic components and potentially the complete community metagenome of soil (Rondon et al., 2000) or other systems (Figure 1). Information about the soil

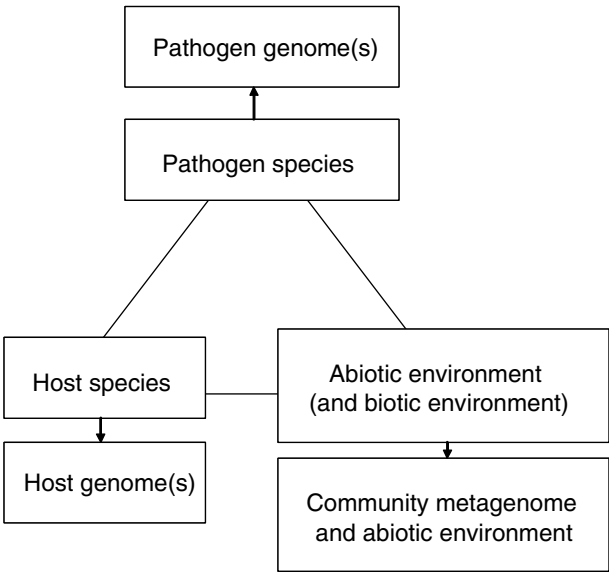


Figure 1. The traditional disease triangle depicts prediction of epidemics based on interactions between pathogen species, host species, and the abiotic environment. It is now much easier to obtain information about the complete genotype and current gene expression of host and pathogen, and there is even the potential to obtain this information for complete communities such as those in soil, the rhizosphere, and the phyllosphere, as well as endophytic communities. Models about a hierarchy of features of "genomic status" might be experimentally evaluated in this context. For example, "The host landscape is sufficiently described to predict epidemic features and outcomes by information about... ● ... host community composition (in terms of species)." ● ... a specific subset of the host genotype sequence(s)." ● ... the host genotype sequence(s)." ● ... a subset of host gene expression." ● ... complete profiles of host gene expression (now and/or in the past)."

Table 1. The temporal and spatial scale of variation in different components of host genomic status

Component of host genomic status	Temporal scale	Spatial scale
In annual monoculture		
Species	Cropping season	Size of field in many conventional systems
Genotype	Cropping season	Size of field in many conventional systems
Gene expression	Less than one hour to cropping season	Part of one individual to size of field
In unmanaged systems		
Species	Days to decades	One individual to majority of plant community
Genotype	Days to decades (potential for somatic mutation)	One individual to majority of species (in clonal species)
Gene expression	Less than one hour or until phenotype expressed (days for defense reaction, months for flowering, etc.)	Part of one individual to majority of area

metagenome may contribute to an understanding of disease suppressive soils that develop over time as microbial populations respond to the buildup of pathogen populations. For example, soils suppressive to the wheat take-all pathogen and potato scab have been described, with fluorescent pseudomonads and streptomycetes, respectively, being the likely causes of suppressiveness (Weller et al., 2002). Advances in genomics also make it possible to characterize the genomic status of host plants at a much finer temporal and spatial scale than in the past (Table 1). The addition of gene expression as a response variable or predictor variable in epidemiological models has the potential to shift the scale of inquiry to hours and millimeters. Monitoring the expression of genes in specific defense pathways, or individual genes that reflect the expression of the pathways, could be used to predict the outcome of pathogen infection in individual plants or plant organs. For most diseases, progress in determining the efficacy of different defense responses for controlling specific pathogens and how the responses become distributed throughout the plant must be made before this information is useful. Then epidemiologists will need to perform a range of exploratory field studies to identify the forms of predictors that are most useful for inclusion in more detailed follow-up studies. For example, if the early induction of senescence-related gene pathways were observed to occur, would this be related to reduced epidemic potential at a field scale?

Characterizations of populations may include the composition of both qualitative features produced by different genotypes and quantitative features produced by different levels of gene expression in what may be the same genotype. Evaluation of qualitative features might be performed using marker or sequence studies, while evaluation of quantitative features might be performed using microarrays or real-time PCR. Studies of gene expression in agriculturally important host plants have expanded remarkably, with microarrays now available for several major crop species. These allow host resistance to be assessed as an outcome of gene expression. In addition, the expression of plant genes in response to non-pathogenic microorganisms may be highly relevant to epidemiology, as it may provide an understanding of how plants select for rhizosphere flora that are antagonistic to pathogens, for example (Smith and Goodman, 1999). Microarray analyses can be used to identify sets of coregulated genes and their common regulatory elements (e.g., Maleck et al., 2000; Chen et al., 2002), which may both reveal different response pathways and allow selection of smaller sets of indicator genes to represent particular stress response pathways. Microarrays developed using genes from one plant species may also be applied, with some caveats, in studies of related species; for example Travers et al. (in preparation) have applied maize microarrays to study gene expression in the related tallgrass prairie grasses *Andropogon gerardii* and

Sorghastrum nutans, and have identified statistically significant responses to simulated climate change in native field populations.

New genetic information can be used to refine state transition models such as Susceptible-Infected (SI) models (e.g., Otten et al., 2003). Rather than modeling host individuals as simply "susceptible and uninfected" or "infected", more details about the state of individuals could be included. The first simple modifications might include broad genotypic resistance and susceptibility. Further refinement could include transitional states of greater or lesser susceptibility based on physiological age, and probabilities of exposure to other organisms that would induce resistance. Matrix-based models of probabilities of transitions from one state to another could be applied to predict long-run states such as disease severity or survival of different genotypes. Such models could potentially be applied to develop both epidemiological theory and better disease management schemes. In the short-run, they could be used to ask questions about the incremental benefits of adding information about host phenotype to epidemic models. In the long-run, these models could contribute to a much deeper understanding of epidemic dynamics.

The more complete genetic information from DNA sequencing can be used to study long-standing questions of population structure, host specificity, and phylogenetics. Due to the growth of sequence databases and the reduction in PCR amplification and sequencing costs, determining the sequence of a specific gene in a microorganism is often the most efficient way to determine the species of the microorganism. Databases now carry information on a huge number of organisms, and matching an unknown sequence to the sequences in a database like GenBank takes only a few minutes, although one must keep in mind that not every sequence accession in GenBank is annotated correctly. Reduced sequencing and DNA amplification costs make the identification of components of large microbial populations feasible. Entire fungal or bacterial communities can be characterized taxonomically by incorporating new techniques such as "shotgun sequencing" of a community's collective genome and using genome database searches to identify species and predict gene function (Venter et al., 2004). At a finer scale, sequencing specific genes in pathogen mutants may give insight into cost of virulence (Vera Cruz et al., 2000; Ponciano et al., 2004). Sequencing can also be used to evaluate the potential repertoire of resistance genes available, to the extent that sequence similarity can predict functional similarity (Bai et al., 2002). Examples include NBS-LRR genes, the largest class of disease resistance genes. Plant genome projects have indicated there are approximately 150 in *Arabidopsis* and more than three times this number in rice. Information about the number of resistance genes available may contribute to resistance gene deployment strategies. The identification of sequences associated with resistance genes may also be applied to related plant species to answer long-standing questions about the number and type of resistance genes in natural populations (Gilbert, 2002). Microbial genome projects are providing similar estimates of the number and types of effector proteins in a single organism, such as the number of gene products transferred into plant cells by the type III secretion system of *Pseudomonas syringae* strain DC3000 (Collmer et al., 2002; Alfano and Collmer, 2004; Rohmer et al., 2004; Chang et al., 2005). These are not only important proteins that make the bacteria successful pathogens, but also the targets of plant disease resistance proteins. These are just a few examples of how partial and whole genome sequencing projects can contribute to understanding host–pathogen interactions.

Studies of gene expression in pathogens are still limited, but, where available, are being used to understand expression of genes during plant colonization, and under various cultural practices. As more whole genomes are sequenced, microarrays using various platforms are becoming available for several pathogens. As examples, arrays exist for the rice blast fungus and for several bacterial plant pathogens. Techniques other than microarrays are also being applied to understand gene expression; for example, serial analysis of gene expression (SAGE) has been applied to study gene expression in response to rice blast infection (http://www.mgosdb.org/). Microarrays can also be used in comparative genomics studies of closely related pathogens using full genome sequences. For example, the gene content of the human pathogen *Yersinia pestis* has been studied as an indicator for adaptation (Chain et al., 2004). Genomes have been evaluated to determine what is missing in a

fastidious, xylem limited species like *Xylella fastidiosa* by comparison to other less fastidious bacteria (Van Sluys et al., 2002). The genomes of *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* are being compared for insights into why the first is systemic in xylem while the second grows in mesophyll (A. Bogdanove, pers. comm.).

The greater availability of genetic information will allow plant pathologists to move ''beyond the inoculation experiment'' in studies of the genetic features of host–pathogen interactions. In the past, painstaking and expensive analyses of genetic expression in host–pathogen interactions have generally been applied to studies of pathogens introduced to hosts either in highly conducive environments, in the case of rust fungi, for example, or directly inoculated into host tissues, in the case of many bacterial pathogens. In contrast, it would be extremely interesting and valuable to have a greater understanding of the genetic basis for the broad range of other epidemiological features that are important in determining population-level interactions between host and pathogen. For example, from the standpoint of the pathogen, aside from direct effects on virulence or aggressiveness, what are the genes most important for features such as survival in soil or on plant surfaces, tolerance for temperature extremes, dispersal capability, or other specialized features such as conversion from production of urediniospores to production of teliospores in rust fungi? At larger epidemic scales, the genetic characteristics most important to dispersal might be those that affect survival of propagules under challenging environmental conditions. These characteristics would help determine whether the long-distance transport events so important to establishment of epidemics in new areas occur or not. From the standpoint of the host, what genes are most important for predicting epidemics aside from direct resistance genes, including features such as the probability of escape through faster or slower movement through developmental stages, ''leaking'' of compounds in the phyllosphere or rhizosphere, and architectural features that affect microclimate? Such information would be useful both for applied crop plant breeding programs and for understanding resistance profiles in natural plant populations.

There is the potential to identify genes predictive of epidemiological features using ''comparative genomics'' to inform ''comparative epidemiology''. For example, Kranz (2002) discusses several disease parameters influenced by host plant resistance that together predict epidemic rates and outcomes: disease intensity, incubation period, latent period, infection efficiency, disease efficiency, infection rate, lesion size, infectious period, and sporulation intensity. In comparative epidemiology, the differences in these parameters between host–pathogen systems can be evaluated both in terms of their typical values and the frequency distribution of these values in response to typical forms of resistance. The availability of gene expression data will also make it possible to study disease parameters as a function of measures of gene expression, given a particular genotype (Figure 2), in the same way that the expression levels of key genes associated with the initiation of flowering have been used to predict flowering time (Welch et al., 2003, 2005).

There is a basic need in epidemiology for improved diagnostic systems and genomic advances will greatly expand the tools available. For example, as models of the risk of invasion by particular plant pathogens are constructed, their validation depends on researchers' ability to determine precisely the abundance of pathogens in a range of environmental settings. In their simplest form, such studies require the ability to detect and identify particular species of pathogens. Diagnosis may also be taken to more sophisticated levels through the ability to detect particular genotypes,

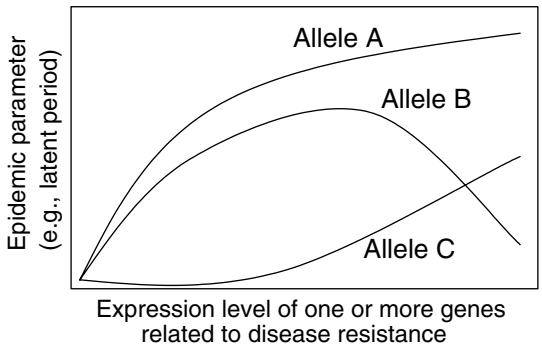


Figure 2. Schematic of possible relationships between gene expression levels and epidemic parameters such as latent period, infection rates, lesion size, etc. Such a relationship could be incorporated in models to refine predictions of epidemic outcomes such as disease severity or disease incidence. The different lines indicate possible differences in response for different alleles.

in particular, those that are capable of causing disease. Presence of genes for these traits, such as genes related to pathogenicity, toxin production, and other epidemiological features, if known, could be used to more reliably measure genotypes in a population responsible for disease. A particularly important application might be the identification of disease, through evaluation of host or pathogen, when infection is still at very low levels, to allow for early management that might, for example, allow an invasive pathogen to be eradicated before it has become well-established. Further refinement for successful diagnosis of gene expression may allow assessment of features such as quorum sensing status (van Bodman et al., 2003). The use of microarrays also opens the possibility of synoptic rapid-throughput diagnostic procedures for huge numbers of organisms for the study of the community characteristics of systems such as disease-suppressive soils, the phyllosphere, and endophytic communities. These approaches could bring great advances in understanding microbial biodiversity, including the potential to find new non-culturable putative pathogens through scans for microbial genes used for taxonomic classification or even genes associated with pathogenicity. Epidemiologists might also make good use of a genomic tool that would allow them to study the past presence of pathogens through on-going expression in host populations. Such a measure of pathogen ''footprints'' could support studies of long-term epidemics and changes in host resistance over time. But it appears that an indicator of past infection is not readily available in plants, or at least researchers have not yet discovered how to recognize it.

How models of populations and communities may apply to systems of gene expression to inform a ''population biology'' of gene expression

A null model for how models of populations and communities apply to the study of gene expression might be ''consilience''; that is, the null model might be that the same models will apply across scales, so tests could be developed to determine where population models do and don't adequately explain patterns of gene expression.

Models from population biology can be applied in the study of gene expression in three general ways. First, at the smallest scale, genes may be conceptualized to interact within a cell comparably to the way that species interact within an ecosystem (Mauricio, 2005). For example, it may be useful to apply such models to the interactions between different defense response pathways. There is evidence the jasmonate (or ethylene) and salicylic acid pathways affect somewhat different pathogens and pests but also interact with each other (Thomma et al., 1998; Glazebrook et al., 2003). Depending on the response examined, they may sometimes be viewed as complementary (van Wees et al., 2000) and in other cases as in competition (Spoel et al., 2003).

Second, an individual plant may be conceptualized as a population of cells or organs across which gene expression occurs. It is now possible to measure gene expression in individual plant cells (Kerk et al., 2003; Nakazono et al., 2003) so the spatial pattern of expression through an individual host can be measured and modeled at whatever spatial grain is motivated by the experimental questions. Spatial patterns of defense responses between cells are relevant both to how effective defense responses are to pathogen challenge and to how the host responds to adjacent or subsequent challenges by the same or other pathogens. Could models of the dispersal of individuals through ecological landscapes be usefully adapted to describe the dispersal of gene products within and between cells? State transition models could be applied to individual plants in cases where it makes sense to treat them as a set of units, such as different tissues and organs, each of which would have its own expression status. This could be addressed using a variation on SI models. Predictions based on these models might include the predicted infection level as well as the predicted plant growth rate.

Third, experiments in epidemiology might begin with models within individuals, predicting infection levels based on the expression of particular genes, and then expand on these to predict infection rates in plant populations based on the gene expression rates in individuals. A null model for such a study might be that the mean expression level of individual hosts is fully predictive of the level of infection in the population. In contrast to the null model, it would be interesting to determine whether the frequency distribution of the expression rates, and perhaps even their spatial pattern,

in the different host individuals would substantially improve predictions of epidemic features, just as different patterns of disease severity across individuals can result in different overall yields for the same mean disease severity (Hughes, 1996). Simulation modeling might be used for initial tests of the sensitivity of epidemic outcomes to such patterns of expression. Addressing questions such as these with an understanding of mechanism will require considerably greater understanding of the relationships between gene expression and gene product physiological function. This third scale is addressed further in a later section.

The study of gene expression offers a new method for measuring integrated effects of environmental variation (Figure 3). Environmental variables such as temperature and precipitation are standard predictors of disease progress in epidemiological models (Jeger, 2004), and integrated forms such as "growing degree days" are already commonly used to predict growth stage as a model component. Different types of host responses may be integrated over different time intervals. Growth stage, or more specialized responses like the development of sun and shade leaves, are the products of the cumulative effects of gene expression, as affected by environment, over a period of time. Younger tissues might only experience "indirect" effects from past environmental conditions, perhaps as an analog to maternal and grandmother effects in individuals. Induced systemic resistance might be an example of short-term gene expression in response to non-pathogens while systemic acquired resistance might be an example of short-term gene expression in response to pathogens or to chemical stimulants. The timing of infection and its effects on losses in productivity can also be evaluated through variations on time-of-infection models for predicting yield loss (Madden et al., 2000) that include explicit descriptions of host gene expression in response to infection. The schematic model in Figure 3 applies most directly in agricultural systems in which a genotype is generally maintained, at least for a season, through removal of competitors. A more complicated model might be developed in which host genotypes can be replaced by other plant genotypes. The schematic might also be adapted to take into account the possibility of thresholds such that long-term changes in phenotype could be produced by short-term gene expression at critical time points in development.

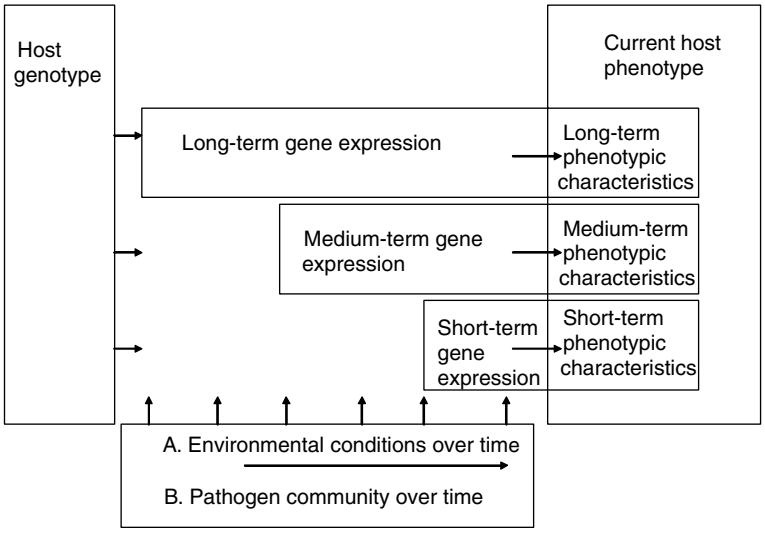


Figure 3. The current host phenotype, at any spatial scale within a host individual, is a form of integration of the individual's environment, including the composition of the pathogen community, acting on the host genotype. Long-term phenotypic characteristics would include features such as physiological age of leaves or roots, forms of specialization such as the development of sun or shade leaves, and other characteristics that may influence disease resistance. Short-term characteristics might include features such as upregulation of pathways contributing to induced resistance. Of course, host gene expression will also influence pathogen populations and even the abiotic microenvironment.

How genomics can contribute to understanding of resistance gene evolution and durability of resistance

A major goal of agricultural plant pathology is the development of durable resistance to plant pathogens in agricultural species. ''Durable resistance'' has been defined as resistance that is still effective after it has been deployed over a wide area, over a long period of time, in a disease-conducive environment (Johnson, 1981). Without durable resistance, plant breeders are forced to continually incorporate new resistance genes in crop varieties as pathogen populations adapt to infect older varieties with previously deployed resistance genes. An understanding of the evolution of host and pathogen genes affecting host–pathogen interactions is needed to form strategies for the durable deployment of resistance in agriculture. It has long been thought that understanding of the relative importance of the disease effector proteins from bacterial and fungal pathogens that are detected by R genes (i.e., the products of avirulence genes) should provide insight into which R genes might have more durable effects, but this idea has had limited impact because of the difficulty of identifying and characterizing these effector protein genes. Comparative genomic approaches for identifying these genes and functional genomic approaches to obtain 'knocking-outs' of their function is making this increasingly feasible (Leach et al., 2001). Some resistance genes, like *mlo* of barley (Buschges et al., 1997), may confer resistance without interacting with specific pathogen effector proteins. These genes may provide non-specific resistance by changing the physiology and gene expression of the plant before pathogen challenge (Wolter et al., 1993). Gene expression analysis has indicated other resistance genes with suspected non-specific effects may alter expression of defense genes before pathogen challenge (Bowden and Hulbert, unpublished). Such analysis should be useful in identifying genes controlling non-specific and thus durable resistance and also provide insight into the possible physiological cost of the resistance.

The isolation and sequence analysis of several resistance genes has provided insight into the evolution of disease resistance in plants (Hulbert et al., 2001). Some of the results of these analyses are consistent with a classical evolutionary arms race model, while others are not (Hulbert, 1998). High levels of sequence variation have been observed at most R gene loci examined. This is consistent with the arms race model's prediction that R genes would evolve rapidly, creating novel alleles with new specificities in response to pressure imposed by rapidly evolving pathogen populations. Loci like *L* of flax (Ellis et al., 1999), which is structurally simple but has multiple resistance alleles, exhibit extremely high levels of polymorphism compared to most genes. At some R gene loci, the patterns of nucleotide substitution between alleles or family members show evidence of the types of diversifying selection that might be predicted by an arms race model. While polymorphic nucleotides are usually synonymous (not affecting the encoded amino acid) at most loci, the opposite is true of certain regions of some R gene loci. This is most often true in regions of R genes thought to code for the ligand recognition part of the protein, like the leucine-rich repeat regions (Parniske et al., 1997; Meyers et al., 2003). Evidence of diversifying selection in other regions of R genes, like the TIR domain-encoding regions of the L alleles, has suggested they may also be involved in recognition (Luck et al., 2000).

One interpretation of an arms race evolutionary progression is that there should be little variation at a given R gene locus at one point in time and that most R gene alleles should be fairly recent in their evolutionary origin. This would be expected if new highly effective R genes arose periodically and replaced the older 'defeated' alleles. The polymorphic nature of many R gene loci indicates this is apparently not the case for most of them. In fact the partitioning of polymorphism between functional alleles and non-functional alleles at the *Rpm1* and *Rps5* loci of *Arabidopsis* indicated that the classes of alleles have co-existed for a long period (Bergelson et al., 2001; Tian et al., 2002), probably the result of some form of balancing selection. While actual estimates of the age of specific resistance gene alleles are not available, this may be an indication that some R gene alleles are ancient. In contrast, no evidence that resistance alleles are ancient has been obtained by sequencing the same resistance allele from multiple germplasm accessions. If resistance alleles are indeed ancient, it should be possible to identify versions that have accumulated extensive neutral sequence polymorphisms. This has not yet been the case in the limited experiments that have been

conducted (Caicedo et al., 1999; Jia et al., 2003; Smith et al., 2004). The low nucleotide diversity among the functional alleles of these loci is consistent with the idea that they could be recently evolved, although other explanations are possible.

The sequence evidence collected to date implies that different R gene loci are evolving in different manners. For example, some appear to be under strong diversifying selection while others do not. The implications of an R genes' evolutionary history for the stability of the resistance it confers is not clear, but the ability to predict durability based on genomic analysis would be quite useful for crop improvement strategies. Molecular analyses of resistance proteins and their corresponding avirulence proteins have indicated that some physically interact directly (Scofield et al., 1996; Tang et al., 1996; Jia et al., 2000, Deslandes et al., 2003) while others detect modifications of other host components (Mackey et al., 2003; Axtell and Staskawicz, 2003). It is possible that whether an R gene recognizes effector (avirulence) genes directly or protects host targets from modification by effector proteins affects the type of selection pressure driving its evolution (Ponciano et al., 2003). This association, however, is not yet clear due to the small number of interactions in which this type of information is known. For R proteins that guard other host components, it is not clear if the nature of the host protein being guarded affects the durability of the R gene, but it might be expected that some targets are more important to the pathogens ability to cause disease than others. The nature of the effector gene, particularly how essential it is to pathogenicity, has been proposed by many to affect R gene durability and recent data indicates this is true (Vera Cruz et al., 2000).

One response to the problem of rapid "breakdown" of resistance in agricultural systems has been a shift by some plant breeders toward greater use of minor resistance genes that each contribute a small amount of resistance and are generally thought to be more durable (Leung et al., 2003; Liu et al., 2004). However, these genes, because of their small effects, are more difficult to study in the field and even to recognize by the phenotypes of individual plants. The use of genetic markers has made the incorporation of minor genes easier, but the problem remains that, because we do not know what genes are responsible for quantitative traits, the association of the markers with the traits is not absolute. Genomic tools will allow discovery of the genes responsible for quantitative traits, and may make it easier to determine whether resistance governed by quantitative traits is truly more durable; whether the effects of QTL are actually less pathotype specific, or whether an apparently more durable effect may be mediated by a weaker selection on individual pathogen genes. To the extent that function can be inferred from sequence, the response of pathogens to particular minor genes may be better predicted as this information becomes available. It will be particularly useful if comparative genomics would allow predictions of the interactions between minor resistance genes and their responses to abiotic and biotic environments. Functional genomics may also contribute to the identification of new minor resistance genes. QTL analysis or the identification of quantitative trait loci provides a powerful tool for assessing the fitness consequences for genes including resistance genes. For example, Newcombe and Bradshaw (1996) used it to identify genes of large effect that changed the resistance of poplar to pathogenic *Septoria populicola* with community level effects.

The study of pathogen and host co-evolution in natural plant populations is also important for understanding what role pathogens may have played in structuring plant communities. In studies demonstrating the importance of genetic variation in host plant species within a larger community that includes pathogens, hybridization of host plants (e.g. willows, sagebrush, oaks) has led to fundamental changes in the species composition of the entire community (reviewed in Whitham et al., 1999). This "extended phenotype" effect would be reflected in the context of epidemiology by the dying out of some pathogens and replacement with others (Whitham et al., 2003b). Agricultural systems and unmanaged systems offer an interesting contrast, because the selection pressures in unmanaged systems are "direct" while selection pressures in agricultural systems are mediated by human decision-making.

The landscape ecological genomics of host populations and communities, in terms of both genotypic and phenotypic resistance

Once meteorological measurements could be collected using automated systems, epidemiologists

were faced with the question of what temporal scale of resolution was needed for understanding epidemic progress. Information about variation in temperature at the scale of minutes is not generally needed for predicting epidemic features. But whether predictions are improved by resolving differences in temperature at the scale of days or weeks will vary from one host–pathogen system to another, based on characteristics such as pathogen generation time, and requires attention for the construction of good models. Information about variation in meteorological features across space is still not so readily collected at very fine scales, though the increasing availability of "smart dust" and other tiny wireless sensor networks may change that (e.g., http://webs.cs.berkeley.edu/). The same question of appropriate scale of variation to include for modeling will arise for the new spatial maps of meteorological features. Similarly, the potentially huge amount of information about gene expression will require research to determine what scale of variation is important to include in prediction of epidemics for particular host–pathogen systems. The cost of microarray analyses limits the number of samples in time and space for now, but as costs become less limiting, epidemiological research will focus more on determining optimal scales of variation in expression data to include in predictive models.

Plant disease epidemiology has developed models of disease foci and how these foci expand in time and space (Zadoks and Vandenbosch, 1994; Waggoner and Aylor, 2000), including studies of the spatial pattern of disease used to draw inference about modes of dispersal and to devise optimal sampling strategies. Landscape ecology also offers methods for studying spatial features with models for describing the relationships between organisms in landscapes and for describing the connectivity of features (With, 2002). In agricultural systems, the spatial pattern of host genotypic resistance is sometimes manipulated through the construction of intercropping systems and/or use of mixed genotypes within a crop species (Garrett and Mundt, 1999; Zhu et al., 2000; Mundt, 2002). And, of course, most unmanaged systems include a mixture of plant species that, with few exceptions (e.g., *Phytophthora cinnamomi*), do not tend to share the same pathogen species. Mixtures of susceptible and other genotypes make models of disease spread through space somewhat more complicated. Some models have assumed that epidemic "waves" move out from an initial point with constant velocity to simplify the modeling of the system, while other researchers predict that epidemic waves accelerate (Scherm, 1996; Cowger et al., 2005).

A genomics approach applied to epidemiology could explore multiple spatial and temporal scales as well as levels of detail in genomic status, perhaps employing cellular automata models (e.g., Kleczkowski et al., 1997; Figure 4). Within a host individual, the local phenotype might be at the scale of a leaf or of a cell. Local gene expression might be at the point of infection; for example, within compared to beyond a green island of host tissue formed around an infection by a rust fungus. Regional gene expression within an individual might be expression in tissues adjacent to infection. Within a host individual and its immediate environment, a wide range of pathogens may be present, specializing on different host tissues. Competition between particular pathogens may play out differently depending on the time of infection and the type of plant tissue (Adee et al., 1990; Al-Naimi et al., 2005). The question for epidemiologists will be what spatial and temporal resolution is needed for predicting epidemics with the new and upcoming abundance of data, as opposed to averaging over host and pathogen individuals' genomic status across space and time.

In host populations, "expression foci" in which host individuals share altered gene expression patterns may form around inoculum sources, with properties related to those of disease foci. Gene expression changes in hosts in response to exposure to pathogens and other microbes may range from increased resistance through SAR or ISR to increased susceptibility because of weakened tissue integrity. The effect of exposure to pathogens that do not infect has the potential to be substantial, at least temporarily; Calonnec et al. (1996) estimated that the infection efficiency of *Puccinia striiformis* was reduced by 44% when plants were previously exposed to an "inducer race" of the pathogen. At increasing distances from a primary inoculum source, exposure to inoculum may have occurred at more recent time points, potentially resulting in waves of different expression patterns surrounding the initial source areas. Spatial patterns of abiotic features, such as differences in topography that produce cooler or wetter local conditions, may

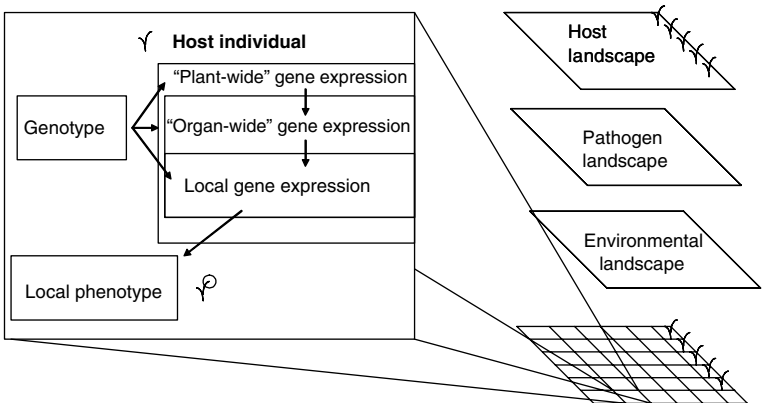


Figure 4. Each host individual is potentially influenced by the landscape of hosts, pathogens, and other biotic and abiotic environmental features. Within a host individual, these influences may play out through "plant-wide", "organ-wide", or more local gene expression, depending on the scale of variation of each feature in the landscape and how it acts upon the host individual. "Plant-wide" gene expression might include responses to factors such as drought and disease that alters water relations within the host. "Organ-wide" gene expression might include responses to factors such as stem or petiole lesions. Local gene expression might include responses such as localized forms of induced resistance. Models related to spatial scale and scale of genetic detail that could be experimentally evaluated as predictors of epidemic features include the following, presented in a hierarchy of increasing complexity. "The host landscape is sufficiently described to predict epidemic features and outcomes by information about... ● ... the abundance of host species" ● ... the abundance of host genotypes" ● ... the abundance and spatial pattern of host species" ● ... the abundance and spatial pattern of host genotypes" ● ... the mean level of gene expression among host individuals" ● ... the spatial pattern of gene expression among host individuals" ● ... the spatial pattern of gene expression within host individuals."

also produce expression foci relevant to epidemics. Studies of gene expression in landscapes may develop distinctions analogous to the distinction between a dispersal gradient and a disease gradient. Disease gradients may differ markedly from dispersal gradients if the success rates per unit of inoculum are low, particularly if the reproductive rates of the pathogen are density dependent (Garrett and Bowden, 2002). There may be similar effects for gene expression, such that thresholds of exposure to pathogen inoculum, for example, must be exceeded before substantial gene expression results. At much smaller spatial scales, gene expression in bacterial populations may give insights into quorum sensing and its implications for density dependent reproduction (van Bodman et al., 2003).

Epidemiologists have developed the terms autoinfection and alloinfection to describe infection of a target host individual by inoculum produced on the same target host individual vs. infection by inoculum produced on other host individuals, respectively (Robinson, 1976). The rate of autoinfection can be an important predictor for epidemics of non-systemic disease in mixed genotype host populations. If some host genotypes are susceptible and others are not, the reduction in epidemic rates on susceptible genotypes that would be predicted by the presence of other genotypes will be reduced if autoinfection rates are high; more inoculum will land on susceptible host individuals rather than being lost through dispersal to non-hosts (Garrett and Mundt, 1999; Mundt, 2002). It may prove useful to develop analogous concepts for gene expression, so that "autoinduction" of gene expression would occur when microbes associated with a particular plant individual disperse to other organs within that individual to induce SAR, ISR, or other reactions. By comparison, "alloinduction" would occur when microbes are dispersed to a different plant individual to induce these reactions. Higher rates of alloinduction compared to autoinduction would tend to result in higher mean levels of SAR or ISR in populations, and the rate of alloinduction would be a function of host size and the dispersal properties of the relevant microbe populations.

Feedback between host and pathogen would occur as pathogens disperse, infect or elicit other responses in plants, and then disperse further through a landscape of phenotypic resistance that has potentially been altered in response to previous dispersal. Good models of such a system would require the ability to predict plant

phenotypic resistance levels in response to the biotic and abiotic environment, pathogen phenotypes in response to "non-host" environmental features, plant phenotypic responses to exposure to pathogens, etc. (Figure 1). Of course, one challenge is simply to be able to describe the current level of phenotypic resistance of a host individual. Adding the spatial pattern of host genotypes provides additional modeling challenges. The level of aggregation of susceptible hosts will produce a particular "genotype unit area" (Mundt and Leonard, 1986), or area occupied by a single host genotype, and help to determine the extent to which microbial populations are spread between host species/genotypes or tend to be dispersed within host species/genotypes. This pattern of host genotypes will also influence the pattern of expression in response to microbes associated with a particular host genotype. The combination of host genotype spatial patterns and the spatial pattern of the biotic and abiotic environment will result in a host "phenotype mixture". Just as the effects of genotype diversity vary for different host–pathogen systems (Lannou et al., 1994; Mundt et al., 1995; Ngugi et al., 2001; Mitchell et al., 2002; Cox et al., 2004), the complex communities of plants and microbes involved in induced resistance may experience different patterns of spatial effects on induction. Models of epidemics in genotype mixtures will be useful in this context, but new models will also be needed.

Conclusion

Epidemiology will benefit from new genomic technologies in several ways. New diagnostic techniques will make the development of a "community epidemiology" more practical, through providing the ability to characterize thousands of microorganisms simultaneously as well as identifying particular genes and alleles. New techniques will make it easier to extend genetic analyses of pathogens beyond virulence genes, by facilitating the study of the population structure and evolution of genes important for other important features such as the ability to survive in non-conducive environments. Functional genomic analysis of pathogen virulence genes and host resistance and defense response genes will enable better predictions of the durability of resistance. New genomic tools will also allow great advances in the study of phenotypic resistance. It will finally be possible to thoroughly evaluate the many ideas put forward about age-related resistance and the effects of the biotic and abiotic environment on phenotypic resistance. Conversely, epidemiology provides the context for understanding the role and significance of pathogen genes and plant genes related to pathogen reproduction and also provides models for evaluating landscapes of plant phenotypes.

Acknowledgements

Thanks to J. Bai, R. L. Bowden, A. H. C. van Bruggen, members of the KSU Ecological Genomics Community, and members of the KSU Plant Disease Ecology Lab for helpful discussions of these ideas, to EJPP reviewers for helpful comments, and to S. P. Dendy for assistance preparing this manuscript. This work was supported by the U.S. National Science Foundation under Grant Nos. DEB-0130692 and DEB-0516046, by the Ecological Genomics Initiative of Kansas through NSF Grant No. EPS-0236913 with matching funds from the Kansas Technology Enterprise Corporation, by the NSF Long Term Ecological Research Program at Konza Prairie, by the U.S. Department of Energy Office of Science (Program in Ecosystem Research) Grant No. DE-FG02-04ER63892, by the US Smithsonian Tropical Research Institute (Center for Tropical Forest Science), and by USDA Grant No. 2002-34103-11746. This is Kansas State Experiment Station Contribution No. 06-22-J.

References

Adee SR, Pfender WF and Hartnett DC (1990) Competition between *Pyrenophora tritici-repentis* and *Septoria nodorum* in the wheat leaf as measured with de Wit replacement series. Phytopathology 80: 1177–1182

Alfano JR and Collmer A (2004) Type III secretion system effector proteins: double agents in bacterial disease and plant defense. Annual Review of Phytopathology 42: 385–414

Al-Naimi FA, Garrett KA and Bockus WW (2005) Competition, facilitation, and niche differentiation in two foliar pathogens. Oecologia 143: 449–457

Antonovics J (2003) Toward community genomics? Ecology 84: 598–601

Axtell MJ and Staskawicz BJ (2003) Initiation of RPS2-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4. Cell 112: 369–377

Bai J, Pennill LA, Ning J, Lee SW, Ramalingam J, Webb CA, Zhao B, Sun Q, Nelson JC, Leach JE and Hulbert SH (2002) Diversity of nucleotide binding site-leucine-rich repeat genes in cereals. Genome Research 12: 1871–1884

Bergelson J, Kreitman M, Stahl EA and Tian DC (2001) Evolutionary dynamics of plant R-genes. Science 292: 2281–2285

Black WC IV, Baer CF, Antolin MF and DuTeau NM (2001) Population genomics: genome-wide sampling of insect populations. Annual Review of Entomology 46: 441–469

Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Topsch S, Vos P, Salamini F and Schulze-Lefert P (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell 88: 695–705

Caicedo AL, Schaal BA and Kunkel BN (1999) Diversity and molecular evolution of the RPS2 resistance gene in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America 96: 302–306

Calonnec A, Goyeau H and de Vallavielle-Pope C (1996) Effects of induced resistance on infection efficiency and sporulation of *Puccinia striiformis* on seedlings in varietal mixtures and on field epidemics in pure stands. European Journal of Plant Pathology 102: 733–741

Chain PSG, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, Georgescu AM, Vergez LM, Land ML, Motin VL, Brubaker RR, Fowler J, Hinnebusch J, Marceau M, Medigue C, Simonet M, Chenal-Francisque V, Souza B, Dacheux D, Elliott JM, Derbise A, Hauser LJ and Garcia E (2004) Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. Proceedings of the National Academy of Sciences of the United States of America 101: 13826–13831

Chang JH, Urbach JM, Law TF, Arnold LW, Hu A, Gombar S, Grant SR, Ausubel FM and Dangl JL (2005) A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae*. Proceedings of the National Academy of Sciences of the United States of America 102: 2549–2554

Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA, Budworth PR, Tao Y, Xie Z, Chen X, Lam S, Kreps JA, Harper JF, Si-Ammour A, Mauch-Mani B, Heinlein M, Kobayashi K, Hohn T, Dangl JL, Wang X and Zhu T (2002) Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. The Plant Cell 14: 559–574

Collmer A, Lindeberg M, Petnicki-Ocwieja T, Schneider DJ and Alfano JR (2002) Genomic mining type III secretion system effectors in *Pseudomonas syringae* yields new picks for all TTSS prospectors. Trends in Microbiology 10: 462–469

Cortesi P, Fischer M and Milgroom MG (2000) Identification and spread of *Fomitiporia punctata* associated with wood decay of grapevine showing symptoms of Esca. Phytopathology 90: 967–972

Cortesi P and Milgroom MG (2001) Outcrossing and diversity of vegetative compatibility types in populations of *Eutypa lata* from grapevines. Journal of Plant Pathology 83: 79–86

Cowger C, Wallace LD and Mundt CC (2005) Velocity of spread of wheat stripe rust epidemics. Phytopathology 95: 972–982

Cox CM, Garrett KA, Bowden RL, Fritz AK, Dendy SP and Heer WF (2004) Cultivar mixtures for the simultaneous management of multiple diseases: Tan spot and leaf rust of wheat. Phytopathology 94: 961–969

Cui J, Bahrami AK, Pringle EG, Hernandez-Guzman G, Bender CL, Pierce NE and Ausubel FM (2005) *Pseudomonas syringae* manipulates systemic plant defenses against pathogens and herbivores. Proceedings of the National Academy of Sciences of the United States of America 102: 1791–1796

Deslandes L, Olivier J, Peeters N, Feng DX, Khounlotham M, Boucher C, Somssich I, Genin S and Marco Y (2003) Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. Proceedings of the National Academy of Sciences of the United States of America 100: 8024–8029

Durrant WE and Dong X (2004) Systemic acquired resistance. Annual Review of Phytopathology 42: 185–209

Ellis JG, Lawrence GJ, Luck JE and Dodds PN (1999) Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. The Plant Cell 11: 495–506

Escriu F, Fraile A and Garcia-Arenal F (2000a) Evolution of virulence in natural populations of the satellite RNA of Cucumber mosaic virus. Phytopathology 90: 480–485

Escriu F, Fraile A and Garcia-Arenal F (2003) The evolution of virulence in a plant virus. Evolution 57: 755–765

Escriu F, Perry KL and Garcia-Arenal F (2000b) Transmissibility of Cucumber mosaic virus by *Aphis gossypii* correlates with viral accumulation and is affected by the presence of its satellite RNA. Phytopathology 90: 1068–1072

Feder ME and Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. Nature Reviews Genetics 4: 651–665

Frick UB and Schaller A (2002) cDNA microarray analysis of fusicoccin-induced changes in gene expression in tomato plants. Planta 216: 83–94

Garrett KA and Bowden RL (2002) An Allee effect reduces the invasive potential of *Tilletia indica*. Phytopathology 92: 1152–1159

Garrett KA and Mundt CC (1999) Epidemiology in mixed host populations. Phytopathology 89: 984–990

Gilbert GS (2002) Evolutionary ecology of plant diseases in natural ecosystems. Annual Review of Phytopathology 40: 13–43

Glazebrook J, Chen W, Estes B, Chang HS, Nawrath C, Métraux JP, Zhu T and Katagiri F (2003) Topology of the network integrating salicylate and jasmonate signal transduction derived from global expression phenotyping. The Plant Journal 34: 217–228

Hughes G (1996) Incorporating spatial pattern of harmful organisms into crop loss models. Crop Protection 15: 407–421

Hulbert S (1998) Arming plants against pathogens-recombination in resistance loci. Trends in Plant Science 3: 121–122

Hulbert SH, Webb CA, Smith SM and Sun Q (2001) Resistance gene complexes: evolution and utilization. Annual Review of Phytopathology 39: 285–312

Iavicoli A, Boutet E, Buchala A and Métraux JP (2003) Induced systemic resistance in *Arabidopsis thaliana* in response to root inoculation with *Pseudomonas fluorescens* CHA0. Molecular Plant–Microbe Interactions 16: 851–858

Jeger MJ (2004) Analysis of disease progress as a basis for evaluating disease management practices. Annual Review of Phytopathology 42: 61–82

Jia Y, Bryan GT, Farrall L and Valent B (2003) Natural variation at the *Pi-ta* rice blast resistance locus. Phytopathology 93: 1452–1459

Jia Y, McAdams SA, Bryan GT, Hershey HP and Valent B (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. The European Molecular Biology Organization Journal 19: 4004–4014

Johnson R (1981) Durable resistance: definition of, genetic control, and attainment in plant breeding. Phytopathology 71: 567–568

Jones PG, Allaway D, Gilmour DM, Harris C, Rankin D, Retzel ER and Jones CA (2002) Gene discovery and microarray analysis of cacao (*Theobroma cacao* L.) varieties. Planta 216: 255–264

Kerk NM, Ceserani T, Tausta SL, Sussex IM and Nelson TM (2003) Laser capture microdissection of cells from plant tissues. Plant Physiology 132: 27–35

Kleczkowski A, Gilligan CA and Bailey DJ (1997) Scaling and spatial dynamics in plant–pathogen systems: from individuals to populations. Proceedings of the Royal Society of London B 264: 979–984

Kranz J, (2002) Comparative Epidemiology of Plant Diseases. Springer

Lannou C, de Vallavieille-Pope C, Biass C and Goyeau H (1994) The efficacy of mixtures of susceptible and resistant hosts to two wheat rusts of different lesion size: controlled condition experiments and computerized simulations. Journal of Phytopathology 140: 227–237

Leach JE, Vera Cruz CM, Bai J and Leung H (2001) Pathogen fitness penalty as a predictor of durability of disease resistance genes. Annual Review of Phytopathology 39: 187–224

Leung H, Zhu Y, Revilla-Molina I, Fan JX, Chen H, Pangga I, Vera Cruz C and Mew TW (2003) Using genetic diversity to achieve sustainable rice disease management. Plant Disease 87: 1156–1169

Liu B, Zhang SH, Zhu XY, Yang Q, Wu S, Mei M, Mauleon R, Leach J, Mew T and Leung H (2004) Candidate defense genes as predictors of quantitative blast resistance in rice. Molecular Plant–Microbe Interactions 17: 1146–1152

Luck JE, Lawrence GL, Dodds PN, Shepherd KW and Ellis JG (2000) Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. The Plant Cell 12: 1367–1377

Mackey D, Belkhadir Y, Alonso JM, Ecker JR and Dangl JL (2003) *Arabidopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. Cell 112: 379–389

Madden LV, Hughes G and Irwin ME (2000) Coupling disease-progress-curve and time-of-infection functions for predicting yield loss of crops. Phytopathology 90: 788–800

Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton KA, Dangl JL and Dietrich RA (2000) The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. Nature Genetics 26: 403–410

Mauricio R (2005) Can ecology help genomics: the genome as ecosystem?. Genetica 123: 205–209

Meyers BC, Kozik A, Griego A, Kuang H and Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. The Plant Cell 15: 809–834

Milgroom MG and Peever TL (2003) Population biology of plant pathogens: the synthesis of plant disease epidemiology and population genetics. Plant Disease 87: 608–617

Mitchell CE, Tilman D and Groth JV (2002) Effects of grassland plant species diversity, abundance, and composition on foliar fungal disease. Ecology 83: 1713–1726

Mundt CC, Brophy LS and Schmitt MS (1995) Disease severity and yield of pure-line wheat cultivars and mixtures in the presence of eyespot, yellow rust, and their combination. Plant Pathology 44: 173–182

Mundt CC and Leonard KJ (1986) Effect of host genotype unit area on development of focal epidemics of bean rust and common maize rust in mixtures of resistant and susceptible plants. Phytopathology 76: 895–900

Mundt CC (2002) Use of multiline cultivars and cultivar mixtures for disease management. Annual Review of Phytopathology 40: 381–410

Mysore KS, D'Ascenzo MD, He X and Martin GB (2003) Overexpression of the disease resistance gene Pto in tomato induces gene expression changes similar to immune responses in human and fruitfly. Plant Physiology 132: 1901–1912

Nakazono M, Qiu F, Borsuk LA and Schnable PS (2003) Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. The Plant Cell 15: 583–596

Newcombe G and Bradshaw HD (1996) Quantitative trait loci conferring resistance in hybrid poplar to *Septoria populicola*, the cause of leaf spot. Canadian Journal of Forest Research 26: 1943–1950

Ngugi HK, King SB, Holt J and Julian AM (2001) Simultaneous temporal progress of sorghum anthracnose and leaf blight in crop mixtures with disparate patterns. Phytopathology 91: 720–729

Otten W, Filipe JAN, Bailey DJ and Gilligan CA (2003) Quantification and analysis of transmission rates for soil-borne epidemics. Ecology 84: 3232–3239

Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BBH and Jones JDG (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell 91: 821–832

Pieterse CMJ and van Loon LC (1999) Salicylic acid-independent plant defence pathways. Trends in Plant Science 4: 52–58.

Ponciano G, Ishihara H, Tsuyumu S and Leach JE (2003) Bacterial effectors in plant disease and defense: keys to durable resistance? Plant Disease 87: 1272–1282

Ponciano G, Webb K, Bai J, Vera Cruz C and Leach JE (2004) Molecular characterization of the *avrXa7* locus from *Xanthomonas oryzae* pv. *oryzae* field isolates. Physiological and Molecular Plant Pathology 64: 145–153

Robinson RA (1976) Plant Pathosystems. Springer-Verlag, New York

Rohmer L, Guttman DS and Dangl JL (2004) Diverse evolutionary mechanisms shape the type iii effector virulence factor repertoire in the plant pathogen *Pseudomonas syringae*. Genetics 167: 1341–1360

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J and Goodman RM (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Applied and Environmental Microbiology 66: 2541–2547

Ryals JA, Neuenschwander UH, Willits MG, Molina A, Steiner H-Y and Hunt MD (1996) Systemic acquired resistance. The Plant Cell 8: 1809–1819

Schenk PM, Kazan K, Manners JM, Anderson JP, Simpson RS, Wilson IW, Somerville SC and Maclean DJ (2003) Systemic gene expression in *Arabidopsis* during an incompatible interaction with *Alternaria brassicicola*. Plant Physiology 132: 999–1010

Scherm H (1996) On the velocity of epidemic waves in model plant disease epidemics. Ecological Modelling 87: 217–222

Scofield SR, Tobias CM, Rathjen JP, Chang JH, Lavelle DT, Michelmore RW and Staskawicz BJ (1996) Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. Science 274: 2063–2065

Smith KP and Goodman RM (1999) Host variation for interactions with beneficial plant-associated microbes. Annual Review of Phytopathology 37: 473–491

Smith SM, Pryor AJ and Hulbert SH (2004) Allelic and haplotypic diversity at the *Rp1* rust resistance locus of maize. Genetics 167: 1939–1947

Spoel SH, Koornneef A, Claessens SMC, Korzelius JP, Van Pelt JA, Mueller MJ, Buchala AJ, Métraux JP, Brown R, Kazan K, Van Loon LC, Dong X and Pieterse CMJ (2003) NPR1 modulates cross-talk between salicylate- and jasmonate-dependent defense pathways through a novel function in the cytosol. The Plant Cell 15: 760–770

Tang X, Frederick RD, Zhou J, Halterman DA, Jia Y and Martin GB (1996) Initiation of plant disease resistance by physical interaction of *AvrPto* and *Pto* kinase. Science 274: 2060–2063

Thomma BPHJ, Eggermont K, Penninckx IAWA, Mauch-Mani B, Vogelsang R, Cammue BPA and Broekaert WF (1998) Separate jasmonate-dependent and salicylate-dependent defense-response pathways in *Arabidopsis* are essential for resistance to distinct microbial pathogens. Proceedings of the National Academy of Sciences of the United States of America 95: 15107–15111

Tian D, Araki H, Stahl E, Bergelson J and Kreitman M (2002) Signature of balancing selection in *Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America 99: 11525–11530

VanSluys MA, Monteiro-Vitorello CB, Camargo LEA, Menck CFM, da Silva ACR, Ferro JA, Oliveira MC, Setubal JC, Kitajima JP and Simpson AJ (2002) Comparative genomic analysis of plant-associated bacteria. Annual Review of Phytopathology 40: 169–189

van Wees SCM, de Swart EAM, van Pelt JA, van Loon LC and Pieterse CMJ (2000) Enhancement of induced disease resistance by simultaneous activation of salicylate- and jasmonate-dependent defense pathways in *Arabidopsis thaliana*. Proceedings of the National Academy of Science of the United States of America 97: 8711–8716

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH and Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304: 66–74

Vera Cruz CM, Bai J, Oña I, Leung H, Nelson RJ, Mew TW and Leach JE (2000) Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. Proceedings of the National Academy of Sciences of the United States of America 97: 13500–13505

vonBodman SB, Bauer WD and Coplin DL (2003) Quorum sensing in plant-pathogenic bacteria. Annual Review of Phytopathology 41: 455–482

Waggoner PE and Aylor DE (2000) Epidemiology: a science of patterns. Annual Review of Phytopathology 38: 71–94

Wan J, Dunning MF and Bent AF (2002) Probing plant–pathogen interactions and downstream defense signaling using DNA microarrays. Functional and Integrative Genomics 2: 259–273

Welch SM, Roe JL, Das S, Dong Z, He R and Kirkham MB (2005) Merging genomic control networks and soil-plant-atmosphere-continuum models. Agricultural Systems 86: 243–274

Welch SM, Roe JL and Dong Z (2003) A Genetic neural network model of flowering time control in *Arabidopsis thaliana*. Agronomy Journal 95: 71–81

Weller DM, Raaijmakers JM, McSpadden-Gardener BB and Thomashow LS (2002) Microbial populations responsible for specific soil suppressiveness to plant pathogens. Annual Review of Phytopathology 40: 309–348

Whitham SA, Quan S, Chang HS, Cooper B, Estes B, Zhu T, Wang X and Hou YM (2003a) Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. The Plant Journal 33: 271–283

Whitham TG, Martinsen GD, Floate KD, Dungey HS, Potts BM and Keim P (1999) Plant hybrid zones affect biodiversity: tools for a genetic-based understanding of community structure. Ecology 80: 416–428

Whitham TG, Young WP, Martinsen GD, Gehring CA, Schweitzer JA, Shuster SM, Wimp GM, Fischer DG, Bailey JK, Lindroth RL, Woolbright S and Kuske CR (2003b) Community genetics: a consequence of the extended phenotype. Ecology 84: 559–573

With KA (2002) The landscape ecology of invasive spread. Conservation Biology 16: 1192–1203

Wolter M, Hollricher K, Salamini F and Schulze-Lefert P (1993) The *mlo* resistance alleles to powdery mildew infection in barley trigger a developmentally controlled

defense mimic phenotype. Molecular and General Genetics 239: 122–128

Zadoks JC and Vandenbosch F (1994) On the spread of plant disease – a theory on foci. Annual Review of Phytopathology 32: 503–521

Zhu Y, Chen H, Fan J, Wang Y, Li Y, Chen J, Fan J, Yang S, Hu L, Leaung H, Mew CW, Teng PS, Wang Z and Mundt CC (2000) Genetic diversity and disease control in rice. Nature 406: 718–722

Zwankhuizen MJ, Govers F and Zadoks JC (1998) Development of potato late blight epidemics: disease foci, disease gradients, and infection sources. Phytopathology 88: 754–763

European Journal of Plant Pathology (2006) 115:53–59
DOI 10.1007/s10658-005-2025-y

The practical considerations of scale in plant pathology

W.W. Turechek
USDA-ARS Fruit Laboratory, Building 010A – BARC-W, Baltimore Avenue, 10300 Beltsville, MD 20705-2350, USA (E-mail: turechekw@ba.ars.usda.gov)

Accepted 28 June 2005

Key words: quantitative epidemiology, strawberry

Abstract

The concept of scale has only recently gained recognition as a central theme in ecology. The rise in significance of scale in ecology can be attributed to the increase in hypothesis-driven experimental ecology over the last quarter century, and the realization that experimental results do not sufficiently explain past, or predict future observations in nature. Plant pathologists, who rely heavily on hypothesis-driven research, have confronted these same issues for nearly a century. In this paper, I will provide a concise presentation and discussion of the important concepts of scale and how they apply to the discipline of plant pathology.

Introduction

The concept of scale has only recently gained recognition as a central or unifying theme in ecology. In an extensive review of the ecological literature, Schneider (2001a) showed a dramatic increase in publications of scale-related research during the 1990s. The rise in importance of scale can be attributed to the increase in hypothesis-driven, experimental ecology over the last quarter century and the realization that experimental results do not sufficiently explain past or predict subsequent observations in nature.

Plant pathologists have confronted these same issues for nearly a century. With few exceptions, this important component of systems, observation, and analysis has been conveniently – unknowingly is perhaps a better choice of words – disregarded, sometimes to the detriment of the hypothesis. Unfortunately, the common leap from laboratory/glasshouse to the field without sufficient consideration of how outcomes may differ vastly when rescaled is a theme repeated to this day. Scaling is generally of little concern to the organismal biologist where processes under study are often clearly defined by the organism's size.

In what follows, I will provide a concise presentation and discussion of the important concepts of scale. Most of what I will present is drawn from the ecological literature, as ecologists are generally at the forefront of advancing concepts and our knowledge of scale. Three books that have helped shape my understanding of scale are 'Quantitative Ecology: Spatial and Temporal Scaling' by Schneider (1994), 'Ecological Scale: Theory and Applications' edited by Peterson and Parker (1998), and 'Scaling Relations in Experimental Ecology' edited by Gardner et al. (2001). I recommend these books to anyone wishing to gain an ecological perspective on scale.

Definitions

"Scale has a good start on contesting niche as one of the vaguest yet most often used words in ecology" (Wiens, 2001). Scale can be defined *correctly* in a number of ways. The definition that likely comes to mind when used in everyday conversation is that of cartographic scale. Cartographic scale is the ratio of the distance on a map to the distance on the ground (Schneider, 2001a). Another common usage defines scale as the

"...physical dimensions of observed entities and phenomena" (O'Neill and King, 1998). Merriam and Webster offer this definition of scale "to arrange in a graduated series" (Merriam-Webster online dictionary). In practice, most ecologists would argue against this latter definition because, unlike the previous two, the notion of quantifiable measurement (i.e., distance, dimension) is not stated explicitly. A number of ecologists have argued recently, and I agree, that any definition of scale must consider scale as a quantity and involve, or at least imply, measurements or measurement units (O'Neill and King, 1998).

Hierarchical scale equates the organizational level in a hierarchy to independent or individual scales. Hierarchical scale forms the essence of hierarchy theory (Allen and Starr, 1982) where it is argued that ecosystem processes operate in a way such that upper level processes, structure, etc. regulate and/or constrain processes at lower scales in a quantifiable manner. Some ecologists have argued that hierarchical levels should not be thought of as scales because 'level' is often an arbitrary, ambiguous or unquantifiable term (O'Neill and King, 1998; Wiens, 2001). Again, the notion of quantity is not defined explicitly in the definition, e.g., in what units does one quantify the 'leaf' level? (Turechek and Madden, 2001).

An idealistic goal in experimental design is to conduct an experiment where measurements or observations are taken at the organisms or phenomenons 'characteristic scale'. This is defined as the system scale at which all relevant ecological and biological processes of a population or community occur. This ideal, however, is likely not achievable as biological processes occur and interact over a range of scales. A better way to approach this concept is to envisage ecological phenomena occurring within upper and lower limits (Schneider, 1994). Even still, this may be easily definable for certain phenomena, such as spore dispersal, but not others, e.g., patterns of plant disease.

Scaling concepts

Given an acceptable definition, a logical next question is how to apply the term scale in practice. In other words, what exactly needs to be scaled? Dungan et al. (2002) distinguished three categories to which scale-related terms are applicable. The first category is the phenomenon (process or structure) under study. For example, the spatial pattern of plant disease and the processes that generate it. The second is the experimental or sampling units used to acquire information or data about the phenomenon under study. The third is the analyzes used to summarize the data to describe the phenomenon. From herein, I will refer to these as the system scale, observational scale, and analysis scale, respectively.

In all cases, scale is bounded by *grain* at the lowest extreme and *extent* at the other. When applied to the systems scale grain refers to the smallest ecological unit to which a phenomenon affects a structure or process; extent is the total area affected by the phenomenon, process, or structure. Applied to the observational scale grain refers to the *resolution* of measurements (Dungan et al., 2002). The frequency of sampling in space and/or time determines the resolution of the data. Extent refers to the range over which experimental measurements are taken (Wiens, 2001). For example, consider measurements of precipitation. This is an important variable in many disease forecasters and one where estimates often need to be sufficiently accurate. The recording frequency of the rain gauge (i.e., hourly, daily, weekly, etc.) sets the temporal resolution or grain. The duration of time over which precipitation is measured determines the extent. Applied to the analysis scale, calculating means and grouping sampling units coarsens grain; subsampling reduces the extent. These are a few examples of how analytical tools alter scale (O'Neill and King, 1998).

The *scope* is defined at the observational scale for experimental design, surveys, model building, etc. as the ratio of extent to grain, and is an important and useful measurement for comparing scales across studies. "Scope can be thought of as the number of steps, once we know the step size" (Schneider, 1994). For example, the temporal scope of an experiment is the ratio of the time from the beginning to the end of the experiment, to the time-step of a single measurement. The spatial scope of a survey is the ratio of the maximum length or distance between measurements relative to the minimum length or distance of a single measurement.

Disease surveys are used commonly in plant pathology to estimate some unknown value or quantity of a population, typically disease severity

or incidence (Turechek and Mahaffee, 2004). In the terminology of survey sampling, the scope of a survey is the ratio of the sampling frame to the sampling unit, where the sampling unit is the smallest item sampled and the sampling frame is the total number of possible sampling units in the survey area. It is always the case that the precision of the estimate increases, and the uncertainly decreases, as the number of samples or observations is increased. Thus, the challenge is to collect an ample number of samples to sufficiently minimize uncertainty under the logistical and natural constraints that often limit the scope of most designed surveys.

For example, a typical raised-bed field of strawberry will contain about 43,750 plants ha^{-1}. An individual strawberry plant could be considered the sampling unit in a sampling frame of 43,750 plants. In a survey for Colletotrichum crown rot (caused by *C. acutatum*) a sample of 500 plants gives a sampling fraction (i.e., the ratio of the number of samples taken to the sampling frame) of 500/43750 or 1.14%. The inverse of the sampling fraction represents the magnification factor (MF). The magnification factor magnifies the result of the sample into an estimate for the entire population. For example, say 20% of the sample (100 plants) was diagnosed with crown rot. Multiplying the number of diseased plants by the MF (i.e., 87.7) informs us that we should expect to find 8770 plants with crown rot in the population, assuming a random distribution of infected plants. The magnification factor can be reduced by sampling more plants, but at a cost. Cluster sampling allows the sampler to observe a greater number of plants with the same number of sampling units (Hughes et al., 1996). This, in effect, will reduce the MF at the cost of decreasing resolution and, consequently, will give a reduced scope. The benefit of the trade-off needs to be determined for each study and can be represented in a scope diagram.

A scope diagram is one way of displaying or comparing the scope among different phenomena, events, or studies. Schneider (1994) demonstrates how a simple line diagram can be used to represent the scope of a survey. The line can be partitioned into two components, one representing the scope of the sample and the remainder the inference component of the sample represented by MF. This simple diagram can be used to compare survey strategies and help decide the best approach. Continuing with the strawberry example, assume that in a hectare of strawberries a single plant occupies a space of 0.1 m^2. Figure 1 depicts the scope diagrams for three 100 unit samples where: (1) single plants are observed or collected in a random sample, (2) groups of 10 plants are collected in a cluster sample, and (3) 10 m^2 grids are observed via aerial sampling. In this example, the number of samples collected or observed remains the same, but the size of the sampling unit changes resulting in a smaller magnification factor and a

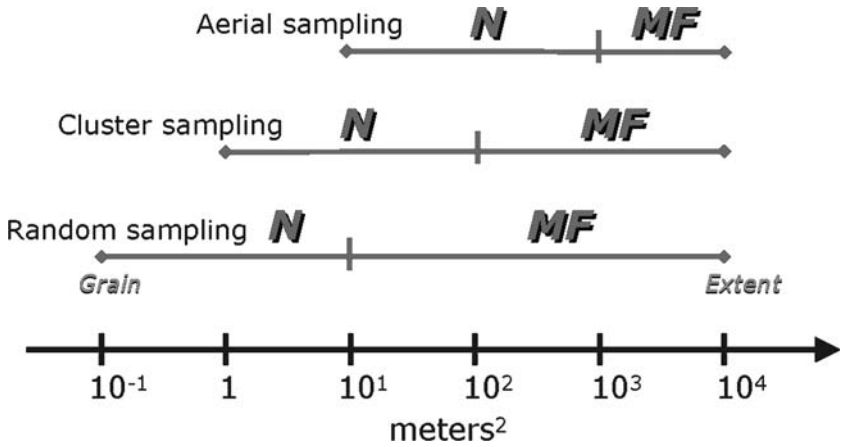


Figure 1. Scope diagrams for three 100 unit samples where: (1) single plants are observed or collected in a random sample, (2) groups of 10 plants are collected in a cluster sample, and (3) 10 m^2 grids are observed via aerial sampling. The starting point of each diagram represents the grain or resolution of the data at the observational scale, and the ending point represents the extent of the study (the sampling frame). The sample size is denoted by *N* and the magnification factor by MF.

reduced scope. Scope diagrams can be much more complex. Seem (2004) gives a good example of scope diagram and its applicability in disease forecasting.

Scaling in practice

The practice of *scaling* involves relating measurements made at one scale to measurements or predictions made at another. A scale-dependent process is one in which the process (e.g., rate) or property (e.g., density) changes with a change of grain or extent. An example of scale dependency encountered often in plant pathology is scale-dependent patterns of plant disease, i.e., patterns differing with sampling unit size and/or extent of the survey. Although these scaling dependencies can be represented through various modelling procedures (Allen, 2001), episodic dynamics, non-Euclidean structures, and/or biotic discontinuities that are typically associated with ecological processes or phenomena make modelling (scaling) a challenge.

Episodic dynamics generally make it difficult to scale temporal events. For example, 30 mm of rainfall over the course of an hour implies a rainfall rate of 0.5 mm min^{-1}; an increase in disease severity from 3 lesions/leaf to 52 lesions/leaf over the period of 7 days implies a rate of disease progress of 7 lesions day^{-1}. The process of calculating rates masks the fact that precipitation and lesion development occur as concentrated events (episodes) over some defined time period. Non-Euclidean structures (such as landscape surfaces) and patchy or heterogeneous environments generally make it difficult to scale spatial properties because they lead to fractal dimensions (i.e., greater than a plane, but less than a volume). For example, wind patterns contributing to larger scale patterns of spore dispersal and microclimatic variability are directly affected by variations in topography in a largely unpredictable manner, much more so than if these events were to occur over a strictly two-dimensional surface.

Biotic boundaries or discontinuities further contribute to modelling headaches. For example, a change of 3 °C from 27 to 30 °C straddles the critical temperature for conidial germination of *Podosphaera macularis*, the causal agent of hop powdery mildew (Mahaffee et al., 2003). Other biotic boundaries are often much more difficult to define and are not necessarily constrained by quantifiable units (Naeem, 2001). For example, the taxonomic groupings of race, subspecies, species, genus, and sometimes higher groupings can often be ambiguous. At the population level, the boundaries between patches, communities, metapopulations, etc. is often blurred due to fluctuations in growth, decline, and interactions among populations.

Considerations in experimental design

"The near absence of prescriptions for incorporating scale in experimental design may partially explain why explicit consideration of scale is not more prevalent in the design of terrestrial field experiments" (King et al., 2001). Readers interested in a statistical treatment of this topic are referred to two chapters by Dutilleul (1998a, b). I will touch on a broader aspect of design: the desire to have experimental results relate directly to natural observation. Every experiment is subject to a compromise between what Manly (1992) defines as internal vs. external validity (Naeem, 2001). "Internal validity concerns whether the apparent effects or lack of effects shown by the experimental results are due to the factor being studied, rather than some alternative factor. External validity concerns the extent to which the results of an experiment can be generalized to some wider population of interest." Naeem (2001) groups experiments into three general classes: (1) field, (2) model-ecosystem (micro-, meso-, and macrocosm), and (3) simulation. Field experiments have the highest level of external validity and, consequently, the lowest level of internal validity. Conversely, simulation experiments have the highest level of internal validity and lowest level of external validity.

More often than not, the type of experiment and the choice of scale (determined by plot size, duration, sampling extent, etc.) are a function of pragmatism. Available funding, personnel, experimental costs, measurement technology, etc. play a more central role in the design of experiments than does the theoretical consideration of scale and validity. In any event, it is likely (and is often the case) that to gain a full understanding of a process, sets of experiments that span the range of what Naeem (2001) refers to as the "scale-validity matrix" must be conducted. That is, a reasonable set

of field, microcosm, or simulation experiments conducted at different spatial, temporal, and possibly biotic scales are often necessary to fully interpret a process or to explain what was observed naturally.

Multiscale analysis

In the (failed) quest for a 'characteristic scale', multiscale analysis has evolved to play a central role in scaling. Multiscale analysis is defined as an analysis with respect to multiples of a unit of measurement (Schneider, 1994). In general, this type of analysis is done by first defining subsystems within a system. In a survey, this might be accomplished by superimposing a grid over the region of study or spatially referencing sampling units in the survey area. As one example, summing the components of the subsystem, grid, or sampling units (with correction factors introduced as needed) can be used to scale to larger areas (Schneider, 2001b). Quantities can be summed by either juxtaposing or superposing values. Summation by juxtaposing values extends the range of scale; summation by superposing leaves the scale unchanged. Summing the number of diseased plants in a series of contiguous plots is an example of juxtaposing; summing the number of newly diseased plants to the number of previously diseased plants in a single plot is an example of superposing.

The variance plays a central role in multiscale analysis. ''One of the major research challenges in ecology is understanding the creation and erosion of spatial variability as a function of spatial scale. Included in this challenge is the question of the degree to which variance generated at one scale is transformed into variance at another scale'' (Schneider, 1994). Across many disciplines, including plant pathology, methods have been developed for relating and/or predicting variance across scales (Hughes et al., 1997; Turechek and Madden, 2003). However, the mere ability to model these relations should not be mistaken as an understanding of how these relations came to be. For the most part, the mechanisms or biological processes generating these differences are only partially understood.

The sample variance is only one measure of spatial variability and has limited interpretation in multiscale analysis. Variances can also be calculated from grouped or lagged measurements or observations. (The term *lag* refers to the interval or spacing between neighbouring units.) Imagining a grid; grouping occurs when contiguous squares of the grid (sampling units) are combined to form larger units and the quantities are combined via juxtaposition (added); under these conditions the resolution of the data changes. Variances are obtained by re-calculating the variance of the combined quantities, and comparing it to the original or ungrouped variance and to variances calculated from successively larger groupings. The blocked quadrat-variance methods, such as the two-term local quadrat variance (TTLQV) method and the paired-quadrat variance (PQV) method, are examples of analyses that use grouped variances (Ludwig and Reynolds, 1988). Lagging, on the other hand, results from calculating deviances between grid components (sampling units) at increasingly greater separations (lags) across the grid. Again, these variances are compared to the original sample variance as well as to variances calculated at different lag distances. Autocorrelation and semivariogram analyses are examples of analyses that use lagged variances (Cressie, 1991). Variances calculated according to the lag manoeuvre can be used to calculate variances that would be obtained via grouping using a Fourier transformation (Schneider, 1994).

Scaled quantities cannot be treated as unitless numbers. The process of summing, multiplying, and taking derivatives of scaled quantities should not be done independently of the unit. For example, the sum of 52 lesions/leaf and 6 diseased trees/orchard is non-sensical. Biologically interpretable sums of scaled quantities are referred to as ensemble quantities (Schneider, 1994). Spatial and temporal averages, variances, and covariances are typical ensemble quantities. This definition differs from the traditional in which an ensemble refers to a collection of 'realizations' of an event or process; the ensemble average, for example, is the mean of the realizations. Although this concept is evident in plant pathology research, the terminology is infrequently used (one exception is Ferrandino, 2004).

Statistical tools

Over the past 20 years, the variety of statistical tools available for multiscale analyses has

increased tremendously; many have been applied to characterize spatial patterns of plant disease. The tests, however, can be categorized based on the general type of analysis, point-pattern vs. correlation (Upton and Fingleton, 1985), or on whether the data consist of mapped or unmapped observations (Diggle, 1983). The point-pattern approach is based on the location of individuals over an area of interest and analyses are conducted either using the distances between individuals (Perry, 1995; Ferrandino, 1998), or using the counts of individuals within sampling units such as quadrats (Pielou, 1977; Madden and Hughes, 1995). The latter methods include the distributional approach that involves fitting observed frequencies of counts per sampling unit to well-defined statistical distributions (e.g., Poisson, negative binomial, binomial, and beta-binomial). The methods based on counts per sampling unit provide direct measurements of heterogeneity of the data at the scale of the sampling units and below, but they do not explicitly define the spatial arrangement of the sampling-unit counts unless several sampling units are grouped in a series of steps (Ludwig and Reynolds, 1988).

Spatial autocorrelation and semivariograms (Cressie, 1991) use lagged variances to produce explicit information on the degree of association of disease intensity among sampling units. Unlike the distributional methods, these methods yield different results for different arrangements of counts within a field, although they are not specifically developed for counts within sampling units. Spatial Analysis by Distance IndicEs (SADIE) is a class of tests developed recently to quantify spatial pattern in the spirit of spatial autocorrelation, but uses data in the form of counts (Perry, 1995; Xu and Madden, 2004). The correlation-based methods characterize pattern at the scale of the sampling unit and above. The results from point-pattern and correlation-type analyses can jointly be used to better interpret patterns and possibly describe the biological phenomenon responsible for generating the observed pattern (Turechek and Madden, 1999).

Simulation and randomizations have also been used to study scale-related processes. For example, Turechek and Madden (2001) used Monte–Carlo methods and randomizations to determine how the variability of strawberry leaf blight at a lower scale impacted the variability at higher scales. Willocquet and Savary (2004) designed a simulation model to determine how varying auto-, allo-leaf, and allo-plant-deposition rates of infective propagules affected epidemic development observed at the plant and leaf level (in both examples, measurement units are implied). As discussed above, simulation studies have a high degree of internal validity and allow researchers to explore a range of conditions that may otherwise take years to observe.

Conclusion

The intent of this paper was to provide an overview of scale-related concepts and how they might apply to plant pathology. Although I did not provide prescriptive advice on how to include scale in designed studies, I hope I made it obvious where scaling is naturally applied in our discipline. I also hope that I impressed the importance of being vigilant in reporting the scale (grain, extent, and scope) of experiments and surveys to allow for drawing valid comparisons across studies.

To summarize briefly, before designing any experiment or survey, it would be prudent to acquire preliminary data on the structure of the population under study so that sampling units (i.e., grain) and the extent of the study can be appropriately chosen (Legendre and Legendre, 1998). Realize that this information may indicate that a single study may not be sufficient to gain a full understanding of the process or characteristic under study. It is also important to consider the units of measurement and how easily information can be rescaled. Pathologists should avoid describing the scale of study as, for example, the 'leaf scale' unless a unit of measurement is clearly implied. In many cases, it is likely that the scale can be defined within some narrow range of values, and these should be used to identify the scale (grain) of study. Lastly, be aware that many field studies or surveys of disease are the result of the interaction between two populations: the population of the host and the population of the pathogen. The scale at which the host population exists should be an important consideration of the pathogen population, because the host represents the possible extent of the pathogen. For example, knowing the scale of spore dispersal distance is not very informative unless the distance significantly overlaps the range of the host. Considering these aspects of scale in the design of an

experiment will help to minimize the possibly large discrepancies in scale between what is observed and what is being tested.

References

Allen TFH (2001) The nature of the scaling issue in experimentation. In: Gardner RH, Kemp WM, Kennedy VS and Peterson JE (eds) Scaling Relations in Experimental Ecology. Columbia University Press, New York

Allen TFH and Starr TB (1982) Hierarchy, University of Chicago Press, Chicago

Cressie N (1991) Statistics for Spatial Data, Wiley, New York

Diggle PJ (1983) Statistical Analysis of Spatial Point Patterns, Academic Press, London

Dungan JI, Perry JN, Dale MRT, Legendre P, Citron-Pousty S, Fortin M-J, Jakomulska A, Miriti M and Rosenberg MS (2002) A balanced view of scale in spatial statistical analysis. Ecography 25: 626–640

Dutilleul P (1998a) Incorporating scale in ecological experiments: Study design. In: Peterson DL and Parker VT (eds) Ecological Scale: Theory and Applications. Columbia University Press, New York

Dutilleul P (1998b) Incorporating scale in ecological experiments: Data analysis. In: Peterson DL and Parker VT (eds) Ecological Scale: Theory and Applications. Columbia University Press, New York

Ferrandino FJ (1998) Past non-randomness and aggregation to spatial correlation: 2DCORR, a new approach for discrete data. Phytopathology 88: 84–91

Ferrandino FJ (2004) Measuring spatial aggregation in binary epidemics: Correlative analysis and the advantage of fractal-based sampling. Phytopathology 94: 1215–1227

Gardner RH, Kemp WM, Kennedy VS and Peterson JE (eds) (2001) *Scaling Relations in Experimental Ecology*. Columbia University Press, New York

Hughes G, Madden LV and Munkvold GP (1996) Cluster sampling for disease incidence data. Phytopathology 86: 132–137

Hughes G, McRoberts N, Madden LV and Gottwald TR (1997) Relationship between disease incidence at two levels in a spatial hierarchy. Phytopathology 87: 542–550

King AW, Gardner RH, Hatfield CA, Naeem S, Peterson JE and Wiens JA (2001) Terrestrial perspectives on issues of scale in experimental ecology. In: Gardner RH, Kemp WM, Kennedy VS and Peterson JE (eds) Scaling Relations in Experimental Ecology. Columbia University Press, New York

Legendre P and Legendre L (1998) Numerical Ecology, Elsevier Science, Amesterdam

Ludwig JA and Reynolds JF (1988) Statistical Ecology: A Primer on Methods and Momputing, Wiley-Interscience, New York

Madden LV and Hughes G (1995) Plant disease incidence: Distributions, heterogeneity, and temporal analysis. Annual Review of Phytopathology 33: 529–564

Mahaffee WF, Turechek WW and Ocamb CM (2003) Effect of variable temperature on infection severity of *Podosphaera macularis* on hops. Phytopathology 93: 1587–1592

Manly BFJ (1992) The Design and Analysis of Research Studies, Cambridge University Press, Cambridge

Naeem S (2001) Experimental validity and ecological scale as criteria for evaluating research programs. In: Gardner RH, Kemp WM, Kennedy VS and Peterson JE (eds) Scaling Relations in Experimental Ecology. Columbia University Press, New York

O'Neill RV and King AW (1998) Homage to St. Michael; or, why are there so many books on scale? . In: Peterson DL and Parker VT (eds) Ecological Scale: Theory and Applications. Columbia University Press, New York

Peterson DL and Parker VT (eds) (1998) Ecological Scale: Theory and Applications. Columbia University Press, New York

Perry JN (1995) Spatial analysis by distance indices. Journal of Animal Ecology 64: 303–314

Pielou EC (1977) Mathematical Ecology, Wiley, New York

Schneider DC (1994) Quantitative Ecology: Spatial and Temporal Scaling, Academic Press, San Diego

Schneider DC (2001a) The rise of the concept of scale in ecology. Bioscience 7: 545–553

Schneider DC (2001b) Spatial allometry: Theory and application to experimental and natural aquatic ecosystems. In: Gardner RH, Kemp WM, Kennedy VS and Peterson JE (eds) Scaling Relations in Experimental Ecology. Columbia University Press, New York

Seem RC (2004) Forecasting plant disease in a changing climate: A question of scale. Canadian Journal of Plant Pathology 26: 274–283

Turechek WW and Madden LV (1999) Spatial pattern analysis of strawberry leaf blight in perennial production systems. Phytopathology 89: 421–433

Turechek WW and Madden LV (2001) Effect of scale on plant disease incidence and heterogeneity in a spatial hierarchy. Ecological Modeling 144: 77–95

Turechek WW and Madden LV (2003) A generalized linear modeling approach for characterizing disease incidence in a spatial hierarchy. Phytopathology 93: 458–466

Turechek WW and Mahaffee WF (2004) Spatial pattern analysis of hop powdery mildew in the Pacific Northwest: Implications for sampling. Phytopathology 94: 1116–1128

Upton G and Fingleton B (1985) Spatial Data Analysis by Example, Vol 1. Point Pattern and Quantitative Data, John Wiley & Sons, West Sussex

Wiens JA (2001) Understanding the problem of scale in experimental ecology. In: Gardner RH, Kemp WM, Kennedy VS and Peterson JE (eds) Scaling Relations in Experimental Ecology. Columbia University Press, New York

Willocquet L and Savary S (2004) An epidemiological simulation model with three scales of a spatial hierarchy. Phytopathology 94: 883–891

Xu XM and Madden LV (2004) Use of SADIE statistics to study spatial dynamics of plant disease epidemics. Plant Pathology 53: 38–49

European Journal of Plant Pathology (2006) 115:61–73
DOI 10.1007/s10658-005-3682-6

Trends in theoretical plant epidemiology

H. Scherm*, H.K. Ngugi and P.S. Ojiambo
Department of Plant Pathology, University of Georgia, Athens, GA 30602, U.S.A
(Fax: +1-706-542-1262; E-mail: scherm@uga.edu)

Accepted 16 September 2005

Key words: decision analysis, mathematical model, population genetics, spatial structure, statistical epidemiology, stochasticity

Abstract

We review trends and advances in three specific areas of theoretical plant epidemiology: models of temporal and spatial dynamics of disease, the synergism of epidemiology and population genetics, and progress in statistical epidemiology. Recent analytical modelling of disease dynamics has focused on SIR (susceptible–infected–removed) models modified to include spatial structure, stochasticity, and multiple management-related parameters. Such models are now applied routinely to derive threshold criteria for pathogen invasion or persistence based on pathogen demographics (e.g., Allee effect or fitness of fungicide-resistant strains) and/or host spatial structure (e.g., host density or patch size and arrangement). Traditionally focused on the field level, the scale of analytical models has broadened to range from individual plants to landscapes and continents; however, epidemiological models for interactions at the cellular level, e.g., during the process of virus infection, are still rare. There is considerable interest in the concept of scaling, i.e., to what degree and how data and models from one scale can be transferred to another (smaller or larger) scale. Despite assertions to the contrary, the linkages between epidemiology and population genetics are alive and well as exemplified by recent efforts to integrate epidemiological parameters into population genetics models (and *vice versa*) and by numerous integrated studies with an applied focus (e.g., to quantify sources and types of primary and secondary inoculum). Statistical plant epidemiology continues to rely heavily on the medical and ecological fields for inspiration and conceptual advances, as illustrated by the recent surge in papers utilizing ROC (receiver operating characteristic), Bayesian, or survival analysis. Among these, Bayesian analysis should prove especially fruitful given the reliance on uncertain and subjective information for practical disease management. However, apart from merely adopting statistical tools from other disciplines, plant epidemiologists should be more proactive in exploring potential applications of their concepts and procedures in rapidly expanding disciplines such as statistical genetics or bioinformatics. Although providing the scientific basis for disease management will always be the *raison d'être* for plant epidemiology, a broader perspective will help the discipline to remain relevant as more resources are being devoted to genomic and ecosystem-level science.

Introduction

It is perhaps somewhat atypical that this commentary on theoretical plant epidemiology is authored by a group of investigators who consider themselves experimentalists rather than theoreticians. However, our role as dispassionate observers allows us to take a bird's-eye view of recent developments in the area and assess their impact on the science and practice of plant pathology without being influenced by predetermined notions.

The Encyclopaedia Britannica defines a scientific theory as a 'systematic ideational structure of

broad scope, conceived by the human imagination, that encompasses a family of empirical (experiential) laws regarding regularities existing in objects and events, both observed and posited' (Anon., 2005). In practice, this 'systematic ideational structure' is usually formalized as a model, either conceptual or mathematical. Both forms of models have been influential in plant epidemiology (Jeger, 2000; Zadoks, 2001). Although theory can be developed without mathematical models, the two concepts are often used synonymously in the epidemiological literature.

One of the most commonly voiced criticisms surrounding the use of theory and mathematical models in the broader ecological literature has been the lack of interaction between modellers and experimentalists during model development, testing, and validation (Caswell, 1988; Hall, 1988b). Theoretical epidemiology has largely escaped this controversy, presumably because some of the most influential epidemiological modellers in plant pathology (or at least some members of their laboratories) are superb experimentalists in their own right. The resulting synergism between models and experimental data in advancing theory is typified in the work of JM Jeger and CA Gilligan, among others (e.g., Jeger, 2000; Gilligan, 2002). It is important to note that testing and validation of models need not occur at the same time for a model to be useful. For instance, the theory of dispersive epidemic waves (focal epidemics that spread with increasing frontal velocity) was formalized by Ferrandino (1993) based on physical principles of spore transport, with limited empirical support. Although additional observational (Scherm, 1996) and experimental (Frantzen and van den Bosch, 2000) backing for this theory was presented in the interim, it took more than a decade after publication of Ferrandino's paper until large-scale disease gradient experiments by Cowger et al. (2005) demonstrated convincingly that epidemics of wheat stripe rust spread consistently with increasing frontal velocity.

In what follows we consider current trends in three specific areas of theoretical plant epidemiology: models of temporal and spatial dynamics of disease, the synergism of epidemiology and population genetics, and advances in statistical epidemiology. The purpose here is not to provide a comprehensive review, but rather to give selected examples illustrating these trends. Inevitably, these examples reflect our personal views of what is interesting and important in theoretical epidemiology. We limit our discussion largely to work published since the last International Workshop on Plant Disease Epidemiology in Ouro Preto, Brazil, in 2001. Selected aspects of theoretical work carried out during the 1990s have been synthesized recently (Jeger, 2000; Gilligan, 2002).

Models of temporal and spatial disease dynamics

The development of mathematical models to describe disease dynamics has been and continues to be the mainstay of theoretical epidemiology. Recent research in the area has focused on incorporating spatial structure, elucidating the consequences of stochasticity and spatial scale, identifying threshold criteria for pathogen or strain establishment, and predicting the effects of selected management strategies on disease dynamics. A detailed account of the use of analytical models to address these objectives has been given by Gilligan (2002). Based on his review and the subsequently published literature, a number of trends may be inferred.

SIR models have entered the mainstream and become more versatile

In its most basic form, an SIR model consists of a set of linked differential equations describing the dynamics of susceptible (healthy), infected, and removed (post-infectious) host tissue; commonly, the infected tissue is divided into exposed (latently infected) and infectious compartments, leading to an SEIR model (Madden, 2005). This type of analytical model, first formalized by Kermack and McKendrick (1927) for human diseases, was popularized by Jeger (1982) for use in plant epidemiology. Almost 20 years later, Segarra et al. (2001) formally derived the SEIR model for plant epidemics from the more general Kermack–McKendrick model based on first principles. In addition, Segarra et al. (2001) provided a detailed comparison of the behaviour of the latter two models with that of Van der Plank's widely used differential-delay equation (Van der Plank, 1963).

Recent work has added considerable complexity to SIR-type models (Gilligan, 2002), including demographic and environmental stochasticity (Park et al., 2003; Gibson et al., 2004; Otten et al., 2004a), seasonal disturbance and multi-year

disease dynamics (Madden and van den Bosch, 2002), dynamics of host growth and susceptibility (Gibson et al., 2004), virus vectoring mode and vector performance (Madden et al., 2000; Holt and Colvin, 2001), interactions with biocontrol agents (Gibson et al., 2004), spatial structure and metapopulation dynamics (Park et al., 2001, 2003), and the presence of pesticide-resistant subpopulations (Hall et al., 2004), among others. The inclusion of stochasticity and spatial structure is especially significant as models featuring these attributes can produce qualitatively very different predictions regarding pathogen establishment and persistence than their deterministic mean-field counterparts. Most importantly, invasion thresholds in stochastic models are higher and the pathogen or strain may be unable to persist following successful invasion due to chance events, especially at low population densities (Gilligan, 2002).

The increased complexity of contemporary SIR models adds realism and allows their application to a wider range of problems. Indeed, models are now routinely formulated to accommodate parameters useful in exploring specific management strategies (Jeger, 2000; Gilligan, 2002; Stacey et al., 2004). For instance, a linked *African cassava mosaic virus*–whitefly vector model (Jeger et al., 2004) includes four management-related parameters, namely the roguing rate of infected host plants, the insecticide-induced death rate of the vector, and the virus acquisition and transmission rates of the vector, both of which are determined by the level of host resistance. Analysis of this model indicated that roguing applied once per month in combination with a modest level of host resistance (specifically one that reduces the product of acquisition rate and transmission rate below 80% of the value of the susceptible host) is sufficient to eradicate the disease, while a combination of roguing and insecticide application is less effective. This example illustrates that analytical models have come a long way in their capacity to provide specific management recommendations that have traditionally been considered in the realm of more complex simulation models.

Nonetheless, a few words of caution are appropriate as there are some well publicized examples from the broader ecological literature where the extension of theoretical models to management has met with disastrous results (Hall, 1988a). Perhaps we need to remind ourselves occasionally that the purpose of theory is to explain rather than to predict, and that theoretical problems without practical applications are just as legitimate as empirical studies that do not contribute to the development of new theories.

Broadened scale of investigation

With few exceptions, epidemiological models have traditionally focused on the field scale, a logical choice considering the importance of individual fields as the spatial unit for tactical disease management by growers. In recent years, however, the scale of analysis has broadened to include both finer and larger scales. At one end of the spectrum is the individual plant scale, where theoretical models have been developed, for instance, to describe transmission of *Rhizoctonia solani* from an infected to a healthy plant based on models of hyphal and colony growth of the fungus through soil (Stacey et al., 2001; Otten et al., 2004b). At the cellular level, effects of phenomena such as viral cross-protection (Zhang and Holt, 2001) and synergism among different viruses (Zhang et al., 2000; Naylor et al., 2003) have been modelled with respect to their effects on field-level disease dynamics. However, epidemiological models that explicitly describe molecular processes and interactions within individual plant cells, e.g., during virus replication or virus- or transgene-induced gene silencing, are still lacking in plant pathology, even though they are common in medical epidemiology (e.g., Phillips et al., 2001).

At the other end of the spectrum are models for disease development at landscape (Park et al., 2001, 2003; Otten et al., 2004a; Stacey et al., 2004) and continental (van den Bosch et al., 1999) scales. With the rising interest in area-wide pest management and the increasing exotic species problem (Scherm and Coakley, 2003), this scale of investigation will become more important in the future. In landscape models, spread among fields has been implemented via percolation theory (Otten et al., 2004a), cellular automata (Gilligan, 2002), or in a metapopulation framework (Park et al., 2001, 2003) in which habitable patches are made up of aggregates of susceptible fields. The models allow for the analysis of disease spread in relation to within-patch pathogen dynamics, the strength of coupling among patches, and patch size, density, and arrangement.

Closely related to the issue of scale is the concept of scaling, i.e., to what degree and how data and models from one scale can be transferred to another (smaller or larger) scale. This has been an active area of research in both theoretical and statistical epidemiology (Turechek, 2005). For instance, statistical procedures to extrapolate disease incidence data from a lower hierarchical level (e.g., leaves) to a higher level (e.g., shoots) and *vice versa* have been developed (Hughes et al., 1997; Hughes and Gottwald, 1998; Madden and Hughes, 1999; McRoberts et al., 2003; Turechek and Madden, 2003) and are now increasingly being applied to develop more efficient sampling and disease assessment protocols through approaches such as cluster sampling and group testing. In a recent example, Xu et al. (2004) used a distribution-based approach to derive relationships between the incidence of spikelet infection and the more easily determined incidence of ear infection for the Fusarium head blight pathosystem on wheat. These particular relationships may be useful for making decisions in cases where management thresholds are based on the incidence of infected spikelets.

On a more process-based level, Stacey et al. (2001) developed a mathematical model to scale up from the behaviour of individual hyphae (of *R. solani* in this example) to fungal colony growth through soil and to infection of individual plants. The approach was based on a spatially explicit model of hyphal expansion incorporating the relationships between hyphal growth and fungal biomass as well as between fungal biomass, proximity of the mycelium to a susceptible root, and the probability of disease transmission. A stochastic, cellular automaton-based model for scaling up from individual plants to plant populations infected with *R. solani* had been developed previously (Kleczkowski et al., 1997), and it may be possible to combine this probabilistic model with the more detailed fungal growth-based model of Stacey et al. (2001) to arrive at estimates of both the mean and variance of the spatio-temporal dynamics of *R. solani*.

In the broader ecological literature, fractal geometry has been applied for scaling among different spatial or temporal hierarchies if the pattern or process of interest is scale-invariant, i.e., repeats itself at progressively larger scales (Brown et al., 2002; Li, 2000). In practice, scale-invariance is suggested by a straight line in a log–log plot of the measure of interest against the scale of observation. The slope of the line is interpreted as the fractal dimension, which summarizes the properties of the pattern across scales. In general, scale-invariance might be expected for organisms occurring at a population density near their lower critical threshold, e.g., due to human intervention (Cousens et al., 2004). In a recent pest management-related example, Cousens et al. (2004) counted numbers of five agricultural weeds in up to 202,500 contiguous 0.2×0.2-m quadrats in a single arable field. Counts from adjacent quadrats were pooled into progressively larger quadrats with up to 90 m-long sides. This allowed for the calculation of incidence values for different quadrat sizes and an understanding of how these incidence values vary with scale. Calculation of the fractal dimension showed that spatial patterns of those weed species that were most aggregated and/or occurred at the lowest densities were scale-invariant, indicating that patterns observed at small scales repeated themselves at progressively larger scales. Although there are theoretical reasons why such scale-invariance would be unlikely for plant pathogens (e.g., different mechanisms for long- vs. short-distance dispersal along with changes in the physical environment at different spatial scales), it would be interesting to test the null hypothesis of scale-variance for different types of pathogens, e.g., those causing aerial vs. soilborne or monocyclic vs. polycyclic diseases. Scale-invariance, if it occurs in plant pathogens, would allow for extrapolation and prediction over a wide range of spatial scales with potentially useful applications in areas such as precision agriculture.

Ferrandino (2004) recently proposed a sampling approach for disease incidence based on a nested fractal design, i.e., one in which sampling points at distances of, say, 1, 2, 4, 8, and 16 m are represented equally. Using simulated spatial epidemics, he showed that this design was more efficient in detecting aggregation than either regular, random, or spatially clustered sampling designs, in addition to providing spatial information over a wider range of scales.

Fascination with thresholds

Van der Plank (1963) expressed his threshold theorem as $iR_c > 1$, which states that an epidemic will not occur unless the product of infectious

period i and basic corrected infection rate R_c exceeds 1. For consistency with the medical and ecological literature, the theorem has been rewritten as $R_0 > 1$, where R_0 is the basic reproductive number, i.e., the number of new infected individuals resulting from one introduced infected individual (Madden, 2005). Although the interest in thresholds for plant epidemics has been long-standing (e.g., Jeger and van den Bosch, 1994a, b), we note a recent surge of activity in this area, mostly derived from analyses with SIR-type models. This has included derivation of threshold criteria based not only on pathogen demographics (e.g., fitness of fungicide-resistant strains; Hall et al., 2004), but also on host spatial structure (e.g., host density, patch size, and coupling among patches; Bailey et al., 2000; Gubbins et al., 2000; Park et al., 2001; Otten et al., 2004a). One of the reasons for the current preoccupation with thresholds in plant epidemiology is the wider availability of stochastic models, which allows for the calculation of not only the risk of pathogen invasion, but also the probability of subsequent persistence in the face of chance events that can lead to extinction at low population densities.

Apart from stochastic forces, the establishment of a pathogen following its successful introduction may be limited by certain demographic features, such as the difficulty to find a compatible mating partner at very low population densities for species with an obligate sexual cycle (Taylor and Hastings, 2005). This feature leads to an intermediate optimum in the relationship between population growth rate and population density (Allee effect; Figure 1). In a deterministic population model of the heterothallic Karnal bunt fungus *Tilletia indica*, inclusion of an Allee effect resulted in a teliospore threshold for establishment about two orders of magnitude higher than in a version of the model without this constraint (Garrett and Bowden, 2002). This finding has potentially important implications for risk assessments of *T. indica* and other quarantine pathogens with an obligate sexual cycle.

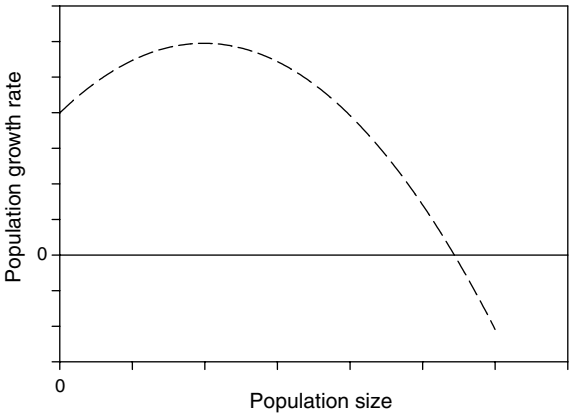


Figure 1. Graphical representation of the Allee effect showing an intermediate optimum in the relationship between population growth rate and population density, e.g., due to the difficulty to find a compatible mating partner at very low population densities for species with an obligate sexual cycle.

Pathogen population biology

Epidemiology is a holistic discipline (Zadoks, 1990), and the development of epidemiological theory thus requires an interdisciplinary approach. This includes not only mathematics and statistics, but also concepts and tools from population genetics. The main focus of population genetics is to understand the evolutionary processes driving and maintaining genetic variation within and among populations (McDonald, 2004). Because host and pathogen populations consist of distinct genetic entities, the fundamental theory of their population dynamics in space and time must coincide with that of their genetic composition. Conceptually, there is thus a considerable overlap between epidemiology and population genetics. Here, we take an heuristic look at the interplay between the two disciplines in the development and application of epidemiological theory and highlight areas that may best be served by an interdisciplinary approach. An in-depth review of the nature of the synergy between epidemiology and population genetics is outside the scope of this paper but is available elsewhere (Milgroom, 2001; Milgroom and Peever, 2003).

It has been suggested that, over the past 20 years, a schism appears to have developed between epidemiology and population genetics (Milgroom, 2001; Milgroom and Peever, 2003). Indeed, it is tempting to conclude that such a split was a consequence of both disciplines becoming more specialized as they responded to new technologies; epidemiology to the availability of advanced modelling techniques and increased computing power, population genetics to advances in molecular biology. In practice, however, the two

disciplines often have been utilized jointly to address applied epidemiological questions such as source and type of primary inoculum (Gobbin et al., 2003; Peever et al., 2004), dispersal of secondary inoculum (Cortesi et al., 2000; Cortesi and Milgroom, 2001; Loskill et al., 2004), or host specificity (Peever et al., 2000; Akimitsu et al., 2003; Flier et al., 2003). A notable example in a theoretical sense is the recent work by plant pathologists on the appropriateness of the application of measures of genotypic diversity to microbial populations (Grünwald et al., 2003; Kosman and Leonard, 2005). From these examples it should be obvious that plant epidemiology can benefit greatly from concepts and tools developed in population genetics (and *vice versa*), including in studies designed to test theoretical ideas.

Eriksen et al. (2001) used numerical simulations to address a question that had been the subject of vigorous theoretical discussions. The problem, broadly put, was to determine the role of ascospores in development and microevolution of septoria tritici blotch of wheat caused by *Mycosphaerella graminicola*. Population genetics studies in the United States in the 1990s had provided indirect evidence for sexual reproduction by *M. graminicola* during the growing season (McDonald et al., 1995; Chen and McDonald, 1996). An important question was how to determine the relative contribution of immigration (gene flow) and sexual reproduction to the genetic structure of the pathogen during the course of an epidemic, and which of these two evolutionary forces is of greater epidemiological importance within a season. This was resolved, not without some debate, through mark-release-recapture experiments (Zhan et al., 1998) and a theoretical analysis (a subject of two letters to the editor in *Phytopathology*) of the data to estimate the rates of recombination and migration (Brown, 2000; Zhan et al., 2000). Nonetheless, these studies did not answer the question of the relative contribution of ascospores vs. pycnidiospores to disease development, nor of the extent of genetic recombination. Through simulation modelling Eriksen et al. (2001) showed that the extended latent period of pseudothecia compared with that of pycnidia leads to the release of ascospores too late in the season to have a major effect on final severity of septoria tritici blotch epidemics in northern Europe. However, ascospores contributed appreciably to the genetic composition of the pathogen population (as indicated by the proportion of sexual descendants among lesions at the end of the season), especially in dry conditions unfavourable for the dispersal of pycnidiospores.

With regard to analytical modelling approaches, one of the key challenges has been to integrate epidemiological parameters into population genetics models (and *vice versa*) while at the same time keeping model complexity at a manageable level. Jeger (1997) illustrated this by incorporating host-pathogen gene-for-gene interactions into an analytical SIR model. This resulted in a set of six linked differential equations, one each for homozygous and heterozygous genotypes of both host and pathogen. Although the model was not very tractable analytically, it allowed for the derivation of threshold criteria for persistence of specific pathogen and host genotypes. Subsequent simplification of the model allowed the effects of host density dependence, fitness cost for virulence in the pathogen, and fitness cost for host resistance to be incorporated and analyzed.

Durability of host resistance, a key concept in population genetics, also has been examined from an epidemiological perspective (van den Bosch and Gilligan, 2003). This analysis considered three epidemiologically based measures of durability of resistance: (1) time to invasion by a virulent pathogen genotype; (2) time taken for the virulent genotype to dominate the pathogen population; and (3) time until a threshold proportion of the host population becomes diseased ('additional uninfected crop growth days'). These metrics differ conceptually from conventional population genetics-based measures of resistance durability in that they emphasize quantitative rather than qualitative aspects, i.e., they focus on the duration of resistance utility rather than the conditions under which durability is maintained. The model showed that if the virulent pathogen genotype is not already present, and the time between introduction (by mutation or immigration) and establishment is considered as a metric of resistance durability, both low and high proportions of resistant genotypes in the crop can prolong durability. This observation might explain the oft-encountered difficulty in trying to predict the durability of resistance genes (Hovmøller et al., 1997; Brown, 2002; Burnett, 2003). The results

also showed that the metric representing additional crop growth days without disease is unaffected by the proportion of the resistant host genotype in a cultivar mixture, thus concurring with data from experimental field studies in which varying the proportion of mixture constituents (of sorghum in this case) had no effect on time of disease onset (Ngugi et al., 2001).

Gudelj et al. (2004) investigated evolution of sibling fungal plant pathogens from an epidemiological perspective using adaptive dynamics methodology. They focused on the role of multiple host species involving a trade-off between the evolutionary benefit of being specialized and its cost (reduced virulence on other hosts). The results showed that this infectivity trade-off accounted for the evolution of only those pathogen siblings with non-overlapping host ranges (i.e., a high degree of host specialization such as observed with obligate parasites), and that other mechanisms (ecological and/or epidemiological) must account for the evolution of generalists with overlapping host ranges and that of groups containing both generalist and specialist siblings.

Several important generalizations about the role of spatial structure in host-pathogen coevolution can be drawn from the work by PH Thrall and JJ Burdon, who integrated population genetics and spatio-temporal analysis of epidemics in natural pathosystems (Burdon and Thrall, 1999, 2004). Key among these is that disease patterns in host–pathogen metapopulations are spatially and temporally asynchronous, whereby the magnitude of pathogen fluctuations varies between host populations but there is clustering of disease levels among populations. This prediction is supported by results of experimental studies (Burdon and Thrall, 2000; Thrall and Burdon, 2000, 2003; Bock et al., 2002; Thrall et al., 2002). Further, disease persistence, and hence its impact on coevolution, is higher at the local level. As a consequence, there is a tight evolutionary link between resistance and virulence of associated host–pathogen pairs whereby pathogen virulence (ability to infect many host genotypes) increases with increasing mean resistance of the host sub-population (Thrall and Burdon, 2003). These studies also provided evidence for a trade-off between virulence and aggressiveness (defined here as spore production per pustule), whereby selection for the former is favoured in resistant host genotypes while that for aggressiveness is favoured in susceptible host genotypes. Although the use of spore production as a measure of aggressiveness may be subject to debate, the study marks an important step toward documenting a virulence-aggressiveness trade-off for which previous evidence has been weak (Mundt, 2002), especially in natural systems.

Statistical epidemiology

Apart from forming a crucial link between theory and data, statistical concepts – in their own right – may result in new theoretical knowledge about plant pathosystems and plant epidemiology. For instance, distribution-based methods to characterize disease aggregation in a spatial hierarchy (Hughes et al., 1997; Madden and Hughes, 1999) have led to novel, testable hypotheses regarding disease dynamics in time and space, e.g., for incidence–severity relationships. With ever increasing computing power and a better understanding of how to utilize contemporary statistical tools, new opportunities for the application of statistics in plant epidemiology, both theoretical and applied, continue to emerge.

Generalized linear mixed models

Garrett et al. (2004) highlighted several statistical methods that are used relatively little but have the potential to improve inference from a range of epidemiological studies. Foremost among these are mixed-effects models, i.e., models to analyze data with fixed and random effects. At a theoretical level, the nature and properties of generalized linear mixed models (GLMMs) have been understood for decades (McCulloch and Searle, 2001), but until recently, without significant input from a specialist statistician, mixed-effects modelling has been very difficult in practice. Now, an increasing number of articles in application-oriented journals provide guidance for setting up mixed models and for implementing them in off-the-shelf statistical packages (Piepho, 1999; Madden et al., 2002; Piepho et al., 2003; Spilke et al., 2005). One of the most important advantages of these models is their applicability to unbalanced designs, for which exact statistical tests are usually not available. Therefore, one needs to resort to approximate methods such as the restricted maximum likelihood approach. Even for experimental designs for which traditional general linear models (GLMs)

are appropriate, analysis using the GLMM can produce more robust results when variances are unequal and/or sample sizes are small (Piepho, 1999). Madden et al. (2002) evaluated various GLMMs and recommended the fixed residual variance model (which is also the simplest GLMM) for analyzing disease incidence data from designed experiments.

Survival analysis

Data on the occurrence and timing of events such as sclerotium germination, disease onset, or leaf abscission are routinely encountered in epidemiological studies. With such time-to-event data, several problems can arise that limit the usefulness of traditional statistical methods: (1) the times are unlikely to be distributed normally; (2) the data set will likely contain censored observations, i.e., observations for which the event has not occurred when the study was completed; and (3) the response may be influenced by time-dependent covariates, i.e., independent variables whose values change during the study period. Because of these properties, time-to-event data are now increasingly being modelled using survival analysis (Scherm and Ojiambo, 2004). This set of statistical methods not only allows the comparison of time-to-event distributions among treatment groups, but also the development of models for the effects of discrete and/or continuous covariates on event times. Recently published examples include analyses of the effects of landscape attributes on the time to invasion by an exotic plant pathogen (Jules et al., 2002); of orchard characteristics, environment, and disease status of neighbouring trees on the time of virus infection of individual orchard trees (Dallot et al., 2004); and of disease severity and other leaf attributes on the time of premature defoliation of diseased plants (Ojiambo and Scherm, 2005).

Decision analysis

Important advances have been made in the area of decision analysis for disease management, especially in relation to the quantitative evaluation of risk algorithms such as disease forecasters (Yuen et al., 1996; Hughes et al., 1999; Yuen and Hughes, 2002; Madden, 2005). Increasingly, ROC (receiver-operating characteristic) analysis is being employed to optimize risk algorithms and thresholds for making decisions. An ROC curve is a plot of the true positive rate (sensitivity) as a function of the false positive rate (1 – specificity) at all possible decision thresholds of the risk algorithm. This curve allows one to identify trade-offs between liberal and conservative thresholds in an attempt to identify the most suitable decision threshold for a given application. ROC analysis is best suited for responses that are inherently dichotomous, for instance the decision whether or not to apply a fungicide. In a recent example, Dewdney et al. (2002) used ROC analysis and historical data to evaluate parameters of MARYBLYT (a forecaster for fire blight of apple and pear) and to identify where key improvements were needed. MARYBLYT and Cougarblight (another fire blight forecaster) have been compared using ROC analysis and found to have equivalent action thresholds and thus perform similarly in their ability to predict blossom blight (Dewdney et al., 2003).

ROC analysis also can be applied in situations where the response is not dichotomous (Patil, 1991), for instance the decision on how much fertilizer to apply or how many fungicide applications to make. However, in plant epidemiology, ROC analysis of responses on a non-dichotomous scale has yet to be demonstrated.

Bayesian analysis

The evaluation of plant disease forecasters based on ROC analysis may be improved further when conducted in a Bayesian framework (Yuen and Hughes, 2002). This is accomplished by considering the prior probability of disease occurrence in addition to the likelihood ratios for positive and negative predictions by the risk algorithm. The latter two are calculated directly from sensitivity and specificity of the forecaster, while the former may be based either on the historical prevalence of the disease in the region of interest, or on growers' subjective estimates of disease risk. In either case, the result is a posterior probability of disease occurrence given the prediction by the forecaster. Yuen and Hughes (2002) illustrate this approach by means of risk algorithms for eyespot of wheat and Sclerotinia stem rot of canola (oilseed rape).

Apart from its application in the specific example of ROC analysis discussed above, Bayes's theorem presents a general framework for incorporating uncertainty and prior information into

epidemiological analyses and for updating current knowledge as new information becomes available (Mila and Carriquiry, 2004). The key feature is the calculation of posterior probabilities for the parameter of interest based on empirically derived prior probabilities in conjunction with the conditional probability of each possible outcome. This use of prior probabilities represents a powerful mechanism for incorporating subjective information such as growers' perceptions. This is illustrated in the work of Mila et al. (2003), who examined the effect of soybean growers' production decisions on Sclerotinia stem rot incidence using decision theory under uncertainty. Predictions of stem rot incidence and soybean yield based on regression-type models were updated with growers' subjective estimates of disease incidence via Bayes's theorem. The resulting posterior probabilities were then used to derive management criteria for profit maximization.

Economic criteria (which often exhibit considerable uncertainty) and growers' perceptions are among the most important drivers affecting disease management decisions, yet they are routinely ignored by plant pathologists developing decision algorithms. The continued penetration of Bayesian analysis into the epidemiological mainstream should lead to a greater appreciation of the importance of these drivers and – it is hoped – their more widespread incorporation into disease management models.

Statistical genetics and bioinformatics

As shown in the above examples, statistical plant epidemiology has relied heavily on the medical and ecological fields for inspiration and conceptual advances. This trend will likely continue in the future as plant epidemiologists become more familiar with the theories and tools of statistical genetics and bioinformatics. In a recent example, Parsons and Te Beest (2004) used genetic algorithms to optimize fungicide applications on winter wheat relative to spray date as well as choice, number, and dose of active ingredients. Genetic algorithms use biologically derived concepts such as inheritance, mutation, natural selection, and recombination to 'evolve' a large population of possible solutions ('individuals') to the best ('fittest') solution ('survivor'). The evolution starts from a population of completely random individuals, and in each subsequent generation multiple individuals are selected stochastically and modified (mutated or recombined) to form a new population. Although the concept of evolutionary computing may be intuitively appealing to biologists, the approach is computationally intensive and effectively treats the optimization problem as a black box. Its theoretical and practical impact on plant epidemiology remains to be seen.

Apart from merely adopting statistical tools from other disciplines, plant epidemiologists should be more proactive in exploring potential applications of their concepts and procedures in rapidly expanding disciplines such as statistical genetics or bioinformatics. Conceptually, for instance, there are many parallels between the dynamics of plant pathogens in populations of plants and those of genetic loci or markers within a genome (Delwiche, 2004). The key here is to remain imaginative and keep an open mind toward broader applications, without being confined to the organismal level that has historically dominated statistical applications in plant epidemiology.

Conclusions

Based on the selected examples given above there can be little doubt that significant progress has been made in theoretical plant epidemiology since the turn of the century. New theories and models continue to be developed, and sincere efforts are being made to relate them to the broader field of theoretical biology on one hand and practical disease management on the other. As analytical models of plant disease dynamics have become more realistic, they also have become considerably more complex, and solutions often can be obtained only numerically. As such, the division between analytical and simulation models, an important distinction some 20 years ago (Jeger, 1986), is narrowing. It seems that we have come back full circle to the medium-sized models advocated by Botkin's (1977).

Although theoretical problems need not be tied to practical applications to be valid, the image of theoretical epidemiology within the larger field of plant pathology could benefit from a clearer documentation of its impact on practical disease management. In medical epidemiology, such evaluations are commonly achieved by comparing

model outputs with long-term morbidity data sets, e.g., in the case of models for the impact of vaccination on childhood diseases (Rohani et al., 2000). We would like to call attention to the need for similar analyses in plant epidemiology, especially with pathosystems for which long-term data are available (e.g., the cereal rusts). Establishment of additional long-term data collection standards, even if only for a limited number of pathosystems, would provide a more solid data base from which to evaluate the impact of interventions suggested by current theoretical knowledge.

While plant epidemiology, by definition, is concerned with the study of populations of pathogens in populations of plants, there exists ample opportunity to broaden the scale of investigation and apply the concepts of theoretical epidemiology to both sub-organismal and ecosystem scales. Examples of such non-traditional applications could include models of virus cross-protection in individual plant cells, temporal and spatial dynamics of molecular markers or of molecules such as mycotoxins, biotechnology risk assessment, microbial forensics, or the quantitative analysis of ecosystem health. We would argue that plant epidemiologists, including theoreticians, are not yet taking advantage of these new opportunities sufficiently. Although providing the scientific basis for disease management will always be the *raison d'être* for plant epidemiology, a broader perspective will help the discipline to remain relevant as more efforts and resources continue to be devoted to genomic and ecosystem-level science. Plant epidemiology, both theoretical and applied, will remain as integrating a discipline as it has ever been, but the individual components that require integration are changing.

References

Akimitsu K, Peever TL and Timmer LW (2003) Molecular, ecological and evolutionary approaches to understanding *Alternaria* diseases of citrus. Molecular Plant Pathology 4: 435–446

Anon. (2005) Encyclopaedia Britannica Online. http://www.britannica.com (accessed 28 August 2005)

Bailey DJ, Otten W and Gilligan CA (2000) Saprotrophic invasion by the soil-borne fungal pathogen *Rhizoctonia solani* and percolation thresholds. New Phytologist 146: 535–544

Bock CH, Thrall PH, Brubaker CL and Burdon JJ (2002) Detection of genetic variation in *Alternaria brassicicola* using AFLP fingerprinting. Mycological Research 106: 428–434

Botkin DB (1977) Bits, bytes and IBP. Bioscience 27: 385

Brown JH, Gupta VK, Li BL, Milne BT, Restrepo C and West GB (2002) The fractal nature of nature: Power laws, ecological complexity and biodiversity. Philosophical Transactions of the Royal Society of London, Series B 357: 619–626

Brown JKM (2000) Estimation of rates of recombination and migration in populations of plant pathogens. Phytopathology 90: 320–323

Brown JKM (2002) Yield penalties of disease resistance in crops. Current Opinions in Plant Biology 5: 339–344

Burdon JJ and Thrall PH (1999) Spatial and temporal patterns in coevolving plant and pathogen associations. American Naturalist 153(Suppl): S15–S33

Burdon JJ and Thrall PH (2000) Coevolution at multiple spatial scales: *Linum marginale–Melampsora lini* – from the individual to the species. Evolutionary Ecology 14: 261–281

Burdon JJ and Thrall PH (2004) Genetic structure of natural plant and pathogen populations. In: Ehler LE, Sforza R and Mateille T (eds) Genetics, Evolution, and Biological Control (pp 1–17) CABI Publishing, Wallingford, UK

Burnett J (2003) Fungal Populations and Species. Oxford University Press, Oxford, UK

Caswell H (1988) Theory and models in ecology: A different perspective. Ecological Modelling, 43: 33–44

Chen RS and McDonald BA (1996) Sexual reproduction plays a major role in the genetic structure of populations of the fungus *Mycosphaerella graminicola*. Genetics 142: 1119–1127

Cortesi P and Milgroom MG (2001) Outcrossing and diversity of vegetative compatibility types in populations of *Eutypa lata* from grapevines. Journal of Plant Pathology 83: 79–86

Cortesi P, Fischer M and Milgroom MG (2000) Identification and spread of *Fomitiporia punctata* associated with wood decay of grapevine showing symptoms of esca. Phytopathology 90: 967–972

Cousens R, Wallinga J and Shaw M (2004) Are the spatial patterns of weeds scale-invariant? Oikos 107: 251–264

Cowger C, Wallace LD and Mundt CC (2005) Velocity of spread of wheat stripe rust epidemics. Phytopathology 95: 972–982

Dallot S, Gottwald T, Labonne G and Quiot JB (2004) Factors affecting the spread of *Plum pox virus* strain M in peach orchards subjected to roguing in France. Phytopathology 94: 1390–1398

Delwiche CF (2004) The genome palimpsest: Genomics in evolution and ecology. Bioscience 54: 991–1001

Dewdney MM, Biggs AR, Lightner G and Turechek WW (2002) MARYBLYT: Can there be improvements? (Abstr) Phytopathology 92: S19

Dewdney MM, Biggs AR and Turechek WW (2003) A statistical comparison of MARYBLYT and Cougarblight using Receiver Operator Characteristic (ROC) analysis. (Abstr) Phytopathology 93: S20

Eriksen L, Shaw MW and Østergård H (2001) A model of the effect of pseudothecia on genetic recombination and epidemic development in populations of *Mycosphaerella graminicola*. Phytopathology 91: 240–248

Ferrandino FJ (1993) Dispersive epidemic waves. I. Focus expansion within a linear planting. Phytopathology 83: 795–802

Ferrandino FJ (2004) Measuring spatial aggregation in binary epidemics: Correlative analysis and the advantage of fractal-based sampling. Phytopathology 94: 1215–1227

Flier WG, Grünwald NJ, Kroon LPNM, Sturbaum AK, vanden Bosch TBM, Garay-Serrano E, Lozoya-Saldana H, Fry WE and Turkensteen LJ (2003) The population structure of *Phytophthora infestans* from the Toluca Valley of central Mexico suggests genetic differentiation between populations from cultivated potato and wild *Solanum* spp. Phytopathology 93: 382–390

Frantzen J and van den Bosch F (2000) Spread of organisms: Can travelling and dispersive waves be distinguished? Basic and Applied Ecology 1: 83–91

Garrett KA and Bowden RL (2002) An Allee effect reduces the invasive potential of *Tilletia indica*. Phytopathology 92: 1152–1159

Garrett KA, Madden LV, Hughes G and Pfender WF (2004) New applications of statistical tools in plant pathology. Phytopathology 94: 999–1003

Gibson GJ, Kleczkowski A and Gilligan CA (2004) Bayesian analysis of botanical epidemics using stochastic compartmental models. Proceedings of the National Academy of Sciences USA 101: 12120–12124

Gilligan CA (2002) An epidemiological framework for disease management. Advances in Botanical Research 38: 1–64

Gobbin D, Pertot I and Gessler C (2003) Genetic structure of a *Plasmopara viticola* population in an isolated Italian mountain vineyard. Journal of Phytopathology 151: 636–646

Grünwald NJ, Goodwin SB, Milgroom MG and Fry WE (2003) Analysis of genotypic diversity data for populations of microorganisms. Phytopathology 93: 738–746

Gubbins S, Gilligan CA and Kleczkowski A (2000) Population dynamics of plant-parasite interactions: Thresholds for invasion. Theoretical Population Biology 57: 219–233

Gudelj I, Fitt BDL and van den Bosch F (2004) Evolution of sibling fungal plant pathogens in relation to host specialization. Phytopathology 94: 789–795

Hall CAS (1988a) An assessment of several of the historically most influential theoretical models used in ecology and of the data provided in their support. Ecological Modelling 43: 5–31

Hall CAS (1988b) What constitutes a good model and by whose criteria? Ecological Modelling 43: 125–127

Hall RJ, Gubbins S and Gilligan CA (2004) Invasion of drug and pesticide resistance is determined by a trade-off between treatment efficacy and relative fitness. Bulletin of Mathematical Biology 66: 825–840

Holt J and Colvin J (2001) Observation and theory of whitefly-borne virus disease epidemics. In: Jeger MJ and Spence NJ (eds) Biotic Interactions in Plant–Pathogen Associations (pp 331–343) CABI Publishing, Wallingford, UK

Hovmøller MS, Østergård H and Munk L (1997) Modelling virulence dynamics of airborne plant pathogens in relation to selection by host resistance in agricultural crops. In: Crute IR, Holub EB and Burdon JJ (eds) The Gene-for-Gene Relationship in Plant-Parasite Interactions (pp 173–190) CABI Publishing, Wallingford, UK

Hughes G and Gottwald TR (1998) Survey methods for assessment of *Citrus tristeza virus* incidence. Phytopathology 88: 715–723

Hughes G, McRoberts N and Burnett FJ (1999) Decision-making and diagnosis in disease management. Plant Pathology 48: 147–153

Hughes G, McRoberts N, Madden LV and Gottwald TR (1997) Relationships between disease incidence at two levels in a spatial hierarchy. Phytopathology 87: 542–550

Jeger MJ (1982) The relation between total, infectious, and postinfectious diseased plant tissue. Phytopathology 72: 1185–1189

Jeger MJ (1986) The potential of analytic compared with simulation approaches to modeling plant disease epidemiology. In: Leonard KJ and Fry WE (eds) Plant Disease Epidemiology: Population Dynamics and Management (pp 255–281)1 Macmillan, New York, USA

Jeger MJ (1997) An epidemiological approach to modelling the dynamics of gene-for-gene interactions. In: Crute IR, Holub EB and Burdon JJ (eds) The Gene-for-Gene Relationship in Plant–Parasite Interactions (pp 191–209) CABI Publishing, Wallingford, UK

Jeger MJ (2000) Theory and plant epidemiology. Plant Pathology 49: 651–658

Jeger MJ and van den Bosch F (1994a) Threshold criteria for model plant disease epidemics: I. Asymptotic results. Phytopathology 84: 24–27

Jeger MJ and van den Bosch F (1994b) Threshold criteria for model plant disease epidemics: II. Persistence and endemicity. Phytopathology 84: 28–30

Jeger MJ, Holt J, van den Bosch F and Madden LV (2004) Epidemiology of insect-transmitted plant viruses: Modelling disease dynamics and control interventions. Physiological Entomology 29: 291–304

Jules ES, Kauffman MJ, Ritts WD and Carroll AL (2002) Spread of an invasive pathogen over a variable landscape: A non-native root rot on Port Orford cedar. Ecology 83: 3167–3181

Kermack WO and McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London, Series A 115: 700–721

Kleczkowski A, Gilligan CA and Bailey DJ (1997) Scaling and spatial dynamics in plant-pathogen systems: From individuals to populations. Proceedings of the Royal Society of London, Series B 264: 979–984

Kosman E and Leonard KJ (2005) Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. Molecular Ecology 14: 415–424

Li BL (2000) Fractal geometry applications in description and analysis of patch patterns and patch dynamics. Ecological Modelling 132: 33–50

Loskill B, Gobbin D and Berkelmann-Löhnertz B, (2004) Vertikale Verteilung bodenbürtiger Infektionen durch *Plasmopara viticola*. (Abstr) Mitteilungen aus der Biologischen Bundesanstalt für Land- und Forstwirtschaft, Berlin-Dahlem 396: 113

Madden LV (2005) Botanical epidemiology: Some key advances and its continuing role in disease management. European Journal of Plant Pathology (this volume)

Madden LV and Hughes G (1999) An effective sample size for predicting plant disease incidence in a spatial hierarchy. Phytopathology 89: 770–781

Madden LV and van den Bosch F (2002) A population-dynamics approach to assess the threat of plant pathogens as biological weapons against annual crops. Bioscience 52: 65–74

Madden LV, Jeger MJ and van den Bosch F (2000) A theoretical assessment of the effects of vector–virus transmission mechanism on plant virus disease epidemics. Phytopathology 90: 576–594

Madden LV, Turechek WW and Nita M (2002) Evaluation of generalized linear mixed models for analyzing disease incidence data obtained in designed experiments. Plant Disease 86: 316–325

McCulloch CE and Searle SR (2001) Generalized, Linear, and Mixed Models. Wiley, New York, USA

McDonald BA (2004) Population genetics of plant pathogens. The Plant Health Instructor, doi:10.1094/PHI-A-2004-0524-01

McDonald BA, Pettway RE, Chen RS, Boeger JM and Martinez JP (1995) The population genetics of *Septoria tritici* (teleomorph *Mycosphaerella graminicola*). Canadian Journal of Botany 73(Suppl 1): S292–S301

McRoberts N, Hughes G and Madden LV (2003) The theoretical basis and practical application of relationships between different disease intensity measurements in plants. Annals of Applied Biology 142: 191–211

Mila AL and Carriquiry AL (2004) Bayesian analysis in plant pathology. Phytopathology 94: 1027–1030

Mila AL, Carriquiry AL, Zhao J and Yang XB (2003) Impact of management practices on prevalence of soybean *Sclerotinia* stem rot in the north-central United States and on farmers' decisions under uncertainty. Plant Disease 87: 1048–1058

Milgroom MG (2001) The synthesis of genetics and epidemiology: Contributions of population biology in plant pathology. Journal of Plant Pathology 83: 57–62

Milgroom MG and Peever TL (2003) Population biology of plant pathogens: The synthesis of plant disease epidemiology and population genetics. Plant Disease 87: 608–617

Mundt CC (2002) Use of multiline cultivars and cultivar mixtures for disease management. Annual Review of Phytopathology 40: 381–410

Naylor M, Godfray HCJ, Pallett DW, Tristem M, Reeves JP and Cooper JI (2003) Mutualistic interactions amongst viruses? In: Hails RS, Beringer JE and Godfray HCJ (eds) Genes in the Environment (pp 205–225) Blackwell, Oxford, UK

Ngugi HK, King SB, Holt J and Julian AM (2001) Simultaneous temporal progress of sorghum anthracnose and leaf blight in crop mixtures with disparate patterns. Phytopathology 91: 720–729

Ojiambo PS and Scherm H (2005) Survival analysis of time to abscission of blueberry leaves affected by *Septoria* leaf spot. Phytopathology 95: 108–113

Otten W, Bailey DJ and Gilligan CA (2004a) Empirical evidence of spatial thresholds to control invasion of fungal parasites and saprotrophs. New Phytologist 163: 125–132

Otten W, Harris K, Young IM, Ritz K and Gilligan CA (2004b) Preferential spread of the pathogenic fungus *Rhizoctonia solani* through structured soil. Soil Biology and Biochemistry 36: 203–210

Park AW, Gubbins S and Gilligan CA (2001) Invasion and persistence of plant parasites in a spatially structured host population. Oikos 94: 162–174

Park AW, Gubbins S and Gilligan CA (2003) Extinction times for closed epidemics: The effects of host spatial structure. Ecology Letters 5: 747–755

Parsons DJ and Te Beest D (2004) Optimising fungicide applications on winter wheat using genetic algorithms. Biosystems Engineering 88: 401–410

Patil GP (1991) Encountered data, statistical ecology, environmental statistics, and weighted distribution methods. Environmetrics 2: 377–423

Peever TL, Olsen L, Ibañez A and Timmer LW (2000) Genetic differentiation and host specificity among populations of *Alternaria* spp. causing brown spot of grapefruit and tangerine × grapefruit hybrids in Florida. Phytopathology 90: 407–414

Peever TL, Salimath SS, Su G, Kaiser WJ and Muehlbauer FJ (2004) Historical and contemporary multilocus population structure of *Ascochyta rabiei* (teleomorph: *Didymella rabiei*) in the Pacific Northwest of the United States. Molecular Ecology 13: 291–309

Phillips AN, Youle M, Johnson M and Loveday C (2001) Use of a stochastic model to develop understanding of the impact of different patterns of antiretroviral drug use on resistance development. AIDS 15: 2211–2220

Piepho HP (1999) Analysing disease incidence data from designed experiments by generalized linear mixed models. Plant Pathology 48: 668–674

Piepho HP, Büchse A and Emrich K (2003) A hitchhiker's guide to mixed models for randomized experiments. Journal of Agronomy and Crop Science 189: 310–322

Rohani P, Earn DJD and Grenfell BT (2000) Impact of immunisation on pertussis transmission in England and Wales. The Lancet 355: 285–286

Scherm H (1996) On the velocity of epidemic waves in model plant disease epidemics. Ecological Modelling 87: 217–222

Scherm H and Coakley SM (2003) Plant pathogens in a changing world. Australasian Plant Pathology 32: 157–165

Scherm H and Ojiambo PS (2004) Applications of survival analysis in botanical epidemiology. Phytopathology 94: 1022–1026

Segarra J, Jeger MJ and van den Bosch F (2001) Epidemic dynamics and patterns of plant diseases. Phytopathology 91: 1001–1010

Spilke J, Piepho HP and Hu X (2005) Analysis of unbalanced data by mixed linear models using the MIXED procedure of the SAS system. Journal of Agronomy and Crop Science 191: 47–54

Stacey AJ, Truscott JE and Gilligan CA (2001) Soilborne fungal pathogens: Scaling-up from hyphal to colony behaviour and the probability of disease transmission. New Phytologist 150: 169–177

Stacey AJ, Truscott JE, Asher MJC and Gilligan CA (2004) A model for the invasion and spread of rhizomania in the United Kingdom: Implications for disease control strategies. Phytopathology 94: 209–215

Taylor CM and Hastings A (2005) Allee effects in biological invasions. Ecology Letters 8: 895–908

Thrall PH and Burdon JJ (2000) Effect of resistance variation in a natural plant host-pathogen metapopulation on disease dynamics. Plant Pathology 49: 767–773

Thrall PH and Burdon JJ (2003) Evolution of virulence in a plant host–pathogen metapopulation. Science 299: 1735–1737

Thrall PH, Burdon JJ and Bever JD (2002) Local adaptation in the *Linum marginale–Melampsora lini* host–pathogen interaction. Evolution 56: 1340–1351

Turechek WW (2005) Spatial and temporal scales in plant disease epidemiology. European Journal of Plant Pathology (this volume)

Turechek WW and Madden LV (2003) A generalized linear modeling approach for characterizing disease incidence in a spatial hierarchy. Phytopathology 93: 458–466

vanden Bosch F and Gilligan CA (2003) Measures of durability of resistance. Phytopathology 93: 616–625

vanden Bosch F, Metz JAJ and Zadoks JC (1999) Pandemics of focal plant disease, a model. Phytopathology 89: 495–505

Vander Plank JE (1963) Plant Diseases: Epidemics and Control, Academic Press, New York, USA

Xu XM, Parry DW, Edwards SG, Cooke BM, Doohan FM, vanMaanen A, Brennan JM, Monaghan S, Moretti A, Tocco G, Mule G, Hornok L, Giczey G, Tatnell J, Nicholson P and Ritieni A (2004) Relationship between the incidences of ear and spikelet infection of *Fusarium* ear blight in wheat. European Journal of Plant Pathology 110: 959–971

Yuen JE and Hughes G (2002) Bayesian analysis of plant disease prediction. Plant Pathology 51: 407–412

Yuen JE, Twengström E and Sigvald R (1996) Calibration and verification of risk algorithms using logistic regression. European Journal of Plant Pathology 102: 847–854

Zadoks JC (1990) Whithering plant disease epidemiology?. Plant Disease 74: 82

Zadoks JC (2001) Plant disease epidemiology in the twentieth century. A picture by means of selected controversies. Plant Disease 85: 808–816

Zhan J, Mundt CC and McDonald BA (1998) Measuring immigration and sexual reproduction in field populations of *Mycosphaerella graminicola*. Phytopathology 88: 1330–1337

Zhan J, Mundt CC and McDonald BA (2000) Estimation of rates of recombination and migration in populations of plant pathogens – a reply. Phytopathology 90: 324–326

Zhang XS and Holt J (2001) Mathematical models of cross protection in the epidemiology of plant virus diseases. Phytopathology 91: 924–934

Zhang XS, Holt J and Colvin J (2000) Mathematical models of host plant infection by helper-dependent virus complexes: Why are helper viruses always avirulent? Phytopathology 90: 85–93

European Journal of Plant Pathology (2006) 115:75–93
DOI 10.1007/s10658-005-4039-x

Establishing priorities for plant science research and developing world food security[☆]

Robert W. Herdt
Adjunct International Professor of Applied Economics and Management, Cornell University, Ithaca, NY, 14853, USA (Phone: +1-607-255-8802; E-mail: rwh13@cornell.edu)

Accepted 30 September 2005

Key words: congruence, crop loss, development assistance, hunger, research priorities, technological change

Abstract

This paper begins with a broad review of food security in the developing world. I argue that technological change has made a key contribution to improving food security wherever it has been achieved and that plant sciences can contribute in the future. Potential contributions by plant scientists will have to be funded through development assistance. A perspective on development assistance and the role of assistance to agricultural research in particular provides a useful background to the consideration of how to set priorities for research using information on *what is needed and what can be done*. Optimizing the contributions of research entails five steps: (1) determine the specific objective, (2) identify alternatives to address the objective, (3) choose a method by which to set priorities, (4) apply the selected method to quantify priorities, (5) allocate available funds among the priority alternatives. Finally, it is important to take a long-term view and continue supporting the research long enough to make a difference. The paper discusses these steps, illustrates how such an approach might be applied and demonstrates the importance of applying economic criteria to research resource allocation.

Introduction

There are nearly 800 million hungry people including 185 million seriously malnourished preschool children in the developing world. All lack adequate food, water and protection from food-related disease, but without the great strides that have been made in reducing hunger in Asia and Latin America over the past 50 years, there would be millions more. Unfortunately, progress has not been achieved everywhere; in many African countries food output per person has fallen over the last decade and in India and Bangladesh large numbers of hungry people remain despite the substantial gains in per capita food production.

Analysis of food production growth of the past 50 years shows that increases in land, labour, irrigation and fertilizer have contributed to the progress that has been made; in addition, intangible factors like efficient marketing systems, dynamic production technology and higher education have played an equally important role in generating long-run growth in agricultural production (Eicher and Staatz, 1998; Hayami and Ruttan, 1985; Mellor, 1966). These intangible factors are the major differences between the low-productivity, traditional production systems that still prevail in much of Africa and the dynamic, high-input, high-output systems that increasingly prevail in Asia. Development assistance has contributed to Asia's ability to keep pace with its demand for food by helping Asian scientists develop suitable new agricultural technology; appropriate development assistance could help Africa begin the same process and agricultural research could play a part (Sachs, 2005).

[☆] Prepared for the International Epidemiology Workshop, Rennes, France, April 13, 2005

This paper briefly reviews the process by which most of the world has achieved food security, identifies the focus of remaining needs, sketches the contributions of development assistance and considers how agricultural research priorities should be established. It discusses five steps involved in priority setting: determining the objective, identifying alternative activities, choosing a method for setting priorities, implementing the method, and allocating resources among the alternatives according to the priorities established. The importance of using economic considerations along with crop loss estimates for establishing research priorities is demonstrated.

Food security and development assistance

The long view

Before the middle of the nineteenth century 'hunger and premature death' was the norm for most of humanity. Eighty percent or more of all people in Europe, North America and elsewhere in the world were farmers, as in today's poor countries. A gradual process of agricultural development and demographic transformation from 1800 to 1950 was required to largely achieve food security in industrialized Europe and North America nearly doubling life expectancy from 35 to 68 years (Fogel, 2004). But in 1950 'hunger and premature death' was the norm in Egypt, India, China and most of the rest of the developing world, with life expectancy hovering around 40 years. Hunger, stunting, nutritional deficiencies and diseases were widespread. Then around the middle of the 20th century this began to change for many poorer countries; average food consumption increased by 20%, real prices of food fell despite a doubling of population and life expectancy increased from 40 to 64 years (FAO, 2002). This remarkable achievement took one-third the time required by the industrialized countries of the north. Chronic food shortages, as manifested in protein-energy malnutrition, fell in much of Asia and Latin America, in large part because grain yields and farm incomes increased through a very similar process used in Europe, North America and Japan during the previous 150 years. Products of the industrial and scientific revolutions were applied to food production; farm incomes grew and per capita food supplies increased (Johnson, 2000).

Today the developing world has twice the population it had in 1960, 150 million fewer hungry people, real prices of food grain one-third as high, and 20% more food available per person. These great advances in food security resulted from a combination of technology, policies and institutions that encouraged production growth in agriculture. As explained by T.W. Schultz, developing world farmers while poor, use the resources and technology available to them efficiently, but without the innovations in policy, institutions and technology needed to generate the ability to accelerate food production and the incentives to use those innovations they are unlikely to increase production much faster than needed to meet their own needs (Schultz, 1964).

Technology embodied in fertilizer and machinery drove the increases in food security in the industrialized countries from about 1850 onwards. But when high rates of fertilizer were applied to rice grown in the tropics at mid-20th century, they caused the plants to grow rank and fall over rather than produce more grain (Herdt and Mellor, 1964). It took the green revolution of the 1970s to provide new varieties sufficiently productive under tropical conditions to generate a growth spurt in Asian agriculture.

Complementary policies to assure greater security of tenure and more stable prices helped. As farming becomes more technologically advanced it requires capital investments like wells and buildings that are attached to the land. To encourage farmers to make such investments they must have assured rights to the land; alternatively, governments may invest in irrigation systems and otherwise subsidize agricultural investments. The institutions that assure land rights, incentive prices and a steady stream of new technology are critical for agricultural development. All these requirements can only be achieved in a stable, non-oppressive political, social, and economic context. Hence well-functioning governments that understand the importance of agriculture, make the necessary investments in agricultural infrastructure and human capital, and encourage a balance between markets and the state, are critical (Hayami, 2001).

International contributions

Crop yields in the developing world have increased substantially over the past 50 years. Wheat, maize and rice yields have more than doubled in most regions with greater increases in Asia and Latin America than in Africa; yields of other crops like sorghum and potato have also increased significantly. New modern crop varieties together with fertilizer and irrigation drove these gains; the greater the adoption of these technologies, the greater the yield increases. Fertilizer consumption grew at over 10% per year in the 1990s and reached 225 kg per hectare of arable land in East and South East Asia and 110 kg in South Asia; it was stagnant in Sub-Sahara Africa at less than 10 kg in 2000 (FAO, 2005a). From 1960 to 2000, public breeding programmes in over 100 developing countries released over 8,000 new varieties of the major food crops (Evenson and Gollin, 2003). More than 35% of these varieties were based on crosses made at the international agricultural research centres funded by development assistance and many of the others were made by plant breeders trained at those centres or stimulated to emulate or exceed the achievements of the centres. Since the 1990s private, local and international seed companies have begun creating varieties for developing countries based on 'platform' varieties generated by these public sector breeding programmes.

In sub-Saharan Africa there were limited contributions from the green revolution. New varieties of most crops did not exceed 30% of planted area and fertilizer application rates remain at five to 10% of the levels used in Asia. Much of the output increase that did occur was achieved by extending the area under cultivation and mining the soil of plant nutrients through shorter fallow periods. Food production did not keep pace with population growth and a decade-long drop in per capita food production continues. Today, Africa faces a food crisis and an environmental crisis, both resulting from low input, low yield agriculture.

While technological change was central to agricultural development, aid for technological change has been a small fraction of agricultural aid and agricultural aid has been a small fraction of total aid. Between 1973 and 2005, total development assistance varied between about $40 billion and $60 billion annually in 2002 dollars, according to data compiled by the Organization for Economic Cooperation and Development (OECD Development Assistance Committee). Development assistance to *agriculture* from all wealthy countries grew from $4.7 billion in 1973 to over $12 billion per year in 1983–87 but since then has fallen back to about the 1973–1977 amounts (Table 1). In the most recent period around one-quarter of all aid to agriculture went to what OECD calls 'agricultural sector policy, planning and programmes; aid to agricultural ministries; institution capacity

Table 1. Official development assistance to agriculture sub-sectors, annual average constant $ 2002, million

	1973–1977	1978–1982	1983–87	1988–1992	1993–1997	1998–2002
Agricultural policy & administration	421	359	857	1468	562	1614
Agricultural water resources	1097	2207	2114	1699	1061	660
Agricultural development & general	735	1251	2307	1188	1081	647
Forestry, not research	149	369	613	880	468	354
Crop production	331	1173	1028	724	388	258
Fisheries, not research	192	471	400	408	285	235
Research	63	275	456	375	184	201
Agricultural inputs	313	684	552	317	309	186
Agricultural land resources	204	253	795	417	271	178
Agricultural finance and coops	425	1127	1549	895	209	132
Extension	104	235	514	230	77	99
Livestock production + vet services	274	379	331	312	124	94
Agricultural services	426	544	1035	840	167	71
Agrarian reform	0	38	31	440	143	63
Total agriculture	4735	9371	12596	10201	5353	4813
Food Aid (not included above)	2681	2858	3000	1502	524	1383

Source: Extracted from OECD (2005), deflated by the total DAC deflator.

building and advice; unspecified agriculture.' The second largest amount went to 'agricultural water resources,' and the third largest to 'integrated projects' and 'farm development.' Agricultural research has received modest support over the years, seldom exceeding 10% of agricultural aid. It is, however, impossible to identify the development assistance support for the plant sciences, which presumably is a fraction of research.

The effect of agricultural development assistance

The effect of this development assistance varies across sub-sectors of agriculture. Irrigation and drainage projects were the largest sub-sector for thirty years through the mid-1980s and evaluation indicates aid for irrigation was usually effective. USAID evaluations of irrigation projects showed that while many problems had to be overcome, the results encouraged continuing investment. For example, a report that summarized the agency's 30 year experience, including evaluations of AID's projects in Sudan, Senegal, Egypt, Morocco, Turkey, Pakistan, Korea, The Philippines, and Indonesia avoided any quantitative generalizations about the rates of return to the aid investments but indicated that while there was evidence that irrigation's contribution to rice yields accounts for about 30% of the factors involved in the Philippines, it is dangerous to generalize about the returns for other areas or other crops (Steinberg et al., 1983).

A 1995 review of irrigation project evaluation by the World Bank focused on 208 Bank-funded irrigation projects. Evaluations rated 67% satisfactory, comparable to the average of 65% for all Bank-supported agriculture projects but worse than the average of 76% for all Bank projects (World Bank Operations Evaluation Department, 1995). A later review of the Bank's strategy for water management summarized results for 336 World Bank water projects completed from 1988 through to 1999 and indicated that their performance was below the Bank average, based on the assessment of project results along three related dimensions – outcome, institutional development impact, and sustainability of project benefits (Pitman, 2002). Just over 40% had satisfactory ratings in 1988; that increased to 53% by 1996. By the 1990s the World Bank considered water projects as part of the social support system rather than as investments intended to generate additional income and low economic rates of return were of much less concern than several decades earlier.

Integrated agricultural or rural development projects made up the second largest area of agricultural development assistance in the 1960s and 1970s. USAID experience was positive but after emphasizing such projects in many countries for about a decade, they fell out of favour (Kumar, 1987). Such projects achieved roughly the same rate of 'success' in World Bank evaluations as irrigation projects. In 1993 the World Bank's data indicated an overall success rate of 49% for such area development projects (World Bank Operations Evaluation Department, 1993). On average they generated a 10.4% economic rate of return, with just over half giving an economic rate of return over 10% (the other half characterized as 'failures' because they produced below 10%). Failures in area development projects were most frequent in Eastern and Southern Africa. Area development projects went out of favour in the 1980s but recently have reemerged in the form of participatory rural development and poverty alleviation work.

Projects to provide subsidized credit and build agricultural cooperatives comprised the third largest proportion of development assistance to agriculture – over 10% of the development assistance portfolio in the 1970s and 1980s. A summary view of experienced analysts based on many evaluations of such projects found that despite the optimistic expectations of their sponsors, the results of such programmes were disappointing. Loan-default problems were serious, poor farmers remained unable to obtain loans, and those who did get credit were often unnecessarily and inequitably subsidized. Many agricultural banks and other specialized formal lenders serving rural areas were floundering as a result of the requirements of the programmes and as a result often limited the range of services they provide (Adams et al., 1984; Meyer and Nagarajan, 1996). Credit projects lost favour in the late 1980s and 1990s and currently make up less than 3% of the agricultural assistance portfolio.

Assistance to agricultural research absorbed around 4% of agricultural development assistance over the past 25 years. Many analytical estimates of the economic rates of return to agricultural research have been made and, contrary to the

conclusion reached for other kinds of agricultural assistance, over 95% of the studies show substantial positive economic return on investments (Alston et al., 2000). Careful examination of nearly 300 studies reporting over 1800 individual rates of return indicate no support for the idea that returns have fallen over time, but rather that returns vary in other ways that make intuitive sense. In particular, research on commodities with longer production cycles like livestock and more diffuse effects like natural resource management have lower rates of return. Overall, the median rate of return to agricultural research investments is nearly 50% and the median rate of return to research and extension combined is nearly 40%. Studies examining the relationship of agricultural growth to research, education, roads, and other important factors in India and China reinforce the importance of research for growth (Fan et al., 1999, 2002; Rosegrant and Evenson, 1992). Much attention has been focused on variety development but it is clear that pathology, entomology, epidemiology and other plant sciences play an important role in the development of new crop varieties.

Establishing priorities for plant science research

Plant scientists interested in contributing to food security in the developing countries face a simple question: 'What should we do?' International assistance can most effectively address research questions while control of epidemics is largely the responsibility of national authorities and except for certain critical pests like desert locust. In contrast, research has been one of the most effective areas of development assistance. The question of how assistance might be allocated to various research options in the plant sciences is the subject of the balance of this paper.

The question of 'how' to allocate research resources is difficult to separate from 'who' should allocate them and there are two views on who should set priorities. One holds that priorities should be set by those who do the research while the other holds that priorities should be set by those who benefit from it or by those who pay for it. But 'who' largely implies 'how.' If researchers decide, they will favour what they believe they can most effectively do. If users decide they will favour research on their 'most important' unsolved problems; but if researchers have no way to address the unsolved problems, there can be no effective research. On the other hand if researchers discover something for which users have no need, it is of no value (although it adds to knowledge and may be valuable 'basic' research).

Having users decide seems eminently reasonable, but in the case of publicly funded research, becomes circular as the bureaucracy involved in directly funding research seeks the optimal allocation by appealing to both the users and doers of research. Dalrymple (2005) provides a useful discussion distinguishing between researchers who provide the supply of scientific goods and the users who represent the demand for such goods. The best approach would take both positions into account, perhaps through a process something like that reflected in Figure 1.

Figure 1 identifies a 'political-bureaucratic structure' that interprets the latent demand for innovations generated by farmers, consumers, processors and other actors in the ' socio-economic structure.' This political-bureaucratic structure might also be characterized as a decision-maker who generates the actual demand for innovations. This structure distributes funds to the 'innovation-producing institutions' that pay researchers to conduct research and thereby generate the supply of innovations. As those innovations are used by the socio-economic structure they generate actual payoff. The supply and demand analogy has some appeal, but even the elaborated view depicted in Figure 1 breaks down because there is no equilibrating price mechanism for publicly funded research so supply and demand are not the right terms. Nonetheless, it seems clear that the two aspects – what new knowledge is *needed by users* and *what can be done* – should be considered in setting research priorities. In the procedure outlined here, both are. Five steps are required to produce an answer to the question of how resources should be allocated:

1. Determine the objective,
2. Identify alternatives, assemble data for each,
3. Choose a method for setting priorities,
4. Establish priorities among the alternatives,
5. Allocate available resources among alternatives.

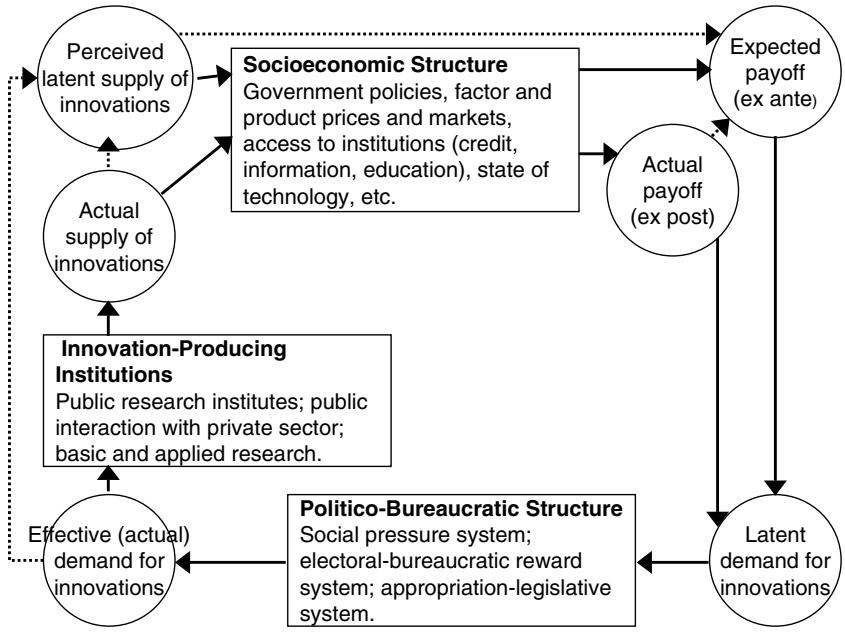


Figure 1. Generalized model of the supply and demand for technological innovation in the public sector.
Source: Adapted by Dalrymple (2005) from: Alain de Janvry, "Social Structure and Biased Technical Change in Argentine Agriculture," in Hans Binswanger and Vernon Ruttan (eds.), *Induced Innovation: Technology, Institutions and Development*. Johns Hopkins University Press, Baltimore, 1978, pp. 301–303. Original referred to both technological and institutional innovation.

Determine the objective

The objective of the whole exercise is presumed to be to put available plant science research resources to their 'best use.' That is, to optimize the use of those resources in generating 'something' – but, exactly what is the something? One option might be to maximize *added output and loss prevented for all crops*. But crops are different, so it is not appropriate to simply add together the prevented losses of grains, vegetables, coffee, and cotton. It is easy enough to aggregate different crops by valuing each agricultural product at its price and adding to get total value of agricultural output. However, production of different crops requires different inputs. For example, it takes much more capital to produce a ton of wine grapes than a ton of rice; more to produce a ton of coffee than a ton of lentils. Such differences in costs suggests the objective to be maximized might be the value remaining after subtracting the cost of inputs, in other words the net value of farm output or *net farm income*.

Some advocate that publicly-supported research should have a special focus on the poor, arguing that maximizing net farm income of the poorest people should be the primary objective. One might accommodate this concern by weighting the income of the poor more heavily in setting priorities, or consider only the farm income of the poorest farmers, ignoring the income of others. If so, it is important to define who the poor are. Some analysts focus on the one-fifth of the population with the lowest incomes – the low income quintile. But is this the low income quintile in each country or in the developing world as a whole? An alternative is to consider the contribution to equity – that is the *income of the poor relative to the wealthy*. This is sometimes done by considering the ratio of incomes of the lowest income quintile to that of the highest income quintile. The Gini coefficient is a measure of equity that reflects the relative income of all units in the population, not just the highest and lowest quintiles. However, it is seldom practical to use the Gini coefficient because of the difficulty in obtaining the data to calculate it.

One might prefer to focus directly on the contribution of research to *nutritional adequacy* of the poor. Like the income of the poorest quintile, this avoids data problems associated with measuring equity but introduces complications associated with defining nutritional adequacy. How can

improvements in calories be aggregated with gains in vitamins or increased protein intake? An index of contribution to nutrition might be devised but in reality the contribution of any particular food to any individual's nutrition depends on that individual's current nutritional status, and, using such a set of weights for aggregating different nutrients is no less arbitrary than applying a set of prices to aggregate across commodities and involves many more computational steps.

These complexities and others lead to using monetary terms to value the productivity of research, which are in any case needed to account for input costs. It introduces the challenge of defining the price for each commodity and input. Commodity prices are different in every territory and most fluctuate on a day-to-day basis and over longer periods. Surprisingly, there is no readily applicable set of international prices by which to value agricultural commodities. The World Bank tabulates monthly prices for rice, maize, wheat, soybeans, rubber, sugar, but not for all agricultural commodities.

Another issue is which price along the marketing chain should be used for aggregation. Price to producers differs from price to consumers by the amount of marketing costs. Marketing costs are likely to be relatively similar among the grains but marketing costs for perishable fruits differ significantly from those for grains. Low income consumers may have different relative values for grains and fruits than high income consumers. A commodity's value in one country may differ from its value in another. Finally, the poorest consumers use a high proportion of their incomes simply to obtain food so the purchase price of food is an important factor in the real incomes of the poor.

For the purpose of this discussion it is assumed that the issues identified above are resolved and there is agreement among the stakeholders and decision maker that the objective to be maximized is the *contribution of each alternative to the real net income of the lowest income quintile in the least developed countries*. For convenience, call this the **real income of the poor.**

Identify alternatives

It is essential to begin with a comprehensive list of research alternatives. The allocation process requires that similar information on all options be considered together. An omitted option cannot simply be added later because all interact and, depending on relationships, an allocation to a new option does not necessarily reduce all others in the same proportion. The scope of alternatives will depend on how broadly one defines the problem. For example, if plant science is taken to include the economics of plant protection, such matters must be included as alternatives. If sociology research on movements to ban chemicals in favour of green agriculture is an alternative that might be funded by the decision maker, then research on such topics must be on the list. If plant breeding for genetic resistance is an option, it must be included. Whether such topics should be included in allocating research resources for plant sciences is a prior decision. Here we make the assumption that the universe of alternatives can be defined along the dimensions of: pest or causal organism, crops/host plants, locations, tasks and approaches.

Manageable interest

In abstract terms, there is almost no limit to the alternatives that might be considered. In practical terms, however, one should restrict the alternatives considered to the set over which the decision-maker has a manageable interest. A manageable interest is the set of issues over which a decision-maker can make and implement a decision. In other words, a decision-maker with responsibility for one province has a manageable interest in alternatives for that province and should restrict considerations to alternatives within the province, while a decision maker with responsibility for a nation must deal with all alternatives relevant for that nation. Likewise, a decision-maker responsible for cereals should deal only with alternatives relevant for cereals, while one with responsibility for all crops has a much larger set of alternatives.

In recent times agricultural research decision-makers have become more attuned to the views of a broad range of people and groups who express interests in food-related matters because of their interests as consumers or simply as members of civil society. These groups, together with farmers, food processors, researchers, taxpayers, research organizations and others are considered as 'stakeholders' in the decisions made about the allocation of public resources and decision-makers often seek ways to incorporate stakeholder views into both the definition

of alternatives and in setting priorities among the alternatives.

Plant hosts

One dimension defining the universe is the set of crops or plant hosts that will be included. It is presumed that the interest is with plants of economic importance, but this covers a wide range. Even limiting consideration to agricultural crops is challenging because there are many 'minor' crops that are of importance in some particular situations. A recent global effort to define plants of international agricultural importance resulted in a list of 64 species (FAO, 2005b). For this exercise the number of host plants is called: H.

Organisms included

A second dimension is the set of pest and disease causal organisms to be included. That is, are all plant diseases to be included – bacterial, viral, fungal and idiopathic? What about nematodes? Will vectors of all diseases be included or only vectors of major diseases? If vectors carry human or animal diseases as well as plant diseases, will research on those animal diseases be included as part of the allocation problem? Will priorities be defined strictly for plant diseases or will non-vector insects and weeds be included? In reality it is difficult to separate out these causal organisms, especially because when a new epidemic breaks out the causal organisms are largely unknown and in many cases a single event has multiple causes. For convenience of discussion call the number of causal organisms: N.

Geography

The third dimension is geographic: over what set of agroecologies, countries or territories are the allocations to be made? Assuming an interest in developing countries, are all developing countries to be covered? The World Bank defines least developed, low-income and middle-income developing countries. Should only countries with a defined minimum amount of crop land be included? Should the former Eastern Bloc countries included? Given the importance of climate in plant diseases, one might argue that it makes most sense to use agroecological regions. Logical though it is, the problem introduced by this is that most data are available for political regions and must be transformed into agroecological categories if they are to be used in that way. For our discussion the number of territories is called: G.

Possible research activities

Contemporary efforts to understand the challenges plant diseases pose to the global food supply roughly follow the above approach of identifying the gains (and losses prevented) from controlling specified sources of loss on specified plants in specified countries. For example, the objective of one ambitious study on the subject reports the scale of losses caused by plant pathogens, animal pests and weeds on eight crops in seven global regions (Oerke et al., 1994). It seems appropriate to follow this lead and define research activities through the target intersections of causal organism, plant host and location. For convenience we call the intersections 'research tasks,' and their number is: R.

Hence:

$$R = N \times H \times G \tag{1}$$

Diseases are controlled through host resistance, pesticides and cultural practices, but all three are probably involved in most successful control systems. Each of the technological control approaches may entail distinctly different activities. For example, host resistance may be pursued through conventional plant breeding or through genetic engineering and may be polygenic or monogenic (Sorho et al., 2005). Biological control may be pursued using native or exotic organisms. The technology for each approach requires quite different resources and, most effective control entails several approaches. The number of such technologies is called: T.

The total possible number of *research activities*, A, is then:

$$A = T \times R \quad \text{or:} \quad A = T \times N \times H \times G \tag{2}$$

A useful notation is to allow each of the elements T, N, H and G assume the form of a subscript that runs from 1 to t; 1 to n; 1 to h and 1 to g. Then any individual research activity can be designated as A with the appropriate subscripts, or in general as:

$$A_{\mathrm{tnhg}}$$

The allocation problem is: to determine the priorities for research among all possible research

activities, that is, among all possible intersections of causal organisms, host plants, geographies, and technologies. To get some idea of the magnitude of the allocation task, suppose that for the whole developing world, there are 50 major causal organisms, 25 plant hosts, 10 territories and 5 technologies, then there will be 62,500 research activities among which to allocate resources. This appears to be an overwhelming task, but of course, some combinations will be 'empty' and others will most efficiently be combined into one activity thereby reducing the number of alternatives. Still, the number will be large, requiring a systematic procedure for organizing all the applicable information.

Priority setting methods

Three broadly different methods have been used to set priorities among research alternatives: scoring, congruence and benefit:cost approaches (Norton and Pardey, 1987). Each has a number of variations.

Scoring approaches

The simplest possible approach is to group alternatives in priority categories such as high, medium and low or rate each alternative on a one to five, one to ten, or some other numerical scale that directly indicates priority by the score of each alternative. More challenging is *ranking* alternatives numerically from the 'most important' to the 'least important,' giving each alternative a unique numerical rank indicating its priority.

Often people are not comfortable with a single number because they believe either there is 'no basis' for making such a judgment or there are several different dimensions to alternatives that would rate differently. They prefer scoring or rating the contribution each alternative is expected to make to several *dimensions* and then aggregating those contributions. For example, individual scores could be assigned to dimensions like output, equity, geographic distribution, women's income, food security or others, and those scores aggregated. The aggregation may be through simple addition or alternatively through weighted aggregation. Table 2 shows how such a system might work.

The first section shows the scores for two research activities; one is high on women's income and low on output and food security; the second alternative scores high on output and food security and the same as the first alternative on the other three characteristics. The aggregate is the simple average or the aggregated value using equal weights. The second section illustrates the effect of differential weights where equity and women's income have higher weights and other characteristics have lower weights. With the unequal weights the aggregate score of the first research alternative is much closer to the second, reflecting the greater weights given for two of the characteristics. In this system both weights and the scores each contribute to the aggregate score.

Any number of variations of scoring approaches can be devised. For example one might use data on production in geographic regions to score the geographic characteristic and value of output to score the output variable. A number of possible weighting schemes may be devised; and a large number of different characteristics may be used as weights. The weights can be determined in a separate exercise from the scores so stakeholders can be involved where they have special knowledge or interests (e.g. in the scores) without completely determining the outcome.

The same versatility that permits the introduction of many characteristics and variations on weighting is one of the limitations of scoring

Table 2. Illustration of alternative scoring approaches

	Activity	Output	Equity	Geo-graphic	Women's income	Food security	Aggregate score
Simple average							
Score	1	2	2.5	3	4	2	2.7
Score	2	4	2.5	3	4	4	3.5
Weighted							
Weight		0.1	0.3	0.1	0.4	0.1	
Score	1	2	2.5	3	4	2	3.05
Score	2	4	2.5	3	4	4	3.45

approaches. One must be careful not to design the weights to skew the results in a particular direction and recognize that the greater the number of weighting characteristics, the more difficult it is to trace the links between characteristics, weights and aggregate score (Alston et al., 1995).

Congruence-based allocation

Another approach is based on the view that research resources ought to be allocated in proportion to the 'importance' of each activity as reflected in the value of crop production or, in the case of plant protection, the value of crop losses attributed to various problems. This approach is known as the 'congruence' method of allocating research resources.

Using Y to represent production loss prevented or yield increase obtained, for every A_{tnhg} there is a corresponding $Y_{\text{tnh.}}$ A critical relationship is the contribution each research activity (A_{tnhg}) is expected to make to increasing Y_{tnh} and in turn, the contribution that increased output makes to the incomes of the poorest quintile. Implementing this approach requires information like: applying $\$X$ to A_{tnhg} over a period of Y years will prevent losses or increase production by Y_{tnh} and raise real income of the poor by $\$Z_{\text{tnh}}$ per year over subsequent years.

An obvious starting point is to know the Y_{tnh} – the yield loss or potential yield increase – for each of the A intersections defined in (1) above. Intuitively, crop losses are the amount of crop lost to various pests or because production factors are used at less than maximum output levels. Here the focus is on losses from pathogens and pests. As with most seemingly simple concepts, complexities lie below the surface, as the literature on crop losses makes clear. Figure 2 illustrates this point.

For any crop a physiologically defined theoretical yield potential can be associated with any genotype and climate regime, unimpeded by limitations of water, nutrients and pests. In any practical situation there is some *unavoidable crop loss*, given the impossibility of controlling all factors that lead to losses. This defines an attainable yield. In general, attaining that yield requires expenditures on inputs or control measures in excess of the profitable levels and so one can define *economically non-recoverable loss* and hence an economic yield. That is the level one would expect to observe if all farmers applied all crop loss control measures at the economically optimal level, but generally the actual yield is somewhat below that level. The actual yield reflects the yield response to crop protection actually used that is, the *prevented loss*. Still lower, assuming some effectiveness of current practices is the yield without crop protection. The distance between these differently defined yields reflect the various loss concepts.

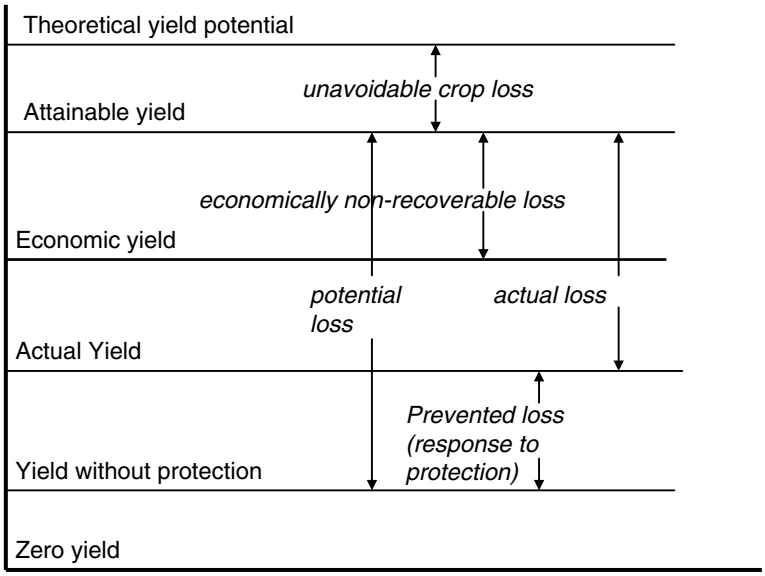


Figure 2. Conceptual model for crop loss assessments (adapted from Oerke et al., 1994 modification of Zadoks and Schein, 1979).

Of course, except for actual yield, all the yield levels and losses identified above are concepts that cannot be observed under production agriculture. But the concepts are so intuitively helpful that some broadly accepted conventions have been developed that permit estimates to be generated. In studies of crop losses, 'attainable yield' is defined or computed using crop growth models taking into account the climate, water availability, yield potential of varieties grown, rates of fertilizer application and other cultural techniques like seedbed preparation and crop density (Oerke et al., 1994). The difference between the estimated attainable yield and actual yield provides an estimate of '*actual loss*' attributed to pests and pathogens. It is also possible to estimate the '*prevented loss*' from knowledge about the plant protection measures used and their effectiveness. In many cases the results from plant protection experiments are used in making such estimates.

An alternative, more participatory approach to using crop modelling and experiments is to draw on the knowledge of farmers who are producing crops in the areas of interest. Clearly, those who make their livelihoods through farming have an interest in anything that reduces yields and they would seem an important resource for identifying yield losses and potentials for increases. As intuitively appealing as this is, a review of the literature reporting such activities identifies at least six limitations (Dalrymple, 2005). First, farmers are likely to be highly influenced by their immediate and highly visible problems and are likely to have a short-term outlook and be less concerned with or aware of the opportunities offered by longer-term research. Second, farmers are more likely to favour research that generates benefits they receive rather than broadly-adapted research that generates product price reductions and benefits to consumers. Third, in most developing nations elites dominate and will naturally direct attention to research that favours them over less powerful groups. Fourth, those who favour participatory approaches in setting priorities generally ignore consumers and consider only farmer participation, despite the evidence showing that consumers are the main beneficiaries of much agricultural research. Fifth, as the geographic scope of the allocation exercise is enlarged it becomes increasingly difficult to get a comprehensive and unbiased view from farmers and consumers. Finally, the wide diversity of clientele and the complexity of the systems necessary to integrate the number and diversity of client views make such approaches inherently difficult to structure. These difficulties could, of course, be overcome and estimates of the relative importance of conducting research in each of the A_{tnhg} could be generated using participatory methods. In practice, more efforts seem to have been devoted toward participatory research than toward participatory priority setting.

Estimated crop losses

An important stimulus to crop loss measurement was given by several major symposia on the subject in the last century. The first was organized in the 1960s by the Food and Agriculture Organization (FAO, 1967) and a second took place in the 1970s in honour of E.C. Stakman (Teng and Krupa, 1980). A more recent study of crop losses (Oerke et al., 1994) provides access to a large amount of systematically organized information. This work was designed to stimulate research on the causes of losses, improve methods to protect crops, enhance the effectiveness of control methods, integrate plant protection with other management practices to optimize methods of crop production, and help generate support for research on effective crop protection. Table 3 provides a summary.

The immense amount of work and detailed, country-by-country, crop by crop information that lies behind the table cannot be overstated. Based on that work, global crop losses are estimated to be about 75% as large as actual production, with the losses almost equally attributed to pathogens, pests and weeds. The lowest estimated losses, about 30%, are in Europe and North America, while Africa and Asia each are estimated to lose nearly 50% of their attainable production. Nearly 60% of global losses occur in Asia, far more than in any other region. This is because Asia produces nearly half of global agricultural production and has a higher rate of loss than other regions.

Using a simple congruence approach to set priorities based on these data would suggest that 60% of research resources should be allocated to preventing losses in Asia and within that allocation, the resources allocated to pathogens, animal pests and weeds should be in the ratio of approximately 14:18:14, the proportion of loss to the three main causal agents. The balance of available resources

Table 3. Estimates of crop losses, in financial terms (US$), occurring during the production of the eight principal food and cash crops in the years 1988–1990, by continent

Continent	Actual production		Loss (%) of production due to			Loss, overall		
	$ bn.	%	Pathogen	Animal pests	Weeds	%	$ bn.	% of global total
Africa	13.3	4.0	15.6	16.7	16.6	48.9	12.8	5.3
N. America	50.5	15.1	9.6	10.2	11.4	31.2	23.0	9.4
L. America	30.7	9.2	13.5	14.4	13.4	41.3	21.8	8.9
Asia	162.9	48.6	14.2	18.7	14.2	47.1	145.3	59.6
Europe	42.6	12.7	9.8	10.2	8.3	28.2	16.8	6.9
U.S.S.R.	31.9	9.5	15.1	12.9	12.9	40.9	22.1	9.1
Oceania	3.3	1.0	15.2	10.7	10.3	36.2	1.9	0.8
Total	335.2	100.0	13.3	15.6	13.1	42.1	243.7	100.0

Source: Oerke et al. (1994, p. 749); final column added by the author of this paper. Prices used in valuing production (by Oerke et al.) were: wheat: US$ 136.2/t; rice: US$ 209.1/t; barley: US$ 79.5/t; maize: US$ 98.1/t; potatoes: US$ 128.7/t; soybeans: US$ 236.1/t; cotton: US$ 490.6/t; coffee: US$ 1934.4/t.

would be allocated similarly, following the proportion of losses in each region to each source.

One can argue that congruence with crop losses is a good approach if dollars spent on every A_{tnhg} have the same effectiveness in increasing the incomes of the poor (given that objective). This is likely to hold if no A_{tnhg} are 'harder' or 'easier' than others so that a given amount of money spent on each would make the same contribution to the value of each crop and that each commodity makes the same contribution to real incomes of the poor. But, the relationship between research input and loss prevented is likely to be complex – some research challenges are harder than others. The research continuum from basic through applied to adaptive implies that a higher degree of uncertainty is associated with 'more basic' research activities that are likely to take longer. 'More difficult' research problems are likely to require more resources and time to generate usable results but are also likely to have the potential to generate higher returns. The contribution of losses prevented for a commodity important in the consumption of the poor or in generating income for the poor is more important for the objective than for a commodity not important to the poor. Technologies not well-suited for adoption by the poor would contribute less than those especially well-suited.

Input–output function of A_{tnhg}

The relationship between research input, expressed in researcher time and funds, and the expected findings or 'solutions' that prevent loss, is defined as the research input–output function. The input–output function for each A_{tnhg} should reflect the difficulty and time required to find a solution through that activity; input–output functions for different activities will reflect differences in the difficulty or time needed for various A_{tnhg}. The input–output function provides an estimate of what research may actually contribute towards the objective while crop loss estimates represent the opportunity for research to contribute – these are the two key factors: *what can be done* and *what is needed*. In Dalrymple's terms, the input–output function reflects the supply of research findings while increased real incomes of the poor from the crop loss thereby prevented reflect the demand for research findings.

To illustrate: a set of pesticides can be screened for their effectiveness against a particular pest in a relatively few growing cycles, say in a matter of 2–5 years. If the pesticide has been approved, a control practice can be recommended to farmers shortly thereafter. An alternative approach, the development of cultivars with genetic resistance to the pest, may take 6–10 years from the beginning of research to release to farmers. Even if the two activities give the same yield effect and remain effective for the same period, they have different input–output functions. Some kinds of research may have a greater inherent requirement for inputs such as laboratory equipment, experimental fields and labour; costs of land, labour and capital vary across locations and other factors affect the cost of

any particular research activity, but these differences can be incorporated in an input–output function.

A number of things can be inferred about the relationship between research input and expected prevented yield loss or expected output (Y_{tnhg}). First, at the beginning of the process and with zero input, expected prevented loss is zero. Second, no matter how great the resources or time taken, there is some maximum value of expected Y_{tnhg} depending on actual losses or yield potential for each $A_{tnhg,}$. Third, at the beginning of the research with small inputs the probability of finding a successful 'solution' is small and therefore the expected prevented Y_{tnhg} is small. The expected value likely increases slowly until some critical minimum amount of resources are applied and at that point increases rapidly over some range of research inputs. Beyond some level of resources the expected Y_{tnhg} is likely to begin to increase at a declining rate.

Figure 3 illustrates a few of the many possible input:output relationships consistent with the inferences stated above. Each curve portrays the relationship for a different research activity designed to prevent losses experienced in a particular intersection defined by equation (1). Research input consists of money, people and time reflected on the horizontal axis as cost per year. Curves A1 and A2 use the same research approach but with more resources applied each year in the case of A1, so the solution is expected to be found sooner. Both are expected to generate knowledge to prevent the entire loss. Curves B1 and B2 represent a different approach expected to be less successful in preventing the yield losses; B2 is expected to take less time than B1 but the latter will ultimately prevent more of the losses, although not as much as the approach used in A1 and A2. Intuitively, such ideas seem consistent with the ways scientists think about research alternatives and incorporate more information in the allocation process than the congruence approach.

Benefit:cost

In addition to incorporating the effects of different input–output relationships, benefit:cost approaches can incorporate variations in resource use, time lags and uncertainty into priority setting. To illustrate, the following example may be helpful. Suppose two different research activities could be targeted at preventing a $900,000 annual crop loss. Suppose the first, A_{tnh1}, costs $50,000 a year, will be completed in 5 years and is expected to prevent half the potential loss while A_{tnh2} costs $25,000 a year, will take 10 years and is expected to prevent 80% of the potential loss after 10 years. In both cases the technologies are assumed to remain effective for 10 years after introduction. They are illustrated in Table 4.

The first line in the Table shows the research cost per year (all numbers in '000). For simplicity cost is assumed to be constant for a defined number of years, but that assumption is easy to relax. The research is aimed at preventing the potential loss depicted in the second line. The percent expected prevented loss shown in the third line is the concept introduced in Figure 3, in percentage

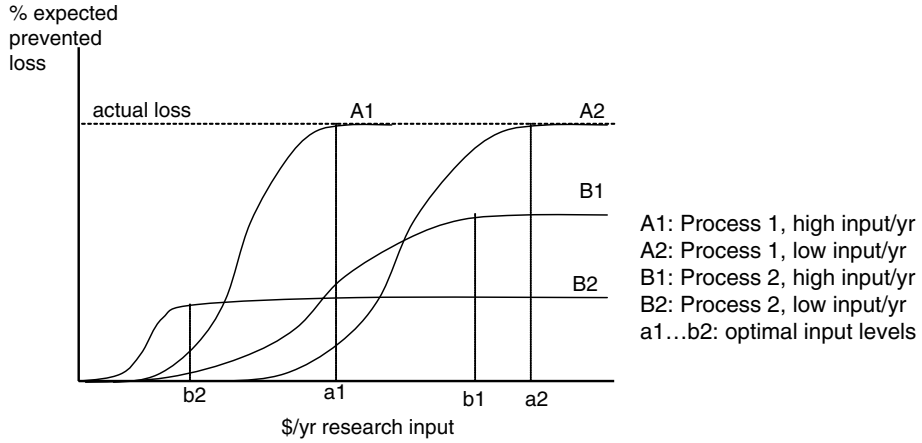


Figure 3. Hypothetical research input/output relationships.

Table 4. Illustration of the calculation of present value of net benefits of two research alternatives at a discount rate of 10%

Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Alternative 1																				
1. Cost	50	50	50	50	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. Potential loss	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900
3. % exp. Prev loss	0	0	0	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0	0	0	0	0
4. Loss prevented	0	0	0	0	450	450	450	450	450	450	450	450	450	450	450	0	0	0	0	0
5. Adoption %	0	0	0	0	0	0.3	0.6	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9					
6. Expected LP	0	0	0	0	0	135	270	405	405	405	405	405	405	405	405	0	0	0	0	0
7. Exp. net benefit	−50	−50	−50	−50	−50	135	270	405	405	405	405	405	405	405	405	0	0	0	0	0
8. PV of net benefit	−45	−41	−38	−34	−31	76	139	189	172	156	142	129	117	107	97	0	0	0	0	0
NPV	1134																			
Alternative 2																				
1. Cost	25	25	25	25	25	25	25	25	25	25	0	0	0	0	0	0	0	0	0	0
2. Potential loss	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900	900
3. % exp. Prev loss	0	0	0	0	0	0	0	0	0	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
4. Loss prevented	0	0	0	0	0	0	0	0	0	720	720	720	720	720	720	720	720	720	720	720
5. Adoption %	0	0	0	0	0	0	0	0	0	0	0.3	0.6	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
6. Expected LP	0	0	0	0	0	0	0	0	0	0	216	432	648	648	648	648	648	648	648	648
7. Exp. net benefit	−25	−25	−25	−25	−25	−25	−25	−25	−25	−25	216	432	648	648	648	648	648	648	648	648
8. PV of net benefit	−23	−21	−19	−17	−16	−14	−13	−12	−11	−10	76	138	188	171	155	141	128	117	106	96
NPV	1161																			

terms for ease of computation. For a greater research cost the expected prevented loss will be achieved sooner and possibly to a greater extent so that for every different research input:output relationship the values of the first and the third lines of the table will be different. The fourth line is the quantity of loss prevented (LP), the product of lines 2 and 3. The time path of expected farmer adoption is incorporated in line 5 of the analysis and reflects the percent of line 4 that is realized each year. If adoption of the results of various A_{tnhg} takes different time paths, incorporating the adoption lag will differentially affect the benefits obtained in any year. The sixth line is the product of lines 4 and 5 and reflects the expected amount of loss prevented each year. Multiplying this line by the output price would give value but for simplicity here we assume a price of 1 per unit of output. The seventh line is line 6 minus line 1, the value of expected net benefit in the year in which benefits are expected to be obtained. Net benefits are, of course, negative in years 1 through 5 until results are obtained; zero in year six when the research is completed but the results have not yet been adopted; and are assumed to reach 30%, 60% and finally 90% over the next three years as adoption occurs. Thereafter net benefits are constant for the next 6 years after which the technology is assumed to be obsolete and no longer prevents yield losses (one could, of course, have a technology with slowly declining benefits). Line 8 shows the 'present value' of expected net benefits, a concept designed to reflect the idea that preventing losses sooner rather than later is better and that money spent today in preventing such losses is worth more than money spent for the same purpose in the future. The 'discount rate' reflects the relative value of solutions today compared to solutions in the future. If it makes no difference when the solution is obtained the discount rate is zero, but in most cases, people would prefer solutions sooner rather than later and the stronger that desire, the higher the discount rate. Of course, all the relationships and parameters in the table are illustrative and the result of simplifying assumptions.

The top part of the table illustrates the early return case: \$50,000 is expended each year for 5 years, at which point the expected results are obtained. In years 6 through 15 the loss prevented could be \$450,000 but because adoption takes time the expected loss prevented is as shown in line 6. The expected net benefit, the difference between cost and expected LP, is equal to cost in the first 5 years; afterwards costs go to zero and net benefit is the loss prevented. The values in line 7 are

'discounted' at 10% to get the present value (PV) in line 8 and those values are added together to get their sum for the entire period, called the net present value or NPV. In the bottom panel alternative 2 is shown. In terms of Figure 3, alternative 1 might be represented by a curve like B2 while alternative 2 is more like curve B1.

In the numerical example the NPV of the first alternative is 1134 and the NPV of the second is 1161. Thus, even though the first alternative has a shorter time to solution (5 vs. 10 years), the second has a slightly greater present economic value, in part because it is lower cost (25 vs. 50) and in part because it prevents a greater proportion of the potential loss (80% vs. 50%).

Each of the numbers that goes into the computation has an impact on the NPV. For example, if the discount rate is 5% rather than 10%, the NPV of the first alternative is 1936 and of the second is 2506 – with a lower discount rate the NPV of the first alternative increased by 70% and that of the second by 115%, illustrating the differential effect discount rate has on more distant income. In the extreme case, if future costs and benefits are not discounted then the NPV is simply the sum of the stream of expected net benefits; in this case 3395 and 5582.

With the discount rate at 10%, if the time to success and adoption are shortened by one year and obsolescence still occurs after 10 years, then the NPV of the first alternative is 1266 rather than 1134. If the solution using the second alternative is found after 8 years rather than 10 and adoption and obsolesce patterns are unchanged, its NPV is 1543 rather than 1161. If, in the second alternative, costs increase to 30 in year two, 35 in year 3, 40 in year 4, 45 in year 5 and 50 in year 6 and beyond, NPV becomes 1068 rather than 1161. Hence, different time paths to success or adoption or costs generate different patterns of returns and higher or lower NPV. While the approach seems complex and requires the specification of numerical values to concepts that normally are little more than 'hunches' of scientists, it has successfully been applied to *help* guide resource allocation in a $110 million programme (Herdt, 1991).

Elaborations to benefit:cost

Increases in output that are large relative to current supply may have the effect of reducing the price of the commodity in question. In fact, the global long term downward trend in grain prices has been ascribed to the success of research in increasing the productivity of grain production in many locations throughout the world. In contrast, the prices of food legumes show no such long term decline, in part because there have been relatively modest productivity gains.

By incorporating appropriate assumptions about the way consumers respond to additional supplies (demand elasticities) it is possible to estimate the impact of a productivity gain on prices and incorporate that into the estimates of benefits and costs. In addition it is possible to calculate how much of the productivity gain remains in the hands of producers and how much goes to consumers. Economists call these the producers' surplus and consumers' surplus and commonly use such concepts in estimating the benefits from technological change (Alston et al., 2000); they can also be used in research resource allocation.

An additional elaboration has been developed to accommodate the idea that it is difficult to give a point estimate of the likelihood that any particular research activity will be successful. This incorporates a 'triangular distribution' into the input–output function using estimates of the maximum likelihood of success, the minimum and the most likely probability of success (Mills, 1998). These numbers are then aggregated into a single one used as the 'probability of success' in the table.

Allocate resources among alternatives

As illustrated, priorities can be established in several ways, from categories to rankings to benefit: cost (with or without considering economic surplus) to subjective scores. Regardless of the method the result will be a set of numbers representing the priority of each A_{tnhg}. However, no matter which approach is used, that set of numbers does not imply any specific allocation of resources. Any of the sets of numbers *could* be used to allocate resources proportionately, but each would be arbitrary, given what is recognized about the research input–output function. Alternatively, the numbers generated *can be* interpreted as a *ranking* of importance if a technique is consistently applied, however such a ranking still does not translate into a particular resource allocation.

Various options might be used to translate the priority into an allocation. One extreme would

be to allocate all available resources to the top ranked activity – a position not likely to get much sympathy. In a strict capital budgeting problem where the NPV is independent of the size of the investment, funds would be allocated to the alternative with the highest NPV, then the second highest, etc. until all funds are used up. Another option is to argue that all alternatives should get equal allocations. This may be the most politically appealing and would make much of steps three through to five unnecessary! However, using some concepts from economic theory, one can easily show that a non-equal allocation can generate higher benefits.

Economic theory shows that the optimal pattern of investment would be to invest in each alternative just the amount that provides an equal incremental (in economic jargon, marginal) return to each alternative and at the same time uses up all the available resources. Applying this concept requires data on the marginal return to each alternative, which can be derived as follows. A larger annual investment in a particular alternative is likely to shorten the time until success (although there is a limit to how short the time can get). On the other hand, a smaller annual investment is likely to lengthen the time to success (although if the annual investment gets too small the probability of success may become zero). The following illustrates what larger or smaller annual investments would do for one research alternative.

Suppose that with a larger annual investment ($50,000) the research phase can be shortened from 10 to 8 years and adoption speeded up so that 10% of farmers adopt in the 8th year and 40% in the 9th, etc. As a consequence the NPV would increase from 1161 to 1605 (details not shown). On the other hand, if the annual investment is smaller ($10,000) and the research phase is consequently stretched out to 15 years with a similar relative pattern of adoption as originally, then the NPV falls to 517. Following this procedure one can estimate the NPVs that correspond to a series of different annual research investments for a given research alternative. These values can be plotted as in Figure 4, alternative 2. In a similar way, for each research alternative there exist a series of NPVs corresponding to different levels of research investments.

Figure 4 shows NPV curves for three research alternatives. Alternative 2 is the case we have been following with NPV of 1161 at $25/year, 1605 at $50/year and $517 at $10/year. Although we computed the NPV for alternative 1 only for one investment level in Table 4, other levels would generate a series of NPVs to trace out the curve shown. A similar curve of NPV vs. annual research investment is shown for one additional alternative and could be plotted for every possible research alternative.

Such curves provide the key to solving the resource allocation problem. Notice that each curve is increasing but at some point the rate of increase falls and the curve eventually flattens out as researchers run out of good ideas to investigate and the work becomes less productive. In other words, the increase in NPV for higher and higher increments of investment eventually declines (and may even become zero or negative). Economic theory says that the greatest total expected gain will be obtained when the additional NPV from each alternative is equated and all resources are used. This can be illustrated as follows.

Suppose the research manager has $150 to invest each year among the three alternatives shown in Figure 4. If it is invested equally, $50 in each alternative, Alternative 1 generates an expected NPV of 1134, alternative 2 an expected NPV of 1605 and alternative 3 an expected NPV of 1175. The total of the three is 3914. On the other hand, if instead of equal allocation, alternative 3 gets $75/year its NPV goes up quite a lot (the curve is steep) and if at the same time alternative 1 gets $25/year its NPV goes down by a lesser amount (its curve is less steep).

The changes in NPV are given in Table 5; each column shows the change in NPV from the lower

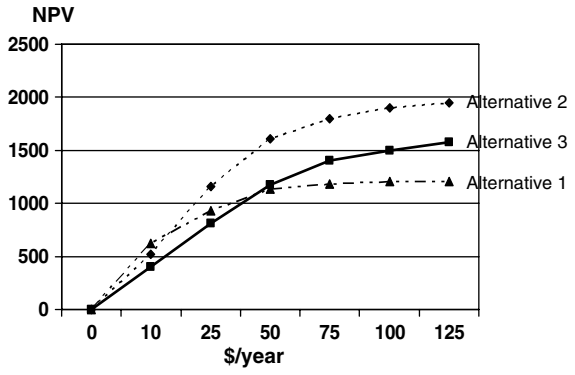


Figure 4. Hypothetical net present value (NPV) of three research alternatives.

Table 5. Change in NPV from changing annual research investment

Investment	0	10	25	50	75	100	125
Alternative 1		625	309	200	50	20	0
Alternative 2		517	544	444	195	100	50
Alternative 3		400	410	365	225	100	75

investment level to the specified one. For example, the first \$10 of investment generates additional expected NPV of \$625 in alternative 1, \$517 in alternative 2 and \$400 in alternative 3. Consider \$50 invested in each alternative: switching \$25 from alternative 1 to alternative 3 reduces expected NPV from alternative 1 by \$200 and increases expected NPV from alternative 3 by \$225 – a net increase of \$25. Of course, if the research decision maker had more funds available, say \$225, it would be better to invest \$50, \$75 and \$100 in alternatives 1, 2 and 3 respectively. The differences are modest because the three curves are quite similar; the more different the input–output curves are, the greater will be the difference in expected output from applying the economic decision rule compared to equal allocation.

In this example, the three alternatives (A_{tnh1}, A_{tnh2}, A_{tnh3}) all apply to preventing losses from a single Y_{tnh} of 900 per year. Incorporating all possible Y_{tnh} and all possible research alternatives to prevent those potential losses would provide a full solution to the allocation problem that follows the rule suggested by economic theory and would maximize the expected value of prevented losses. Of course, implementing such an allocation procedure requires input–output functions for all possible research alternatives and a computational algorithm to solve the entire system. But with modern computers this is possible.

Summary

In 1800 most people in most countries were chronically hungry, life expectancy was 35–40 years and misery was the accepted lot of most people. By 1950 a few countries in Europe and North America had achieved a remarkable improvement in living standards; food production and consumption reached adequate levels for most and people lived to their mid-60s on average. But in Asia, Africa, most of Latin America and the Middle East, things were not much changed from 150 years earlier. Poverty, short lives, high rates of hunger and low-yield, low technology agriculture prevailed except for a few enclaves. A remarkable change has occurred since 1950. The world's population has doubled but there are 150 million fewer hungry people; per capita food availability in the developing world has increased by 20% and the real price of food worldwide has fallen by half. World food production has more than kept pace with growing food demand. Still, there are far too many poor, hungry and ill-clothed people in developing countries, with by far the greatest proportion in sub-Saharan Africa and South Asia.

Technology, policies and institutions designed to encourage economic growth of agriculture and ensure the poor are included in growth are the important necessary conditions to overcome hunger and poverty. Far from being tradition-bound and resistant to change, millions of farmers in poor countries have accepted new technologies in the form of seed varieties, fertilizers and irrigation and driven the rate of food production ahead of the demand for food. Experience shows that such technology must be carefully designed to fit the situations where it is to be used, but once systems for doing such research became operational, a green revolution spread through Asia and Latin America. But the necessary combination of policies, technology and government institutions have proven elusive in sub-Saharan Africa. That part of the word remains the challenge for the 21st Century.

Development assistance from wealthy countries has contributed in significant ways to help improve conditions in poor countries with agricultural research among the most successful of aid efforts. The technology and cultivation practices developed by the international agricultural research centres of the CGIAR spread widely through Asia, the Middle East and Latin America and provided the basis for a green revolution in many countries. While poverty still is the lot of too many, food availability and incomes are much improved.

Plant science research has contributed to the improved management and control of many plant pests and diseases but crop losses continue to claim over 40% of potential production having an estimated value of nearly \$250 billion. Appropriately directed research could develop systems and products to save much of that potential. In order to best allocate available research resources to

address those challenges, decision makers must: carefully define objectives; specify possible research alternatives in quantitative terms; choose among several different approaches for setting priorities; apply the method to establish priorities; and allocate the resources among alternatives. The optimal allocation of research resources can only be established by applying economic principles to estimates of the research input–output functions that quantify how alternatives are expected to prevent crop losses.

Most allocations simply take the previous year's resources and make small adjustments; some allocations use scoring approaches; some allocate resources in proportion to value of production, contribution to incomes of the poor or in proportion to the value of crop losses. These intuitively appealing procedures all have the drawback of failing to take into account either the likely degree of success research may have in addressing each alternative or the importance of each possible solution to poor farmers and consumers.

Economic principles offer tools that can incorporate many considerations important to stakeholders. Allocations that use marginal productivity variations on benefit:cost approaches require large amounts of data and require researchers to make their assumptions explicit. These are difficult to apply and are seldom used. However, they would keep decision makers from overlooking potentially large contributions to the ultimate goal of improving the lives of the poor through agricultural research. The paper demonstrates that two aspects related to future research – *what can be done* and *what is needed by users* of new knowledge – should be considered by any decision-maker in setting research priorities and that using economic principles together with such information generates a higher expected return on research investments than alternative methods.

References

Adams DW, Graham DH and Pischke JDV (eds.) (1984) Undermining Rural Development with Cheap Credit. Westview Press, Boulder, CO

Alston JM, Chang-Kang C, Marra MC, Pardey PG and Wyatt T (2000) A Meta-Analysis of Rates of Return to Agricultural R&D: Ex-Pede Herculem? International Food Policy Research Institute, 113, Washington, DC

Alston JM, Norton GW and Pardey PG (1995) Science Under Scarcity: Principles and Practices for Agricultural Research Evaluation and Priority Setting. Cornell University Press, Ithaca, NY

Dalrymple DG (2005) Setting the agenda for science and technology in the public sector: The case of international agricultural research. Science and Public Policy, forthcoming

Eicher C and Staatz J (1998) International Agricultural Development. The Johns Hopkins University Press, Baltimore and London

Evenson RE and Gollin D (eds.) (2003) Crop Variety Improvement and its Effect on Productivity. CABI Publishing, Wallingford, United Kingdom

Fan S, Hazell P and Thorat S (1999) Linkages Between Government Spending, Growth, and Poverty in India. International Food Policy Research Institute, 110, Washington, DC

Fan S, Zhang L and Zhang X (2002) Growth, Inequality, and Poverty in Rural China. International Food Policy Research Institute, 125, Washington, DC

FAO (1967) Background Papers for the Symposium on Crop Losses. FAO Symposium on Crop Losses. Rome. Food and Agriculture Organization of the United Nations

FAO (2002) World Agriculture: Toward 2015/2030. Food and Agriculture Organization of the UN, Rome, Italy

FAO (2005a) Compendium of Agriculture – Environmental Indicators (1989–1991 to 2000). Food and Agriculture Organizations of the United Nations

FAO (2005b) International Treaty on Plant Genetic Resources. Food and Agriculture Organization of the United Nations

Fogel RW (2004) The Escape from Hunger and Premature Death 1700–2100: Europe. America and the Third World. Cambridge University Press

Hayami Y (2001) Development Economics: From the Poverty to the Wealth of Nations. Oxford University Press, London

Hayami Y and Ruttan V (1985) Agricultural Development: An International Perspective. The Johns Hopkins University Press, Baltimore and London

Herdt RW (1991). Research priorities for rice biotechnology. In: Khush G and Toenniessen G (eds.) Rice Biotechnology CAB International, Wallingford, UK

Herdt RW and Mellor JW (1964) The contrasting response of rice to nitrogen: India and the United States. Journal of Farm Economics 46: 150–160

Johnson DG (2000) Population, food and knowledge. American Economic Review 90: 6–21

Kumar K (1987) AID's experience with integrated rural development projects. USAID Program Evaluation Report No. 19, July 1987, Washington, DC

Mellor JW (1966) The Economics of Agricultural Development. Cornell University Press, Ithaca, NY

Meyer RL and Nagarajan G (1996) Evaluating credit guarantee programs in developing countries. Department of Agricultural Economics and Rural Sociology Working Paper 2322 . Ohio State University, Columbus, Ohio

Mills B (ed.) (1998) Agricultural Research Priorities: Information Investments for the Improved use of Research Resources. International Service for National Agricultural Research, The Hague

Norton GW and Pardey PG (1987) Priority-setting Mechanisms for National Agricultural Research Systems: Present

Experience and Future Needs. International Service for National Agricultural Research, 7, The Hague

OECD Development Assistance Committee (2005) International Development Statistics on-line; CRS online database in aid activities. Organization for Economic Cooperation and Development

Oerke E-C, Dehne H-W, Schonbeck F and Weber A (1994) Crop Production and Crop Protection: Estimated Losses in Major Food and Cash Crops. Elsevier, Amsterdam

Pitman GK (2002) Bridging Troubled Waters: Assessing the World Bank Water Resources Strategy. The World Bank, OED, Washington, DC

Rosegrant MW and Evenson RE (1992) Agricultural productivity and sources of growth in South Asia. American Journal of Agricultural Economics 74: 757–761

Sachs JD (2005) The End of Poverty: Economic Possibilities for Our Time. The Penguin Press, New York

Schultz TW (1964) Transforming Traditional Agriculture . Yale University Press, New Haven

Sorho F, Pinel A, Traore O, Bersoult A, Ghesqiere A, Hebrard E, Konate G, Sere T and Fargette D (2005) Durability of natural and transgenic resistances in rice to *Rice yellow mottle virus*. European Journal of Plant Pathology 112: 349–359

Steinberg DI, Clapp-Wincek C and Turner AG (1983) Irrigation and AID's Experience: A consideration based on evaluations. USAID, Washington, DC

Teng PS and Krupa SV (1980). Assessment of crop losses which constrain production and crop improvement in agriculture and forestry. In: Teng PS and Krupa SV (eds.) E. C. Stakman Commemorative Symposium. University of Minnesota, Minneapolis, Minnesota (pp. 327) Department of Plant Pathology, University of Minnesota,

World Bank Operations Evaluation Department (1993) Annual Review of Evaluation Results 1992. The World Bank, Washington, DC

World Bank Operations Evaluation Department (1995) The World Bank and Irrigation. The World Bank, 14908, Washington, DC

European Journal of Plant Pathology (2006) 115: 95–103
DOI 10.1007/s10658-005-1230-z

Disease assessment concepts and the advancements made in improving the accuracy and precision of plant disease data

Forrest W. Nutter Jr.[1] Paul D. Esker[1] and Rosalee A. Coelho Netto[2]
[1]*Department of Plant Pathology, Iowa State University, 351 Bessey Hall, Ames, IA, 50011 USA (Fax: +1-515-294-9420; E-mail: fwn@iastate.edu);* [2]*National Research Institute of the Amazon, Caixa Postal 478, CEP 69011-670, Manaus, AM, Brazil*

Accepted 13 July 2005

Key words: phytopathometry

Abstract

New concepts in phytopathometry continue to emerge, such as the evolution of the concept of pathogen intensity versus the well-established concept of disease intensity. The concept of pathogen severity, defined as the quantitative measurement of the amount of pathogen per sampling unit has also emerged in response to the now commonplace development of quantitative molecular detection tools. Although the concept of disease severity, i.e., the amount of disease per sampling unit, is a well-established concept, the accuracy and precision of visual estimates of disease severity is often questioned. This article will review disease assessment concepts, as well as the methods and assessment aides currently available to improve the accuracy and precision of visually-based disease severity data. The accuracy and precision of visual disease severity assessments can be improved by quantitatively measuring and comparing the accuracy and precision of rates and/or assessment methods using linear regression, by using computer-based disease assessment training programmes, and by developing and using diagrammatic keys (standard area diagrams).

Introduction

"How can plant pathologists apply advanced statistical procedures or develop quantitative predictive models based upon disease assessment data of unknown accuracy and precision?" David R. Mackenzie, 1979

The efficient application of any integrated disease management programme requires accurate and precise information concerning the quantitative measurement of the disease and/or pathogen population, yet the accuracy and precision of quantitative disease/pathogen assessments in plant pathology is often taken for granted (Main, 1977; Zadoks and Schein, 1979; Gaunt, 1987; Kranz, 1988; Nutter et al., 1991; Nutter and Schultz, 1995; Nutter and Gaunt, 1996). Accurate and precise disease (or pathogen) assessments provide a quantitative link between disease management theory and practice (Shokes et al., 1987; Nutter et al., 1991; Nutter and Schultz, 1995). An integral component of studying the interactions of host and pathogen populations in time and space is the ability to accurately discriminate between levels of injury (disease intensity) caused by plant pathogens.

Disease intensity is a generic term for the amount of disease in a host population. Disease intensity can be either the independent variable or the dependent variable in stimulus–response models; however, in both cases, disease intensity needs to be quantified with a high degree of accuracy and precision if meaningful predictive models are to be developed (Nutter, 1990; O'Brien and van Bruggen, 1992; Nutter and Gaunt, 1996; Guan and Nutter, 2003). For example, quantifying disease intensity–crop yield (loss) relationships demands a high degree of accuracy and precision

with regards to disease assessments because disease intensity is used as the independent variable (X) in single-point or area under the disease progress curve (AUDPC) – crop yield (loss) regression models (Guan and Nutter, 2001, 2004). The stimulus (disease intensity) must be measured accurately in order to develop yield response or yield loss models that have adequate predictive capabilities.

The symbol Y is often used to represent a measure of disease (or pathogen) intensity because disease intensity assessments are often graphed on the y-axis with respect to time (X-axis). The graphical representation of disease intensity versus time is referred to as a disease progress curve, whereas the graphical representation of pathogen intensity versus time is referred to as a pathogen progress curve (Nutter, 2001). A disease (or pathogen) progress curve is the signature of an epidemic and represents the integration of all host, pathogen, and environmental effects (including pathogen vectors and disease management tactics) that occur during the period of host–pathogen interaction (Campbell and Madden, 1990; Nutter, 1997b).

Quantification of disease intensity also requires a high degree of accuracy and precision when disease intensity is the dependent variable (Y) to quantify the rate of disease progress with respect to time (X), or the change in disease intensity with respect to distance (X). In both cases, X (time or distance) can be measured with great accuracy, and therefore, the accuracy and precision of disease intensity assessments (Y) directly affect how much of the variation in Y can be explained by X in such models.

Disease assessment defined

The branch of plant pathology that deals with the theory and practice of quantitative disease (and/or pathogen) assessment is known as phytopathometry (Main, 1977; Zadoks and Schein, 1979; Campbell and Madden, 1990). Disease assessment is defined as the act (or process) of quantitatively measuring disease intensity (Campbell and Madden, 1990; Nutter et al., 1991; Nutter and Gaunt, 1996). In plant pathology, there are two basic and distinct populations that can be quantitatively assessed: the pathogen population and the disease population (Nutter, 1997b, 1999). Because plant disease epidemics result from the interaction of host and pathogen populations in time and space, as affected by the environment, quantification of the disease population usually involves an assessment of visible injury (disease symptoms). This is true because disease injury is often directly proportional to the size of the pathogen population (Nutter et al., 1991; Nutter and Guan, 2001). On the other hand, pathogen assessments can be obtained by directly measuring the pathogen population (e.g. the number of spores, sclerotia, nematodes, etc.) per unit area or volume, or the use of a detection method to determine the presence or absence of a pathogen for each sampling unit (e.g., ELISA or PCR to detect the presence of a pathogen in or on a sampling unit). Thus, researchers can perform disease assessments or pathogen assessments (or both); however, these terms should not be used interchangeably because they represent different populations being assessed (Nutter, 1997b, 1999, 2001).

Disease versus pathogen intensity

Disease intensity (Y) is a general (generic) term used for quantifying the amount of disease in a population (Campbell and Madden, 1990; Nutter et al., 1991; Nutter and Gaunt, 1996). In plant pathology, the three most common measures of disease (and pathogen) intensity are: (i) prevalence, (ii) incidence, and (iii) disease severity. Disease prevalence is a term that is often used interchangeably (and mistakenly) with disease incidence. Prevalence is defined as the number of geographical sampling units (fields, farms, counties, states, regions, etc.) where a disease or pathogen has been detected, divided by the total number of geographical sampling units assessed (Zadoks and Schein, 1979; Campbell and Madden, 1990; Nutter et al., 1991; Nutter and Gaunt, 1996). It is important to distinguish disease prevalence from pathogen prevalence. Disease prevalence measures the proportion (or percentage) of geographical sampling units (fields, counties, countries, etc.) where a disease (expressing symptoms) has been found to occur, divided by the total number of geographical sampling units inspected or surveyed, whereas pathogen prevalence is a measure of the number of geographical sampling units where the pathogen has been detected (e.g., by direct plating, inspections for the presence of

pathogen signs, ELISA, PCR, etc.), divided by the total number of geographical sampling units that were inspected, tested, or indexed (Nutter, 2001). Prevalence data are often multiplied by 100 to express as a percentage. A single diseased or infected plant (or plant part) is all that is required to change the status of a geographic sampling unit from negative (−) to positive (+), provided the sensitivity of the method is sufficient to detect the presence of a pathogen in a bulked (diluted) sample. Bulking samples is particularly useful when pathogen incidence is low because the number of bulked samples tested or indexed is often less than the number of individuals sampled and processed, thus reducing the cost of detection per sampling unit (Nutter and Gaunt, 1996; Nutter, 1997b).

Disease incidence is defined as the number of sampling units (e.g., leaflets, leaves, stems, tillers, whole plants, seeds, etc.) that are diseased (expressing symptoms), divided by the total number of sampling units sampled and assessed (Nutter et al., 1991; Nutter, 1997b, 1999). As with prevalence, it is important to make a clear distinction as to whether incidence is based on detection of the pathogen or on the basis of disease (visual symptoms) in a host population (Nutter, 2001). Progress curves based on pathogen detection (indexing) methods, such as ELISA, may closely mirror progress curves based on disease symptoms (Padgett et al., 1990; Nutter 2001); however, in many instances, the use of different disease assessment or pathogen detection methods may result in progress curves with quite different shapes and rates (Nutter, 1997b, 2001).

Disease severity is a measure of the amount of disease per sampling unit and it is this particular type of measurement that this article will focus upon (Nutter et al., 1991; Nutter, 1997b). Researchers should clearly define disease severity by providing not only a descriptive definition, but also an operational definition that includes the dimensions that were used to assess disease. In plant pathology, disease severity is most often operationally defined as the diseased leaf area (l^2), divided by the total leaf area of a leaf or sampling unit (L^2)$\times$100, i.e., ($l^2/L^2\times100$) to obtain percentage disease severity (James, 1971; Nutter et al., 1991). Other common measures of disease severity include the number of lesions/leaf (or sampling unit), the number of lesions/cm^2 of leaf (or other sampling unit), or the area of non-green tissue of a sampling unit divided by the total area of the sampling unit $\times$ 100 (Nutter, 2001). Disease severity could be also defined as the volume of a disease-induced gall (using the equation and dimensions for a cylinder or a sphere, etc.), as is done in human epidemiology for cancerous tumors (Nutter, 1999).

The concept of pathogen severity is becoming more widespread as new methods are developed to quantify the amount of a pathogen present in a sampling unit. Examples include the use of quantitative PCR methods to estimate the amount of virus (or pathogen) present per gram of leaf tissue, or the number of nematodes per gram of root tissue. In spite of advances concerning the concept of pathogen severity, the majority of severity assessments employed today usually involve visual estimates of disease severity ($l^2/L^2\times100$), and yet, a number of critical questions still remain. At the top of the list of such questions is: how can we do better to improve the accuracy and precision of visual disease assessments? This article will address specific ways to improve the accuracy and precision of diseases assessments. These are: (i) use of regression to quantify the precision of disease raters and/or assessment methods, (ii) the use of computer-based assessment training programmes, and (iii) the use of standard area diagrams (diagrammatic keys) in colour.

Use of linear regression to assess and compare the precision of assessment methods

The expenditure of time and money to develop, evaluate, and compare disease assessment methods can prevent serious flaws (e.g., rater bias) in data acquisition. Disease assessment methods should provide accurate and precise information that satisfies the goals and needs of the research (Nutter and Gaunt, 1996). Campbell and Madden (1990) have defined precision as the lack of variation in disease estimates when the same sampling units are evaluated by other raters. However, this definition of precision excludes another potential source of error, i.e., the repeatability of individual raters (Nutter et al., 1993). Shokes et al. (1987) proposed using a test-retest procedure using correlation analysis to quantify rates repeatability; however, this method provides a measure of precision

(agreement) among raters and does not quantify the degree of bias among raters.

Simple linear regression provides a powerful method to quantify the degree of error (bias) due to raters or assessment methods (Nutter et al., 1993). Regression analysis has been used to determine the relative precision of a visual assessment method (disease severity) and a remote sensing assessment method (reflectance at 600 nm) in which a hand-held, multispectral radiometer was employed (Nutter et al., 1993). The disease assessed was dollar spot of bentgrass (caused by *Sclerotinia homoeocarpa*). The precision of different disease assessment methods and raters can be evaluated and compared by operationally defining intra-rater repeatability and inter-rater reliability (Nutter and Schultz, 1995). Intra-rater repeatability for different assessment methods can be determined by regressing one set of measurements (Y) (obtained by each rater) with a repeated set of measurements (X) performed on the same set of sampling units. The parameters and statistics used to compare the intra-rater repeatability of different assessment methods and/or raters are: slope, intercept, coefficient of determination (R^2), coefficient of variation (CV), and the standard error of the estimate for Y (SEEy) (Nutter et al., 1993; Nutter and Schultz, 1995). A slope significantly less than or greater than 1.0 would indicate the presence of systematic bias and the greater the deviation from 1.0, the greater the systematic bias for a specific rater and/or method. This is because for each 1% increase in estimated disease severity the first time a set of sampling units is assessed, there should be a corresponding 1% increase in estimated disease severity when the same set of sampling units are assessed a second time by the same rater or method (Nutter et al., 1993). An intercept significantly different from zero indicates the presence of another form of bias that is constant for all disease levels evaluated. The use of R^2, CV, and SEEy values to quantify and compare the precision of disease assessment methods or raters has been previously described (Nutter et al., 1993; Nutter and Schultz, 1995; Nutter, 2001). Likewise, linear regression can be used to quantify precision among raters or methods (inter-rater reliability) by having two or more raters (and/or methods) assess the same set of sampling units, and then evaluating the slopes, intercepts, R^2, CV, and SEEy values (Nutter et al., 1993).

Disease assessment training with computer programmes

The accuracy and precision of disease severity assessments have come into question due to the measurable bias that different raters have shown when evaluating the same set of diseased sampling units (Sherwood et al., 1983; Forbes and Jeger, 1987; Kranz, 1988; Nutter et al., 1993). Accuracy can be defined as the measure of the closeness of a disease assessment to the true value (Nutter et al., 1991; Zadoks and Schein, 1979). When assessing disease severity, the stimulus (X) is the actual disease severity of a sampling unit and the rater's estimate of disease severity (Y) is the response. For each 1% increase in actual severity, we would expect a rater to also to estimate a 1% increase, i.e., the slope should be equal to 1.0 (no systematic bias) and the intercept should not be significantly different from zero (no constant bias present). Accuracy cannot be properly evaluated unless the researcher is confident that the actual (true) disease severity can be measured absolutely. This is easily achieved using computer-generated images of diseased leaves because the computer can be programmed to calculate the number of non-green (diseased) pixels in an image, divided by the total number of pixels in the image $\times$ 100 to obtain a true measure of percentage disease severity (Nutter and Litwiller, 1998; Nutter et al., 2000). The use of computer programmes to enhance learning has become widely accepted for several reasons (Nutter, 1997a). One advantage of computer-aided disease assessment training is that a full range of disease severity levels can be presented as stimuli to which operators of the programme respond. Nutter and Worawitlikit (1990) built upon the computer-based disease assessment training concept by developing a computer programme to assess diseases of peanut called Disease.Pro. Recognizing the tremendous potential to improve the accuracy and precision of disease assessments through computer-based training programmes, Nutter and Litwiller (1998) later developed a more generic disease assessment training programme (Severity.Pro) that allowed the user to select from a menu of leaf shapes (alfalfa, apple, barley, cucumber, grape, tomato, etc.) and lesion types (anthracnose, blotch, downy mildew, target spot, powdery mildew, etc.) to mimic almost any foliar pathosystem. Severity.Pro was recently rewritten

in Java to be more compatible with present-day operating systems.

The most current version of Severity.Pro allows raters to: (i) choose whether or not they want the actual severity to be immediately displayed (feedback), (ii) choose the number of leaves to be assessed and the size of the lesions that will appear on diseased leaf images (small, medium, large, or random), (iii) view graphs of the absolute error (Y) versus the actual severity (X) (absolute error is defined as the estimated severity minus the corresponding actual severity), and (iv) regress the rater's severity estimates (Y), against the actual disease severities (X). These changes allowed for a more powerful training tool because: (i) raters can take a pre-test (without feedback before training) to provide a baseline of how different disease severity levels are perceived, (ii) the data can be viewed in graphical form and analyzed by regression, (iii) raters can execute a drill and practice session and receive feedback as to the actual level of disease severity immediately after the estimated severity is keyed in, (iv) rater improvement and, more importantly, the degree of rater improvement can be documented by having raters take a post-test (without feedback after training) and then compare pre- and post-test regression parameters and statistics, and (v) the results of pre- and post-tests can also be used to evaluate and compare rater performance.

Computer-based disease assessment training programmes provide a useful platform for teaching disease assessment theory and hands-on practice. For example, the results of computer training for six raters using Severity.Pro are shown in Table 1. The six raters evaluated computer-generated images of grapevine leaves infected by downy mildew by assessing 30 images before and after training. In pre-tests, one rater (Rater 2) generally overestimated downy mildew severity throughout the range of the severities tested, with rater error being as high as 21% (solid circles, Figure 1a). Following training, Rater 2's estimates were within 5–10% of the actual severity levels (open circles, Figure 1a). Figure 1b shows that this rater also had a constant bias of 6.8% prior to training (Y-intercept) and that this bias was reduced to near zero (−0.14%) after computer-based training using Severity.Pro. Based on R^2 values, the precision of Rater 2 was also significantly improved following training (R^2 was 95% following training compared with 85% prior to training). As a group, five of the six raters demonstrated improvement in precision following computer training, as measured by improvement in the coefficients of determination (R^2) (Table 1). Three of the raters had R^2 values that were 6–11% higher after training. The other three raters (rater 1, rater 4, and rater 6) were already highly precise and their R^2 values increased 2, 1, and 0%, respectively.

Table 1. Y-intercepts[a], slopes[b], coefficients of determination (R^2)[c], and standard errors of the Y-estimate (SEE_Y)[d] for six raters before training (pretest) and after training (posttest) using a computer programme that simulates downy mildew of grapevines (adapted from Nutter, 2001)

Rater	Pretest				Posttest			
	Intercept	Slope	R^2	SEE_Y	Intercept	Slope	R^2	SEE_Y
1	−7.19	1.11	0.94	2.74	−1.52	1.01	0.96	1.95
2[e]	6.83	1.02	0.85	3.80	−0.14	0.94	0.95	1.97
3	−6.30	0.91	0.91	2.83	7.21	0.83	0.97	1.41
4	1.89	1.06	0.91	2.97	0.61	0.82	0.92	2.25
5	−1.13	1.27	0.84	4.69	−8.46	1.05	0.94	2.61
6	1.77	1.01	0.97	1.62	−3.10	1.03	0.97	1.76
Improved					3/6	2/6	5/6	5/6

[a] Y-intercepts that deviate from zero indicate the presence of a constant source of rater bias with regards to accuracy.
[b] Slopes that deviate from 1.0 indicate the presence of a systematic source of rater bias with regards to accuracy.
[c] The higher the coefficient of determination (R^2), the higher the precision of rater estimates.
[d] The lower the standard error of the Y-estimate, the higher the precision of rater estimates.
[e] Data for Rater 2 are shown in graphical form in Figure 1 as this data would appear in the disease assessment computerized training programme Severity.Pro (Nutter and Litwiller, 1998).

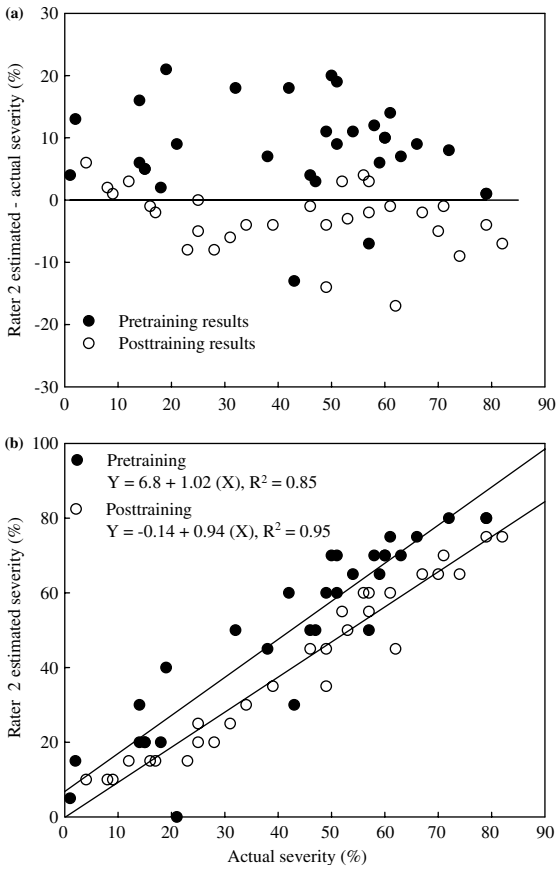


Figure 1. Improvement in the (a) absolute error (estimated minus actual disease severity) and (b) accuracy (slope, intercept) and precision (R^2, SEEy) of Rater 2 (from Table 1), before and after disease assessment using the computer program Severity.Pro (Nutter and Litwiller, 1998).

As stated earlier, the standard error of the *Y*-estimate is another important measure of rater precision. This statistic provides information concerning the degree of error associated with a predicted value of *Y*. Therefore, the lower the SEE_Y, the higher the precision (Nutter and Schultz, 1995). For five of the six raters, SEE_Y values decreased following disease assessment training by an average of 40% (3.41 to 2.04). Rater 6 showed no significant change in pre-test versus post-test SEE_Y values because this rater was already very precise (i.e., rater 6 had very low SEE_Y values in both pre- and post-training tests). Thus, computerized disease assessment training programmes provide an important educational tool that can be used to teach disease assessment theory and concepts, as well as to substantially improve both the accuracy and the precision of disease assessment data.

Use of standard area diagrams to improve the accuracy and precision of disease severity assessments

Disease assessment keys, also known as diagrammatic keys or standard area diagrams are pictorial diagrams that depict the true amount of injury (usually disease severity) on individual sampling units (quadrats, whole plants, leaves, fruit, tubers, etc.). Disease severity of each individual diagram is expressed as a percentage of the total surface area of each sampling unit (disease area/total area of the image $\times$ 100) (Nutter and Esker, 2001). Standard area diagrams (SADs) provide raters with a series of reference images that are accepted to be the truth in terms of the actual amount of injury (severity) depicted on each disease diagram. Clive James developed and marketed the first series of black and white standard area diagrams (James, 1971). More recently, Nutter and Litwiller (1998) developed and tested a computer programme (Severity.Pro) that generates standard area diagrams in colour. Thus, Severity.Pro provides a powerful tool to generate, capture and print diseased leaf images with known severity levels in colour (Nutter et al., 1998). This enables researchers to create a series of pictorial colour diagrams that can be used as an assessment aid to improve the accuracy and precision of disease assessment data. Although it has long been assumed that the use of standard area-diagrammatic keys will help to improve the accuracy and precision of visual disease severity assessments performed by raters (James, 1971; Horsfall and Cowling, 1978; Kranz, 1988), only recently have definitive studies been conducted to demonstrate that the accuracy and precision of disease assessments are actually improved when standard area-diagrammatic keys are used (Godoy et al., 1997; Nutter et al., 1998; Leite and Amorim, 2002; Gomes et al., 2004). As part of a class exercise for students enrolled in a course in plant disease epidemiology at Iowa State University, 10 raters were asked to assess 30 diseased leaf images (representing a range of disease severities) of downy mildew of grape, both with and without the use of colour-standard area diagrams (Nutter and Litwiller, 1998). When individual rater estimates

were regressed against the true disease severity levels as calculated by the computer programme, it was found that rater estimates of disease severity were much closer to the actual (true) severity levels when raters used standard area diagrams as an assessment aid to assess disease severity (Nutter and Esker, 2001).

As an example, Figure 2 shows a typical situation regarding accuracy and precision of visual assessments performed by one rater with, and without, the use of standard area diagrams for grapevine downy mildew. When using the standard area diagrams, this rater had greater accuracy (less systematic and constant bias) as indicated by a slope closer to 1.0 (1.05) and a *Y*-intercept closer to zero (−0.28%) compared to the slope (0.86) and intercept (5.34%) when not using the standard area diagrams. Moreover, R^2 values were higher and SEE_y values were lower when standard area diagrams were used, indicating there was a significant increase in the precision of the assessment data when using the standard area diagrams.

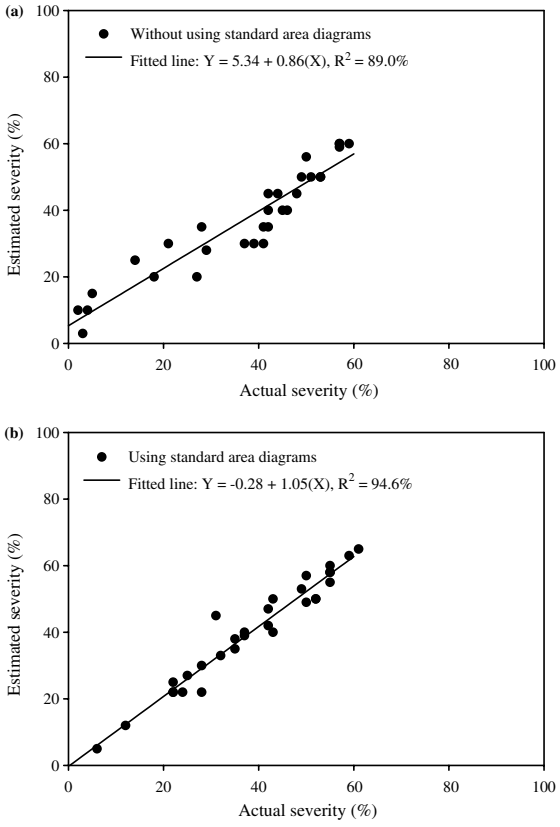


Figure 2. (a) Estimated severity of grapevine downy mildew compared with actual (true) severity when assessing computer images without the use of standard area diagrams and (b) Estimated versus actual severity when using standard area diagrams. Improvements were apparent as both systematic and constant bias were reduced.

As a class, statistical analyses for accuracy showed that eight of the ten raters achieved intercepts closer to zero (less constant bias) and that seven of the ten raters achieved slopes closer to 1.0 (less systematic bias) when standard area diagrams were used (Nutter and Schultz, 1995). Statistical analyses for precision showed that seven of the ten raters achieved higher coefficients of determination (R^2), and eight of the ten raters had lower standard errors of the estimate for *Y* (SEE_y) when standard area-diagrammatic keys were used as an assessment aid. Thus diagrammatic standard area-assessment keys can substantially improve both the accuracy and the precision of visual disease assessments.

Summary and conclusions

The potential for rater bias (under- or over-estimation of the actual level of disease severity) is an ever-present concern that should receive serious consideration by researchers when raters are making visual disease assessments and will use that information as the basis to develop stimulus–response models, or to evaluate and compare disease management tactics, strategies, or integrated disease management systems (Zadoks and Schein, 1979; Gaunt, 1995; Nutter, 1997b, 1999, 2001). Rater bias, however, can be effectively reduced. Disease assessment training programmes using computer-generated images of disease leaves have been shown to improve both accuracy and precision (Nutter and Schultz, 1995; Nutter and Parker, 1997; Nutter and Litwiller, 1998). Moreover, studies by Godoy et al. (1997), Gomes et al. (2004), Nutter (2001), and Nutter and Esker (2001) have documented that the use of standard area diagrams as an assessment aid for visually assessing disease severity can also significantly improve the accuracy and precision of disease severity assessment data. The use of both computer-based disease assessment training programmes and standard area diagrams to improve the accuracy and precision of disease assessment data are not mutually exclusive, as both methods should be used to obtain the best disease assessment data possible. Finally, the use of

regression to evaluate and compare the accuracy and precision of: (i) colour disease assessment protocols (e.g., use of a linear scale versus a logarithmic disease assessment scale) (Nutter and Esker, 2005), (ii) disease assessment instruments (e.g., image analysis or remote sensing sensors and instruments) (Nutter, 1990; Guan and Nutter, 2004), and/or (iii) disease raters, can provide researchers with statistical methods to determine the accuracy and precision of disease assessment data (Nutter et al., 1993; Nutter and Littrell, 1996; Nutter, 1997a; Guan and Nutter, 2003). Thus, researchers can and should place greater focus upon evaluating, comparing, and selecting the best disease assessment protocols, instruments, and/or raters that best meet the goals of the research.

References

Campbell CL and Madden LV (1990) Introduction to Plant Disease Epidemiology, John Wiley, New York.

Forbes GA and Jeger MJ (1987) Factors affecting the estimation of disease intensity in simulated plant structures. Zeitschrift fur Pflanzen Krankheiten und Pflanzenschutz 94: 113–120.

Gaunt RE (1995) New philosophies and technologies in disease measurement and appraisal. Canadian Journal of Plant Pathology 17: 185–189.

Godoy CV, Carneiro MTPG, Iamauti MT, Dalla Pria M, Amorim L, Berger RD and Bergamim Filho A (1997) Diagrammatic scales for bean diseases: development and validation. Journal of Plant Diseases and Protection 104: 336–345.

Gomes AMA, Michereff SJ and Mariano RLR (2004) Development and validation of a diagrammatic key for Cercospora leaf spot of lettuce. Summa Phytopathologica 30: 38–42.

Guan J and Nutter FW, Jr. (2001) Factors affecting the quality and quantity of sunlight reflected from alfalfa canopies. Plant Disease 85: 865–874.

Guan J and Nutter FW, Jr. (2003) Quantifying the intra-rater repeatability and inter-rater reliability of visual disease and remote sensing assessment methods in the alfalfa foliar disease pathosystem. Canadian Journal of Plant Pathology 25: 143–149.

Guan J and Nutter FW, Jr. (2004) Comparison of single-point alfalfa yield models based on visual disease intensity and remote sensing assessments. Canadian Journal of Plant Pathology 26: 314–324.

Gaunt RE (1987) Measurement of disease and pathogens. In: Teng PS (ed.) Crop Loss Assessment and Pest Management (pp. 6–18) The American Phytopathological Society, St. Paul, MN.

Horsfall JG and Cowling EB (1978) Pathometry: the measurement of plant disease. In: Horsfall JG and Cowling EB (eds) Plant Disease: An Advanced Treatise. Vol. 2 (pp. 120–136) Academic Press, New York.

James C (1971) A Manual of Assessment Keys for Plant Disease. Canadian Department of Agriculture Publishing.

Kranz J (1988) Measuring plant disease. In: Kranz J and Rotem J (eds) Experimental Techniques in Plant Disease Epidemiology (pp. 35–50) Springer-Verlag, New York.

Leite RMVBC and Amorim L (2002) Development and validation of a diagrammatic scale for Alternaria leaf spot of sunflower. Summa Phytopathologica 28: 14–19.

Main CE (1977) Crop destruction – The raison d' être of plant pathology. In: Horsfall JG and Cowling EB (eds) Plant Disease: An Advanced Treatise (pp. 55–78) Academic Press, New York.

Nutter FW, Jr. (1990) Remote sensing and image analysis for crop loss assessment. In: Crop Loss Assessment in Rice (pp. 93–105) International Rice Research Institute, Manila, Philippines.

Nutter FW, Jr. (1997a) Disease severity assessment training. In: Francl L and Neher D (eds) Exercises in Plant Disease Epidemiology (pp. 1–7) APS Press, St. Paul, MN.

Nutter FW, Jr. (1997b) Quantifying the temporal dynamics of plant virus epidemics: a review. Crop Protection 16: 603–618.

Nutter FW, Jr. (1999) Understanding the interrelationship between botanical, human, and veterinary epidemiology: The Ys and Rs of it all. Ecosystem Health 5: 131–140.

Nutter FW, Jr. (2001) Disease assessment terms and concepts. In: Maloy OC and Murray TD (eds) Encyclopedia of Plant Pathology (pp. 312–323) John Wiley and Sons, Inc, NY.

Nutter FW, Jr. and Esker PD (2006) Disease assessment keys. In: Maloy OC and Murray TD (eds) Encyclopedia of Plant Pathology (pp. 323–326) John Wiley and Sons, Inc, NY.

Nutter FW, Jr. and Esker PD (2005) The role of psychophysics in phytopathology: the Weber-Fechner law revisited. European Journal of Plant Pathology 114: 199–213.

Nutter FW, Jr. and Gaunt RE (1996) Recent developments in methods for assessing disease losses in forage/pasture crop. In: Chakraborty S, Leath DT, Skipp RA, Pederson GA, Bray RA, Latch GCM and Nutter FW Jr. (eds) Pasture and Forage Crop Pathology (pp. 93–118) American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, WI.

Nutter FW, Jr., Gleason ML, Jenco JH and Christians NC (1993) Assessing the accuracy, intra-rater repeatability and inter-rater reliability of disease assessment systems. Phytopathology 83: 806–812.

Nutter FW, Jr. and Guan J (2001) Disease losses. In: Maloy OC and Murray TD (eds) Encyclopedia of Plant Pathology (pp. 340–351) John Wiley and Sons, Inc, NY.

Nutter FW, Jr. and Littrell RH (1996) Relationships among percent defoliation, canopy reflectance, and pod yield in the peanut-late leafspot system. Crop Protection 15: 135–142.

Nutter FW, Jr. and Litwiller D (1998) A computer program to generate standard area diagrams to aid raters in assessing disease severity. Phytopathology 88: S117.

Nutter FW, Jr., Litwiller D and Wegulo SN (1998). Do standard area diagrams improve the accuracy and precision of disease assessments? Proceedings of the 7th International Congress of Plant Pathology. p. 2.1.16.

Nutter FW, Jr. and Parker SK (1997) Fitting disease progress curves using EPIMODEL. In: Francl L and Neher D (eds) Exercises in Plant Disease Epidemiology (pp. 24–28) APS Press, St. Paul, MN.

Nutter FW, Jr. and Schultz PM (1995) Improving the accuracy and precision of disease assessment: selection of methods and use of computer-aided training programs. Canadian Journal of Plant Pathology 17: 174–184.

Nutter FW, Jr., Teng PS and Shokes RM (1991) Disease assessment terms and concepts. Plant Disease 75: 1187–1188.

Nutter FW, Jr., Wegulo SN, Litwiller D, Emmitt RW and Magarey PA (2000) A computer program to improve the accuracy and precision of severity assessments in grapevines. In: Magarey, PA, Thiele, SA, Tschirpig, KL, Emmet, RW, Clark, K and Magarey RD (eds) Proceedings of the Third International Workshop on Grapevine Downy and Powdery Mildew (pp. 64–69). South Australian Research and Development Institute Series Report No. 50.

Nutter FW, Jr. and Worawitlikit O (1990) Disease. Pro: A computerized disease assessment training program. Proceedings of the American Peanut Research and Ed. Society 22: 38.

O'Brien RD and van Bruggen AHC (1992) Accuracy, precision, and correlation to yield loss of disease severity scales for corky root of lettuce. Phytopathology 82: 91–96.

Padgett GB, Nutter FW, Jr. and Kuhn CW (1990) Quantification of disease resistance that reduces the rate of tobacco etch virus epidemics in bell pepper. Phytopathology 80: 451–455.

Sherwood RT, Berg CC, Hoover MR and Zeiders KE (1983) Illusions in visual assessment of Stagonospora leaf spot of orchardgrass. Phytopathology 73: 173–177.

Shokes FM, Berger RD, Smith DH and Rasp JM (1987) Reliability of disease assessment procedures: A case study with late leafspot of peanut. Oleagineux 42: 245–251.

Zadoks JC and Schein RD (1979) Epidemiology and Plant Disease Management, Oxford University Press, New York.

European Journal of Plant Pathology (2006) 115:105–122
DOI 10.1007/s10658-005-1250-8

Relation between soil health, wave-like fluctuations in microbial populations, and soil-borne plant disease management

Ariena H.C. van Bruggen[1], Alexander M. Semenov[2], Anne D. van Diepeningen[1], Oscar J. de Vos[1] and Wim J. Blok[1]
[1]*Biological Farming Systems, Wageningen University, Marijkeweg 22, 6709 PG Wageningen the Netherlands; (Fax: +31-317-468213; E-mail: ariena.vanbruggen@wur.nl);* [2]*Department of Microbiology, Biological Faculty, Moscow State University, 119899 Moscow, Russia;*

Key words: biological control, disease management, harmonic fluctuations, resilience, soil health, soil-borne pathogens

Abstract

A healthy soil is often defined as a stable soil system with high levels of biological diversity and activity, internal nutrient cycling, and resilience to disturbance. This implies that microbial fluctuations after a disturbance would dampen more quickly in a healthy than in a chronically damaged and biologically impoverished soil. Soil could be disturbed by various processes, for example addition of a nutrient source, tillage, or drying-rewetting. As a result of any disturbance, the numbers of heterotrophic bacteria and of individual species start to oscillate, both in time and space. The oscillations appear as moving waves along the path of a moving nutrient source such as a root tip. The phase and period for different trophic groups and species of bacteria may be shifted indicating that succession occurs. DGGE, Biolog and FAME analysis of subsequent populations in oscillation have confirmed that there is a cyclic succession in microbial communities. Microbial diversity oscillates in opposite direction from oscillations in microbial populations. In a healthy soil, the amplitudes of these oscillations will be small, but the background levels of microbial diversity and activity are high, so that soil-borne diseases will face more competitors and antagonists. However, soil-borne pathogens and antagonists alike will fluctuate in time and space as a result of growing plant roots and other disturbances, and the periods and phases of the oscillations may vary. As a consequence, biological control by members of a single trophic group or species may never be complete, as pathogens will encounter varying populations of the biocontrol agent on the root surface. A mixture of different trophic groups may provide more complete biological control because peaks of different trophic groups occur at subsequent locations along a root. Alternatively, regular addition of soil organic matter may increase background levels of microbial activity, increase nutrient cycling, lower the concentrations of easily available nutrient sources, increase microbial diversity, and enhance natural disease suppression.

Abbreviations: BCA – biocontrol agent; CFUs – colony forming units; DGGE – denaturing gradient gel electrophoresis; FAME – fatty acid methyl esters; GFP – green fluorescent protein

Introduction

Health is a necessary condition for the survival of individual living organisms, communities, ecosystems, and for nature in general. Ecologists have long recognized that the state of health of terrestrial, edaphic and aquatic ecosystems is important. However, it is not so easy to define, let alone measure, ecosystem health. This is also true for soil, which is considered as a living system, where

many physical and chemical properties are mediated by biota, which are primarily responsible for its health (Brussaard et al., 2004). The living components of a soil possess both stable and dynamic characteristics. Recently, we proposed to use the resistance and resilience of microbial communities in response to a disturbance as quantitative indicators for soil health (van Bruggen and Semenov, 1999, 2000).

Soil as a productive system was developed and continues to develop jointly with plants that inhabit the soil. The condition of the soil, including its microbial community, reflects the condition of the (past) vegetation. In agricultural ecosystems, especially in conventional farming systems, most natural plants are considered weeds. Decreasing the vegetation diversity leads to pauperization of soil inhabitants, decreasing of interconnectedness and functional interchangeability. An extremely simplified vegetation, such as a monoculture, selects a specific microbial community, including plant pathogenic microorganisms and sometimes also their parasites or antagonists. However, such a simplified ecosystem may be very sensitive to the slightest disturbance and cannot be considered healthy.

High quality soil is the main production resource for many societies. However, this resource is disappearing at an alarming rate due to loss of organic matter as a result of erosion, oxidation, compaction, and biological impoverishment. In particular, agricultural systems with minimal biological diversity and large inputs of synthetic fertilizers and pesticides have problems with poor soil health and associated plant diseases. Action must be taken urgently to restore the balance of the soil ecosystem and its health status.

In this review we present a dynamic view of microbial populations, soil health, and disease suppression. In the following paragraph an introduction is given on soil health and its connection to disease suppression. The next two paragraphs deal with temporal and spatial oscillatory responses of bacterial communities to various disturbances. Then, we demonstrate that soil-borne pathogens respond with similar oscillations to a disturbance from a growing root. Next, agricultural management practices to control soil-borne diseases, like the use of organic amendments, tillage and biocontrol agents, will be discussed from the point of view of dynamic microbial oscillations. Finally, conclusions will be presented regarding soil health, microbial oscillations and soil-borne diseases.

Soil health and disease suppression

Rapport (1995) defined a healthy ecosystem as an ecosystem with the following characteristics: (1) integrity of nutrient cycles and energy flows, (2) biological diversity, (3) interconnectedness between functional units, (4) stability and resilience when faced with a disturbance or stress, and (5) limited plant and animal disease outbreaks. A soil ecosystem is considered healthy if it has a good balance of mineral and organic substances and living components. Such a balance is reached when an ecosystem comes to a climax condition, characterized by high biodiversity and low concentrations of easily available organic and inorganic nutrients (van Bruggen and Semenov, 1999, 2000). To maintain soil health, it is necessary to promote high primary productivity, high microbial biomass, activity and diversity, high nutrient turnover rates, and low residual nutrient pools; in other words, oligotrophic conditions. In particular, losses of mineral nitrogen and dissolved organic carbon from soil and soil biological complexity have been used to assess the functioning of soil ecosystems (Liiri et al., 2002).

Soils of natural ecosystems are generally thought of as being healthier than those of agroecosystems. Indeed, cultivated soils generally have lower microbial diversities and more severe disease problems than they had as a natural habitat (Ko, 1982; Buckley and Schmidt, 2001). Organically managed soils, where synthetic fertilizers and pesticides are not used, are closer to natural soils than conventionally managed soils even though soil fertility is maintained by regular additions of organic materials (van Bruggen, 1995; van Bruggen and Termorshuizen, 2003). Especially chlorinated pesticides have had negative impacts on microbial diversity (Mas et al., 1996). Although some authors found no differences in soil microbial diversity between organically and conventionally managed soils (Lawlor et al., 2000; Franke-Snyder et al., 2001), most researchers reported a higher biological diversity for organically than for conventionally managed soils with respect to various taxa, namely bacteria (Sivapalan et al., 1993; Drinkwater et al., 1995; Mäder et al.,

2002; van Diepeningen et al., 2005), arbuscular mycorrhizal fungi (Ryan et al., 1994; Oehl et al., 2003), nematodes (Mulder et al., 2003; van Diepeningen et al., 2005), earthworms (Mäder et al., 2002), and arthropods (Drinkwater et al., 1995; Mäder et al., 2002). Also, a higher microbial activity (Workneh et al., 1993; Mäder et al., 2002) and microbial biomass (Workneh and van Bruggen, 1994; Mäder et al., 2002; Mulder et al., 2003) were found in organically managed soils.

High microbial biomass, activity, and diversity in natural or agricultural soils have been associated with suppression of soil-borne plant diseases (Nitta, 1991; Workneh and van Bruggen, 1994; Mäder et al., 2002). This kind of suppression may be due to general competition or antagonism, which may be non-specific and active against a wide range of soil-borne pathogens (Gerlagh, 1968; Whipps, 1997). However, in a few cases, no relationships were found between microbial biomass, activity or diversity and disease suppression. Boehm et al. (1993, 1997) found that the level of *Pythium* root rot suppression in peat mixes was not related to microbial biomass, activity or diversity but to the composition of the rhizosphere bacterial population. The seemingly unpredictability of disease suppression in relation to microbial community parameters may be due to a greater specificity of the relationship between pathogen and antagonist than sometimes thought, due to influences of varying soil physical and chemical characteristics (Hoper and Alabouvette, 1996), or due to variation in soil microbial communities in time and space (van Bruggen and Semenov, 2000).

Soil microbial populations generally fluctuate, and start to oscillate regularly in response to a disturbance, such as addition of organic material to soil (van Bruggen and Semenov, 2000). The amplitude of the waves in microbial populations (a measure of stability of the soil ecosystem), their frequency, and the time needed to return to initial conditions before organic amendment (a measure of the resilience of the system) may be used as indicators for soil health (van Bruggen and Semenov, 1999, 2000). The strongest wave-like response of microbial communities occurs in soils low in organic matter (Semenov et al., 1999). In high-organic matter soils with higher microbial biomass and activity, wave-like responses are also noticeable but the amplitudes and periods of these waves are less pronounced (Semenov et al., 1999). Stability and resilience of microbial communities after exposure to a disturbance could possibly also be related to disease suppression (van Bruggen and Semenov, 1999, 2000). Indeed, soils with a higher biological diversity and activity, such as natural or organically managed agricultural soils are frequently more suppressive to soil-borne diseases than conventionally managed agricultural soils (van Bruggen, 1995; van Bruggen and Termorshuizen, 2003).

Temporal wave-like fluctuations of microbial populations

Fluctuations in soil microbial populations have been observed many times, both in laboratory experiments and in the field with native bacterial communities (Aristovskaya, 1980; Zvyagintsev and Golimbet, 1983; Semenov, 2001). Under natural conditions, microbial fluctuations in soil appear irregular, and generally do not correlate with variations in external environmental characteristics, such as temperature and moisture content of the soil.

Irregular fluctuations can turn into regular oscillations with distinct waves after a disturbance such as addition of fresh organic matter to soil (Doebeli and Ruxton, 1997, 1998; Clarholm, 1981). Soil is generally low in easily available nutrients, especially fallow arable soil. Any disturbance providing a nutrient impulse under these conditions, such as incorporation of fresh organic matter or rewetting after drying, is likely to initiate a wave-like response of the microbial community (van Bruggen and Semenov, 1999; Caldéron et al., 2000). Hints of wave-like fluctuations were obtained in a field experiment after incorporation of cover crop debris into soil (van Bruggen and Semenov, 2000), but the observations were too sparse for time series or harmonical analysis (Shumway, 1988) to prove that regular oscillations occurred.

Only recently, we demonstrated the occurrence of regular oscillations over time using appropriate statistical techniques (Zelenev et al., 2004). Temporal oscillations of microbial populations (CFUs and microscopic cell counts) were observed for one month in soil amended with fresh plant material (grass–clover) incubated at constant temperature

and moisture. Bacterial populations fluctuated with different periods and amplitudes, depending on the specific conditions of each experiment, but immediately after the disturbance they revealed remarkable oscillations with large amplitudes. The patterns of the oscillations were quite predictable, always with a small and large peak within one week after incorporation of a grass–clover mixture in soil (Zelenev et al., 2004).

Various mechanisms underlying oscillations in microbial populations could be envisaged. Ecologists would first of all think of predator–prey interactions. Total bacterial-feeding nematodes did not oscillate, but increased monotonously in the second week after grass–clover incorporation into the soil (Zelenev et al., 2004). Daily changes in active numbers of bacterial-feeding nematodes did oscillate with a frequency similar to that of bacterial oscillations due to intermittent activation of the dormant juveniles (Dauerlarvae). However, the response of bacterial-feeding nematodes was still too slow to explain the decline after the first peak in bacterial populations within two days after the incorporation of grass–clover material (Zelenev et al., 2004). Similarly, protozoa were likely to be too slow to be responsible for the first decline in bacterial populations, suggesting that bacteria initiate the oscillations, and that their predators follow suit (Zelenev et al., 2004). During the experiments with grass–clover amended soil, various chemical and physical parameters were measured, such as ammonium and nitrate concentrations, pH, and redox potential. None of these parameters oscillated over time (Zelenev et al., 2004). In a simulation model, bacterial populations started to oscillate due to a temporary shortage of easily available substrate (Zelenev, 2004). Indeed, substrate availability is a plausible explanation for initiation of the oscillations. Yet, local oxygen deprivation after intensive bacterial growth has not been excluded but remains as a potential mechanism underlying the initiation of bacterial oscillations.

Another aspect of the mechanisms underlying bacterial oscillations is whether all taxa oscillate simultaneously, or if each peak represents a different microbial community corresponding with different organic components that are decomposed subsequently, or if there are repetitive successions within each peak. This question was addressed in another time-series experiment with and without grass–clover incorporated into a sandy soil. The response of copiotrophic bacterial CFUs (de Vos and van Bruggen, 2001; Zelenev et al., 2005) to the disturbance was determined daily over a period of nine days, both for the grass–clover (GC) and the non-amended control series (CO). Copiotrophic bacteria are fast-growing bacteria, with a relatively low substrate affinity and high half saturation constant. Copiotrophic CFUs oscillated over time in a wave-like fashion after amendment of the soil, whereas in the non-amended soil the CFUs fluctuated only very slightly (Figure 1). Microbial communities were characterized daily by determining DGGE profiles using eubacterial primers, FAME composition, and physiological profiles (Biolog, Hayward, CA, USA) on mixtures of copiotrophic colonies removed from agar plates (de Vos and van Bruggen, 2001). The patterns of DGGE bands (Figure 2), fatty acid composition and Biolog profiles indicated a succession in taxonomic and functional groups over time. Discriminant analysis of the DGGE band intensities, percentages of individual fatty acids, and intensities of physiological reactions on Biolog plates (Figure 3) showed that there were repetitive cycles in the succession of bacteria over time: communities at times when CFUs increased were more similar to each other than to those when CFUs decreased and *vice-versa* (de Vos and van Bruggen, 2001).

In an attempt to relate amplitudes and periods of the oscillations (representing stability and resilience of the soil ecosystem) to soil health, grass–clover mixtures were added to γ-irradiated and non-irradiated soils, a filtered (0.8 μm) soil suspension was added to the irradiated soil, and microbial populations were enumerated daily. In the γ-irradiated soil the amplitudes and periods of the wave-like fluctuations in microbial communities in response to the disturbance by grass–clover were larger than those in the non-irradiated soil, supporting the notion that non-irradiated soil is healthier (Zelenev et al., 2004). The amplitudes of microbial populations were also generally higher in conventionally than in organically managed soils (unpublished results) and higher in a fallow soil than in a cover-cropped soil after addition of the same amount of cover crop plant material (van Bruggen and Semenov, 2000). Thus, the amplitude and period of microbial oscillations after a disturbance may indeed be good indicators for soil health (van Bruggen and Semenov, 2000; Orwin and Wardle, 2004).

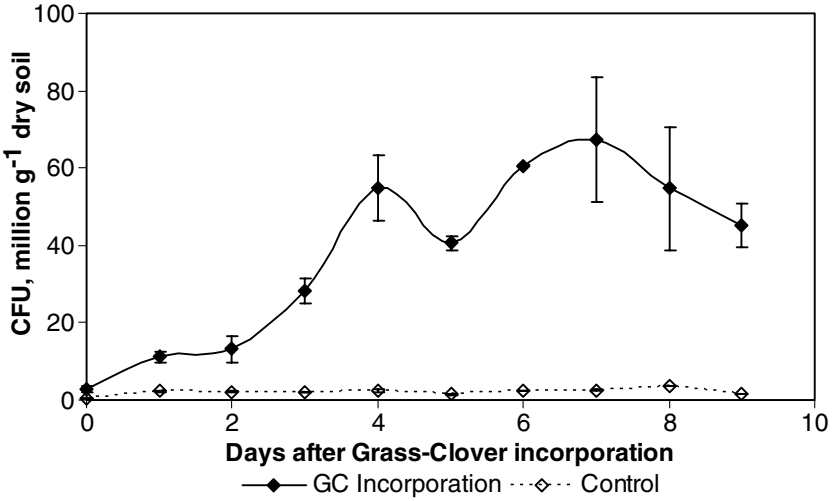


Figure 1. Wavelike fluctuations in numbers of copiotrophic bacteria isolated from sandy soil zero to nine days after incorporation of a grass–clover (GC) mixture and the relative stable numbers in the non-amended control soil.

Soil health is also frequently associated with limited disease outbreaks (Rapport, 1995; van Bruggen and Semenov, 2000), and indeed, root disease suppression is generally greater in non-irradiated than in γ-irradiated soil (Workneh and van Bruggen, 1994), in natural than in agricultural soil (Ko, 1982), and in organic than in conventional agricultural soil (van Bruggen, 1995; van Bruggen and Termorshuizen, 2003). This leads to the following questions: do plant pathogens also fluctuate in soil after a disturbance, and are the amplitudes greater in less healthy soils? Incorporation of vetch/oats cover crop debris in fallowed *versus* cover-cropped soil resulted in temporal fluctuations in copiotrophic bacterial CFUs over the next five weeks (van Bruggen and Semenov, 1999, 2000), and in similar fluctuations in damping-off incidence of tomato seedlings (caused by naturally occurring *Pythium ultimum* and *Pythium aphanidermatum*) in soil samples taken daily from the same experiment (Figure 4). The oscillations showed similar periods but were shifted in time: disease incidence increased when copiotrophic – and possibly antagonistic - populations decreased. It would be interesting to

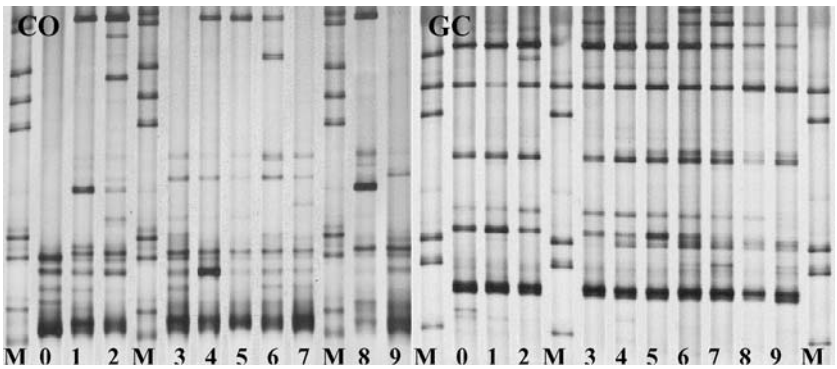


Figure 2. DGGE patterns of PCR products derived from DNA from copiotrophic bacterial colonies. The numbers represent the number of days after the incorporation of a grass/clover mixture in soil (GC) and in a non-amended control soil (CO). M represents a set of eubacterial marker strains. The urea/formamide denaturing gradient was between 40% and 48%. Note that the bacterial composition of CO does not change in time, while the composition of GC changes over time, the composition being similar after 0, 5 and 9 days and different on the other days.

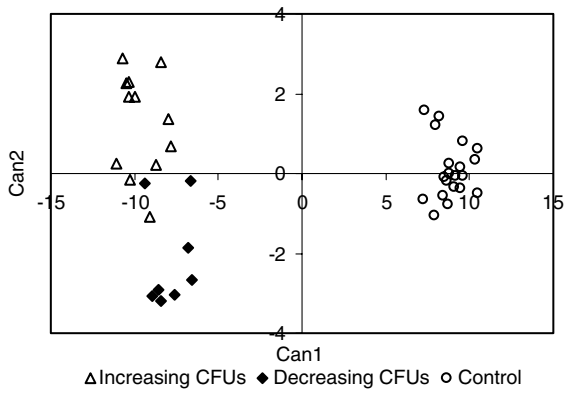


Figure 3. Discriminant analysis of the Biolog data of the copiotrophic bacterial population zero to nine days after grass–clover incorporation into a sandy soil and in the control. All samplings were done in duplicate; daily data were compared to the data on the next sampling day to determine whether CFUs were increasing or decreasing. The clustering of increasing and decreasing data shows that there are different, repetitive, stages in the succession of bacteria in time.

investigate if the decline in bacterial CFUs and the increase in *Pythium* infections were associated with a decrease in oxygen availability.

Spatial wave-like fluctuations of microbial populations in the rhizosphere

Distribution patterns of microbial populations within root systems have been investigated extensively (Schippers and van Vuurde, 1978; van Vuurde and Schippers, 1980; Scott et al., 1995; Kim et al., 1997; Semenov et al., 1999). High microbial densities have generally been observed close to the root tip and in middle and upper sections of the roots, and patterns in microbial density along roots have been thought to be a direct reflection of patterns of exudation and sloughing off of cortex cells (Rovira, 1973; Schippers and van Vuurde 1978; van Vuurde and Schippers, 1980; McCully and Canny, 1985). This is a rather static viewpoint in which growth and death of microbial populations is not explicitly considered. After a series of experiments on the distribution of microbial populations along roots, we arrived at a very different and dynamic concept of microbial community development in the rhizosphere, namely that bacterial communities respond to the influx of nutrients from a root tip with growth and death cycles at any location where the root tip passes, resulting in wave-like patterns along each root (Semenov et al., 1999; van Bruggen et al., 2000; Zelenev et al., 2000).

In the above-mentioned experiments, wheat plants (*Triticum aestivum* L.) were grown in 60 or 90 cm long root observation boxes with soil high or low in fresh organic matter. After two, three, and five weeks, 2 cm root sections were cut at 4 cm intervals. Copiotrophic and oligotrophic bacteria were isolated from the rhizosphere and corre-

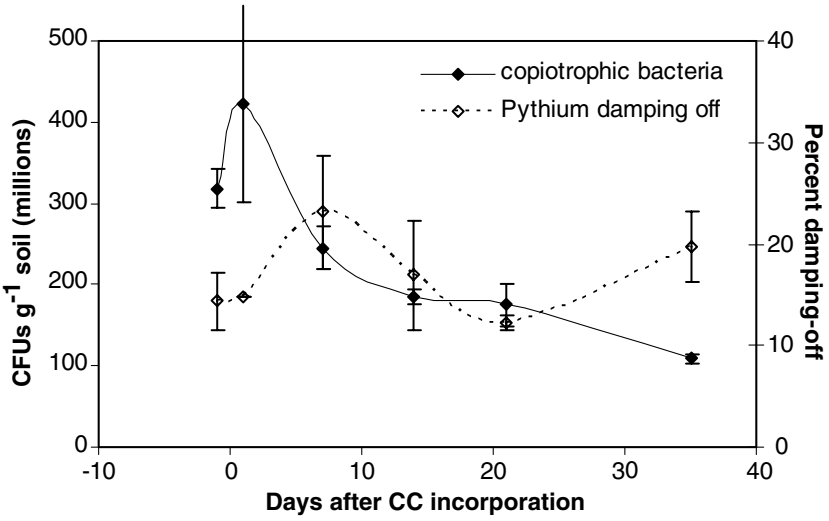


Figure 4. The CFUs of copiotrophic bacteria and the percentage damping-off of tomato seedlings caused by *Pythium ultimum* and *Pythium aphanidermatum* naturally occurring in soils collected one day before, one day after and one, two, three and five weeks after incorporation of a vetch/oats cover crop (after van Bruggen and Semenov, 2000).

sponding bulk soil on carbon-rich and carbon-poor media, respectively (Semenov et al., 1999). For the first time, wave-like distributions of bacterial populations were demonstrated along plant roots using harmonics analysis (Semenov et al., 1999). Peaks in oligotrophic populations were slightly shifted upwards on the root compared to those of copiotrophic populations, indicative of the possibility that oligotrophs would follow copiotrophs in a succession starting from the tip. There were no (cross) correlations of either bacterial group with number of mature lateral roots per section, or with concentrations of soluble total organic carbon (TOC) in the rhizosphere (Semenov et al., 1999). The oscillations shifted from week to week, and were justifiably called 'moving waves'.

To ascertain that the spatial pattern in microbial populations was not related to lateral root formation, we did an experiment with an artificial nutrient source moving through soil. A tube, through which a solution with sugars and amino acids was pumped, was pulled at a speed of 1 or 4 cm per day through a dialysis sleeve buried in soil. This experimental setup gave the expected wave-like patterns in bacterial populations along the path of the moving nutrient source similar to the patterns in real rhizospheres along wheat roots (van Bruggen et al., 2000). Oscillations in space were transformed to oscillations in time, taking the moving rate of the tube into account. This resulted in oscillations with similar periods, regardless of the moving rate of the tube, indicating that the periods are dictated by growth and death rates of the bacteria, not by the growth rate of a root.

These experiments led to the so-called moving-wave hypothesis for bacterial populations in the rhizosphere: 'Waves originate from bacterial growth on nutrients from the root tip, followed by death when nutrients become exhausted and re-growth from recycled carbon sources plus substrate from soil organic matter'. This hypothesis was visualized by means of the results of a simulation model (Zelenev et al., 2000). We envisage the following scenario. As the root tip moves into bulk soil, releasing nutrients, dormant bacteria (and probably fungi) are activated, grow, and then die as nutrients become exhausted; dead bacteria lyse and a new generation grows on recycled nutrients (plus additional substrate from soil and roots). Thus, there are growth and death cycles at any point where the nutrient source passes resulting in waves in space (Figure 5).

Not only total bacterial communities, but also individual bacterial strains exhibit wave-like fluctuations along roots. The biocontrol agent *Pseudomonas fluorescens* 32-*gfp*, marked with the green fluorescent protein gene, was added to soil samples from neighbouring conventional and organic farms at Heelsum, the Netherlands, and re-isolated from the rhizosphere along the total length of wheat roots after three weeks of growth. Both CFUs on selective media and fluorescent microscopic counts oscillated significantly and similarly along the length of the roots (Semenov et al., 2004). The oscillations had a much greater amplitude and period in the conventionally than in the organically managed soil (Figure 6). In the last soil, *P. fluorescens* 32-*gfp* populations were zero towards the root tip. The reason was the lower survival of *P. fluorescens* 32-*gfp* in the organically than in the conventionally managed soil, presumably due to more intense competition in the organic soil. Apparently, when root tips reached a depth of 10–35 cm below soil level, the majority of the intro-

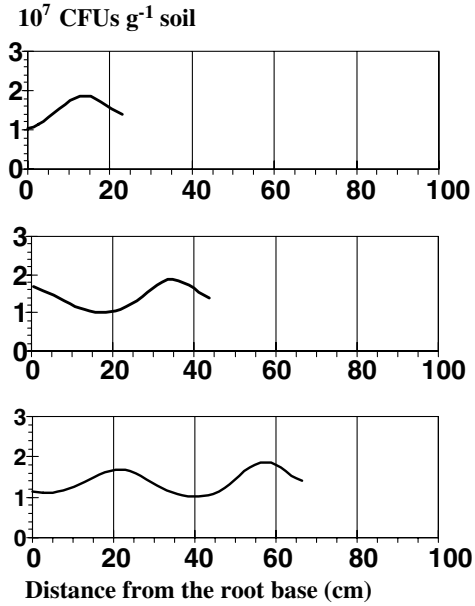


Figure 5. Moving waves: Oscillations in numbers of copiotrophic bacteria along the root in distance from the root base after two, three and four weeks of growth of wheat. With the growing root tip (to the right of each graph) the population moves to a maximum, followed by a harmonic iteration of minima and maxima.

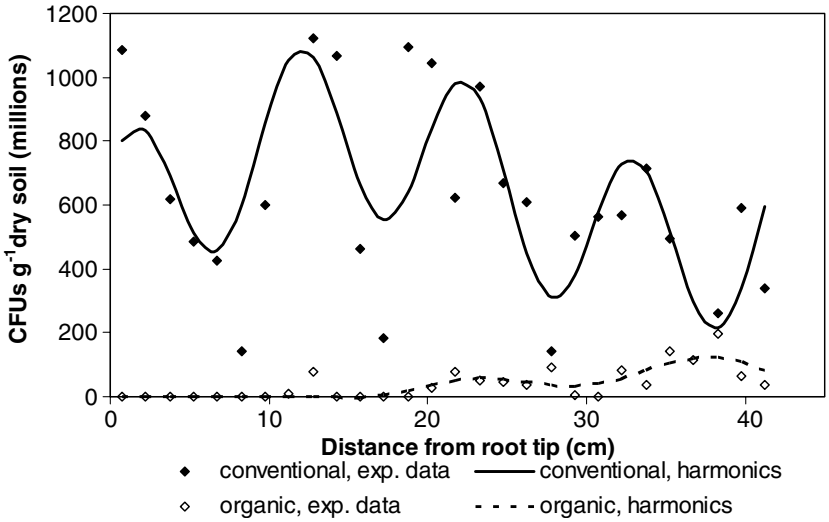


Figure 6. Experimental data and harmonic fluctuations of a GFP-labelled strain of biocontrol agent *P. fluorescens* (CFUs/g^{-1}dry - soil) along the root of wheat in a conventionally and an organically managed sandy soil plotted against distance (cm) from the root tip. The conventional soil had a higher available N and K content and a higher pH than the organic soil. Organic C contents were similar. The contribution of the harmonics to the total variance was 75.6% for the conventional soil and 82.5% for the organic soil. The amplitudes of *P. fluorescens* in the organic soil were lower than in the conventional soil and *P. fluorescens* could not be detected any more around the organic root tip, probably due to a reduced survival in the organic soil compared to the conventional soil.

duced cells had died, so that no cells or CFUs were detected in this region at the time of sampling.

Furthermore, we investigated if the bacterial communities fluctuated as a whole along the wheat root or whether there is a succession in bacterial composition from peak to peak or within peaks (van Diepeningen, pers. comm.). Therefore, rhizosphere microbial communities along roots of wheat were studied in detail (20–25 rhizosphere and bulk soil samples along the total root length) by colony enumeration and DGGE analysis of extracted DNA in the same organic and conventional soils used for the experiments with *P. fluorescens* 32-*gfp*. Similar to our previous findings, the numbers of copiotrophic and oligotrophic bacteria oscillated with significant harmonics along each root, independent of soil moisture or lateral roots. The oscillations and rhizosphere effects were more pronounced in the conventionally managed soil. For amplified eubacterial 16S rDNA fragments from DGGE analysis three different groups could be distinguished: those fluctuating in intensity in phase with CFU oscillations (19.2% of the total number of bands, representing 37.4% of the total band intensity); those fluctuating in intensity in opposite phase with CFU oscillations (26% of the total number of bands, representing only 25.0% of the total band intensity), and remaining bands whose intensity showed no relationship with CFU oscillations or that were restricted to certain root zones (54.8% of the total number of bands, representing only 37.5% of the total band intensity). Discriminant analysis of the bacterial populations in root sections with increasing and decreasing phases in the oscillations showed that the community compositions of waxing populations are more similar to each other than to those of waning populations, especially in conventionally managed soil (Figure 7). Again the succession appeared to be cyclic, in space as well as over time.

Two measures of bacterial biodiversity in soil, species richness S and the Shannon index H, were calculated based on the DGGE data. Both biodiversity measures oscillated with significant harmonics along the root in opposite phase to total bacterial CFUs. The bacterial diversity along the root was negatively correlated with the numbers of oligotrophic and copiotrophic bacterial CFUs in the conventional soil and with oligotrophic bacterial CFUs in the organic soil

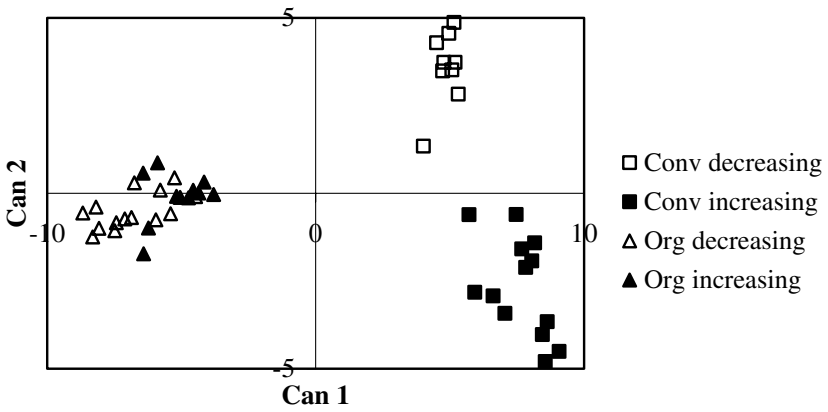


Figure 7. Discriminant analysis of populations in the increasing and decreasing parts of the wave-like oscillations in the rhizosphere of wheat, in an organic and conventional sandy soil. Soil samples originated from neighbouring organic and conventional farms. The conventional soil had a higher available N and K content and a higher pH than the organic soil. Organic C contents were similar.

(Table 1). This indicates that a limited number of fast growing taxa were growing and dying over time and in space.

Besides fluctuations in the vertical direction along the root, running waves of bacterial populations have also been observed in the horizontal direction away from the root surface. Kozhevin (1989) observed fluctuations in cells of introduced *Bradyrhizobium japonicum* 1021 perpendicular to the length of soybean roots (*Glycine max*) at a microscopic scale (up to 1 mm from the root surface), using immunofluorescence. The pattern of the oscillations shifted over time, and these spatial–temporal distributions were described as 'running waves' of bacteria literally moving towards the root surface (Kozhevin, 1989). In seawater, chemotactic bacteria were shown to occur in concentric spheres with alternating higher and lower bacterial densities around point sources of diffusing nutrients, forming wave-like patterns both in space and over time at scales of a few μm and seconds. These patterns were attributed to the combined effects of molecular diffusion of the attractant, congregation and subsequent dispersal of the motile bacteria (Blackburn et al., 1998), and were simulated by nonlinear diffusion-reaction models. The running waves observed by Kozhevin (1989) may also be the result of diffusion-reaction mechanisms.

As mentioned by Kozhevin (1989), there must be a connection between bacterial oscillations in space and time. Indeed, we demonstrated that spatial wave-like fluctuations of microbial populations along the path of a moving nutrient source could be transformed to spatial moving waves by taking the rate of root growth into account (van Bruggen et al., 2000). The connection between spatial and temporal oscillations was used to create a simulation model to describe and predict microbial dynamics in the rhizosphere (Zelenev et al., 2000). This model could also be used to predict the distribution of infections by pathogens in a root system, since infection could possibly take place more easily when microbial abundance and activity decline, at the waning phases of microbial waves.

Table 1. Cross-correlation coefficients (CCF) between oligotrophic and copiotrophic bacterial CFUs at various depths in the rhizosphere of wheat grown in conventional and organic soil, and band intensity in DGGE gels of amplified 16S rDNA fragments from DNA isolated from corresponding rhizophere samples

		CCF	lag[a]
Conventional soil	Oligotrophic CFUs	−0.552[b]	2
	Copiotrophic CFUs	−0.635	2
Organic soil	Oligotrophic CFUs	−0.466	−1
	Copiotrophic CFUs	n.s.[c]	

Soil samples originated from neighbouring organic and conventional farms. The conventional soil had a higher available N and K content and a higher pH than the organic soil. Organic C contents were similar.
[a]One lag corresponds to 1.5 cm.
[b]Significant at $P = 0.05$.
[c]Not significant.

Wave-like fluctuations of plant pathogens in the rhizosphere

The occurrence of microbial growth and death cycles at any point along a root could have important consequences for infection by plant pathogens. Infections by plant pathogenic fungi are rarely uniformly distributed in the root system. Some fungi preferentially infect in the vicinity of the root tip, while others infect primarily older sections of the roots. Given that there are wave-like patterns of saprotrophic microbial populations in space and time in the rhizosphere along roots, and that pathogens often have a saprotrophic phase before infecting a host, it is likely that there are also wave-like patterns in root infections. Indeed, when sclerotia of *Rhizoctonia solani* were placed uniformly along the total length of wheat roots growing in root observation boxes, the proportions of root sections (of eight roots) from which *R. solani* were isolated showed wave-like fluctuations when detrended data were plotted *versus* distance from the root tip (van Bruggen et al., 2002). Similarly, the proportions of root sections from which naturally occurring *Pythium ultimum* was isolated were distributed in a wave-like fashion along the root (Figure 8). The first peak in *Pythium* infections was closer to the root tip than that of *R. solani* (van Bruggen et al., 2002). In the same experiments, copiotrophic bacteria were enumerated at the time of inoculation with *R. solani*, one week before isolation of the pathogens from root sections. For comparison of peaks in infection with those in bacterial populations, the bacterial curves were shifted 14 cm to the right since the root tip moved down 2 cm day^{-1} during the week since the bacterial populations had been assessed. Both pathogens oscillated in a different phase relative to the bacterial oscillations. There were negative correlations between densities of copiotrophic bacteria and *R. solani* infections at 0 cm lag, while there were positive correlations between copiotrophic bacteria and *Pythium* infections at a lag of 6 cm (Figure 8). Infection by *R. solani* was probably inhibited when large bacterial populations were encountered on the root surface at the time of inoculation. It is not known when *P. ultimum* infection took place, but possibly a few days after passing of the root tip, when the first wave of copiotrophic bacteria was already declining. This shift in *Pythium* infection relative to the first peak in copiotrophic bacteria after passing of the root tip is similar to the shift in the *Pythium* damping-off peak relative to the first peak in bacterial CFUs after grass–clover incorporation in soil, as discussed in a previous section of this paper.

In addition to the oscillations in root infections along the length of a root, there is the probability of root infection by plant pathogens located at increasing distances from the root surface that can fluctuate in space. The probability of infection generally declines with perpendicular

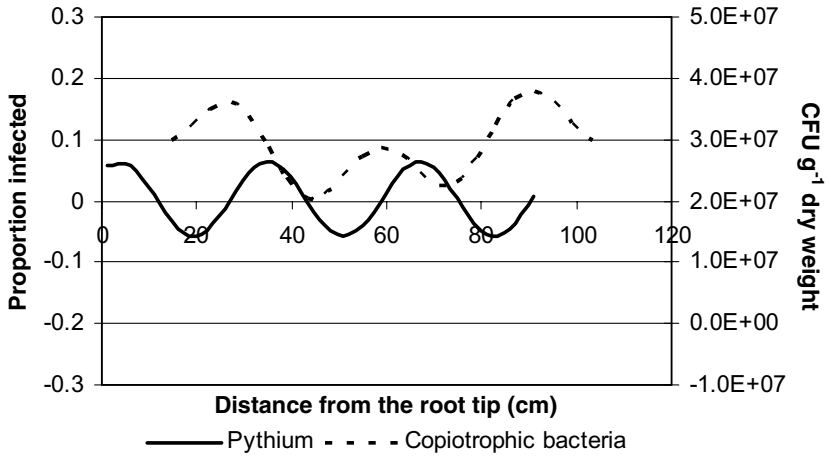


Figure 8. Harmonic fluctuations of the proportion of natural *Pythium* infection and the numbers of copiotrophic bacteria along the wheat root.

distance of propagules from the root, but the decline is generally not monotonous (Mol and van Riessen, 1995; Gilligan and Bailey, 1997). Gilligan and Bailey (1997) placed propagules of *R. solani* at 1–4 mm intervals for a maximum distance of 15 mm horizontally from the host (radish seed) and calculated probability of infection with distance from the host. Close to the host, there was frequently a small peak in the probability of infection, followed by a decline in this probability with distance from the host. After detrending and estimating some missing data by linear interpolation, we re-analyzed their data with harmonics analysis. Most of the data presented by Gilligan and Bailey (1997) actually declined in a wave-like fashion with distance from the host (Figure 9). Harmonics analysis showed that there were significant waves in the horizontal direction, similar to the waves found for *B. japonicum* by Kozhevin (1989). Similarly, the data of Mol and van Riessen (1995) on the probability of infection by *Verticillium dahliae* seemed to decrease in a wave-like fashion with increasing distance from the root surface of various crop species and potato cultivars; unfortunately, the published data were not suitable for harmonics analysis.

The realization that wave-like patterns of saprotrophic and pathogenic microorganisms occur both in the horizontal and vertical direction in the rhizosphere, led to a new view of root infection in relation to microbial population dynamics. Although not demonstrated, it is likely that wave-like distribution patterns of the microbial community are generated along any roots, including lateral roots, initiated from the growing root tips.

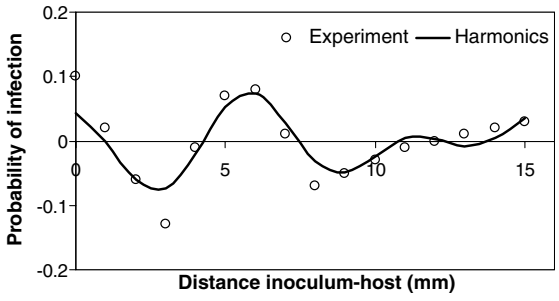


Figure 9. The probability of infection by *Rhizoctonia solani* in relation to distance from the host. Detrended values are shown (derived form Gilligan and Bailey, 1997).

Exudation from root tips leads to creation of waves both in the vertical direction (macro-waves along the root) and in the horizontal direction (micro-waves perpendicular to the root).

The reasons for fluctuating probabilities of infection with distance of propagules from the root are not clear. If propagules of a pathogen are randomly or regularly distributed in space and a root passes releasing exudates, then hyphae growing towards the roots arrive at different distances from the tip depending on their original distance from the host. Upon arrival at the root surface they encounter more or less bacteria, depending on the phase of bacterial waves, so that they have a lower or higher chance of infecting the host. This would result in fluctuating probabilities of infection with distance from the root, and waves of infection along the length of the root.

Alternatively, if there are waves of nutrients moving into the rhizosphere as a result of waves in substrate utilization and release at the root surface or due to a day–night rhythm in exudation, the probability of infection may also fluctuate. If zoospores, for example, are pulled three steps forward and one backward as the waves pass, synchronization of zoospores would take place, so that they arrive in waves at the root surface just behind the root tip. These horizontal dynamics of pathogens would result in waves of infection along the root as the root tip moves on (provided that the root tip does not die from *Pythium* infection). Thus, the occurrence of microbial growth and death cycles at any point along a root does indeed seem to have important consequences for infection by plant pathogens along the root.

Management of soil-borne disease taking microbial oscillations into account

Based on the premise that microbial communities in healthy soils have strong resistance and resilience against disturbances and suppress disease outbreaks (Rapport, 1995), we would need to manage microbial communities so that the amplitudes of the oscillations (resistance) and the time to return to quasi-stationary conditions (resilience) are minimized. To accomplish this, we would need to enhance the biological buffering capacity of a soil by enhancing the background level of microbial activity and food web complexity, for example by

regular additions of relatively stable organic matter to soil. On the other hand, we may want to stimulate a specific part of the microbial community to antagonize certain plant pathogens by applying appropriate external disturbances to stimulate oscillations and succession, for example by tillage. In this paragraph, we will only discuss those management strategies that influence short-term oscillations of microbial populations, and saprotrophic growth and infection by plant pathogens. Management practices that affect long-term population dynamics of microbial communities, such as crop rotation, will not be discussed here.

Organic matter management

Soil organic matter has generally not been managed explicitly by conventional farmers. Addition of synthetic fertilizers over many years has resulted in loss of organic matter by stimulating decomposition of native soil organic matter and enhancing microbial respiration. Combined with the negative effects of synthetic pesticides on components of the microbial communities and associated food webs, conventionally managed soils are generally biologically impoverished. Any addition of substrate for microbial growth such as crop residues, dead organisms after tillage, or root exudates, will result in large oscillations in microbial populations, including fast-growing plant pathogens with saprotrophic abilities such as *Pythium* spp. (van Bruggen and Semenov, 2000).

Large pools of mineral nitrogen in soil may even exacerbate the fluctuations. Soil and plant nitrogen concentrations can have a profound effect on both ecosystem health and disease severity: high levels of nitrogen in the soil, particularly in the form of nitrate, may enhance several fungal diseases (Workneh et al., 1993; van Bruggen, 1995; Tamis and van den Brink, 1998, 1999; Clark et al., 1999).

Several things can be done to restore the biological buffering capacity and enhance internal nutrient cycling in soil. Regular addition of fairly stable organic matter, including solid animal manure, composts of plant and animal origin, and lignified roots of deep-rooted plants such as alfalfa, rye or grass–clover would enhance microbial biomass, activity, and diversity, and food web complexity in soil (Sivapalan et al., 1993; Ryan et al., 1994; Workneh and van Bruggen 1994; Mäder et al., 2002; Schjønning et al., 2002; van Diepeningen et al., 2005). It would also enhance suppression of many soil-borne pathogens (van Bruggen, 1995; van Bruggen and Termorshuizen, 2003). For example, densities of *Phytophthora* and *Pythium* propagules in soil were lower and those of the antagonist *Trichoderma* higher in soils amended with various organic materials (composted cottongin trash, composted yard waste, or cattle manure) than with synthetic fertilizer (Bulluck et al., 2002). Wave-like responses of these pathogens to introduction of the organic materials were not investigated.

Considering the reaction to a disturbance by incorporation of a winter cover crop or weeds into soil, sowing of a subsequent crop needs to be timed so that the inoculum of a facultative saprotrophic pathogen is not at its peak at that time. The quality of the organic matter in terms of easily available substrate and the C:N ratio, and the activity of the microbial community will determine if facultative saprotrophic pathogens can multiply in this material. Pathogens such as *Pythium* and *Rhizoctonia* species multiply easily in fresh substrate, and may cause serious damping-off problems when a crop is sown within three to four weeks after incorporation of fresh plant material (van Bruggen and Semenov, 2000). For example, when a mixture of vetch and oats was incorporated in soil that had been fallow or had been cover-cropped, damping-off of tomatoes by *Pythium* species was most severe seven days after incorporation of the plant material, and five days after the first peak in copiotrophic bacteria (Figure 4). The peaks in bacterial CFUs and damping-off incidence were higher in the previously fallowed soil than in the cover-cropped soil, indicating that the cover-cropped soil was more stable (van Bruggen and Semenov 1999, 2000). In another study where a vetch–oats cover crop was incorporated in organically and conventionally managed soils, *in vitro* growth of *P. aphanidermatum* peaked after 7–10 days while that of *R. solani* peaked after 21–35 days (Grünwald et al., 1997, 2000). Microbial measurements were generally lower and *in vitro* growth of the pathogens higher in the conventionally compared to the organically managed soils, but these differences were temporarily nullified after cover crop incorporation (Grünwald et al., 2000).

Different from the effects of cover crop incorporation, high-nitrogen-containing organic amendments such as cattle and poultry manure or soy meal had an immediate suppressive effect on several root pathogens and nematodes, as a result of ammonia release immediately after initiation of microbial decomposition (Lazarovits et al., 2001). Bulluck et al. (2002) also documented immediate suppressive effects of various types of compost and cattle manure applied at moderate to high rates on southern blight of tomatoes. However, these materials may not be suppressive to *Pythium*, which thrives well under high ammonium concentrations (van Bruggen and Semenov, 1999).

After repeated applications of organic materials, higher organic matter and microbial activity in bulk soil would result in a 'masking of the rhizosphere effect' (Gilbert et al., 1994), reduce the microbial oscillations along roots, and limit substrate concentrations seeping into soil, thereby reducing the attraction of root pathogens to the root surface and decreasing the chance of infection by many pathogens. Reducing the amount of easily available mineral nutrients and soluble carbon compounds by reducing fertilizer applications and the addition of stable carbon sources, would lead to oligotrophication. This promotes mycorrhizal infections, which can also suppress various root diseases (Sharma et al., 1992; Ryan et al., 1994). Whether mycorrhizal infections also occur in wave-like patterns along the roots is not known.

Tillage

No-till or reduced tillage has been promoted in recent years primarily to reduce soil erosion. However, tillage practices also have pronounced effects on survival of fungi and micro- and macrofauna in soil. Deep tillage can enhance the bacteria to fungi ratio and eliminate predatory nematodes, affecting especially the *k*-strategists (Berkelmans et al., 2003). No-till or reduced tillage is often associated with higher microbial biomass and activity and a more complex food web in the upper soil layers compared to regular tillage, i.e. plowing (van Diepeningen et al., 2005).

Tillage is a form of disturbance resulting in clear fluctuations in microbial activity and biomass. Caldéron et al. (2000) showed clear fluctuations in microbial biomass during the first eight days after simulated tillage in the laboratory, namely mixing of soil samples collected from a grassland and a vegetable field. Such a disturbance may also give facultative saprotrophic pathogens a chance to grow, but wave-like fluctuations may be dampened sooner than in the case of a disturbance by fresh plant materials. A report on a lower incidence of *Pythium* damping-off of sugar beet in a farm with reduced tillage than in a conventional farm with regular tillage (El Titi and Richter, 1987) is in agreement with the notion that reduced tillage decreases the chance that *Pythium* would grow explosively in fresh substrate after tillage. On the other hand, pathogens that survive in stubble could become problematic in no-till fields. Roget (1995) demonstrated that after conversion from regularly tilled to no-till wheat production *Rhizoctonia* root rot increased in the first few years. However, this increase was followed by a decline in *Rhizoctonia* root rot after about five years of no-till.

(Partial) soil sterilization

It is well known that soil-borne plant pathogens can wreak havoc when introduced into steamed greenhouse soil or fumigated field soil due to the existence of a biological vacuum (Bollen, 1974; Kreutzer, 1965). Any disturbance of a recently sterilized (or γ-irradiated) and re-colonized soil leads to wild fluctuations in microbial populations (Zelenev et al., 2004), and may lead to similar fluctuations in facultative saprotrophic plant pathogens. A good alternative to soil sterilization may be biological soil disinfestation (Blok et al., 2000), which does not result in enhanced disease pressure when pathogens are re-introduced, and provides long-lasting disease control (Goud et al., 2004).

Mixed cropping

Mixed cropping – a system where two or more crops are grown in the same field – can enhance food web diversity and decrease severity of foliar plant diseases (Finckh and Wolfe, 1997). Although positive correlations between above-ground and below-ground biodiversity have seldom been demonstrated (De Deyn et al., 2004), suppression of root disease (Burdon and Chilvers, 1976; Villich, 1993) and enhanced soil microbial diversity in mixed cropping systems have sometimes been found (G.A. Hiddink, pers. comm.).

Microbial composition in the rhizosphere is strongly dependent on plant species (Smith et al., 1999). Thus in a mixture of roots of different species it may be more difficult for a pathogen to find its host and the saprotrophic phase of pathogens may be limited by a greater variety of antagonists.

Cultivar selection

Choice of crops and cultivars will influence the microbial communities that are selectively enhanced or suppressed in the rhizosphere by the quality of root exudates (Grayston et al., 1998; Kowalchuk et al., 2002; Garbeva et al., 2004). Differential interactions between plant genotypes and beneficial microorganisms have been demonstrated for species of mycorrhizal fungi, rhizobia, and general plant growth-promoting rhizobacteria (PGPR) (Smith et al., 1999). PGPR have been used as biofertilizers and biological control agents (Germida, 1996). They can be directly antagonistic towards plant pathogens or can stimulate systemic induced resistance in the plant (Kloepper et al., 1997).

Besides exudate quality, exudation rates can also vary per cultivar, and these rates determine to a large extent the amplitude of the ensuing microbial oscillations. However, plant breeders have generally not taken exudation rates and exudate quality into account. There would be a great opportunity to select cultivars for their ability to stimulate specific microbial communities that can contribute to disease suppression.

Biological control

Many microorganisms have been found with biological control potential against various plant pathogens. Biological control agents may use a variety of inhibitory and suppressive mechanisms: (1) competition for resources and space, (2) antibiotic production, (3) removal of pathogenicity factors produced by the pathogen, (4) production of degrading enzymes that target the pathogen and (5) the induction of resistance in the host plant (Whipps, 2001). However, many biological control agents perform poorly under field conditions (Fravel, 1999) and only few biocontrol species have been registered for field use (Copping, 2001). Biocontrol of soil-borne pathogens has been more successful under controlled environmental conditions using simplified potting mixes presumably low in microbial diversity (Fravel, 1999). Moreover, inoculation of soil with a single strain of a biocontrol agent rarely leads to a high level of protection and often the positive effect is inconsistent (Weller, 1988; Koch, 1999). Better results have been obtained with combinations of strains or species (e.g. Pierson and Weller, 1994; Guetsky et al., 2001, 2002; Szczech and Shoda, 2004).

These results with biocontrol agents can now be interpreted in view of the general occurrence of microbial oscillations in time and space in the rhizosphere. Introduced biocontrol agents are likely to oscillate similarly to the native soil microbial communities. Densities of *Pseudomonas fluorescens* introduced on wheat seed seemed to form wave-like patterns along the length of the root, the amplitudes tapering off towards the root tip (Scott et al., 1995). In an experiment in our laboratory with GFP labelled, phloroglucinol-producing *P. fluorescens* mixed into soil we conclusively proved a wave-like distribution of this bacterium along growing wheat roots (Figure 6), similar to the oscillations of native bacterial populations (Semenov et al., 1999). We also showed that there is a succession in microbial communities within each wave, repeating from wave to wave. Thus, microorganisms that may be good antagonists *in vitro*, may take a different position in the succession compared to the target pathogen, and may therefore not be effective as biocontrol agents. This might mean that biocontrol can only be accomplished if waves in populations of the biocontrol agent coincide more or less with potential waves in pathogen populations (unless there is systemic induced resistance). Potential biocontrol agents may need to be selected so that their populations are maximal in the region along the root where the target pathogen invades the root. Differences in succession and position along the root may also explain the greater success of biocontrol mixtures than of single biocontrol agents (Guetsky et al., 2001, 2002; Szczech and Shoda, 2004).

In organically managed soil with high microbial diversity and activity (and therefore low concentrations of easily available nutrients) introduced biocontrol agents may survive less compared to biologically impoverished conventional soil (van Bruggen et al., 2000). The difference between the effectiveness of biocontrol

agents in biologically impoverished soil *versus* biologically diverse and active soils, may be explained by reduced availability of easily decomposable substrate in these latter soils. The rhizosphere effect would be masked and microbial oscillations subdued in such soils (van Diepeningen, pers. comm.). In biologically impoverished soils, high substrate concentrations can be expected where wildly oscillating microbial communities are at a minimum; biocontrol agents with a slightly different niche compared to the majority of these oscillating communities may have a chance to survive and even grow. However, they may not grow in a rhizosphere with very diverse microbial communities. An introduced phloroglucinol-producing, gfp marked *P. fluorescens* strain declined faster in three organically managed soils than in three neighbouring conventionally managed soils (van Bruggen et al., 2004). The same strain showed only mild wave-like oscillations along a wheat root in an organically managed soil compared to a conventionally managed soil (Figure 6), and had less effect on take-all disease of wheat in soils from three organic farms than in soils from three neighbouring conventional farms with a lower microbial diversity (van Bruggen et al., 2004). Thus, it is questionable if inundative biological control can be effective in soils with a high microbial biomass, activity and diversity, and low levels of easily available substrate.

Conclusions

In this review we showed that populations of different trophic groups of bacteria develop in a wave-like fashion with repetitive growth and death cycles, both in time and space after an impulse of readily utilizable substrate. Oscillatory development of bacterial populations may be a universal phenomenon after a disturbance, which could possibly be used to compare soils in terms of stability and resilience, and consequently soil health. Indeed, the amplitudes of the oscillations are smaller and decline more quickly in soils with a high microbial biomass, activity and diversity, and low levels of easily available substrate. These are characteristics of soil health. It is argued that healthy soils are more suppressive to soil-borne plant pathogens than biologically impoverished soils.

Single species of saprotrophic bacteria, biocontrol agents and phytopathogenic fungi also show wave-like fluctuations in bulk soil and along plant roots. Different trophic groups and species may fluctuate with different periods and phases. A cyclic succession occurs in response to nutrient input; in the waxing phases of successive oscillations, microbial communities are taxonomically and physiologically more similar to one another than to the communities in the waning phases. This has consequences for the selection of biocontrol agents and cultivars. Introduction of a single biocontrol agent to a soil may not lead to the expected results due to wave-like fluctuations in the rhizosphere of the biocontrol agent and the target pathogen, if they are out of phase. A mixture of biocontrol agents of different trophic groups may be more successful.

The main strategies to control soil-borne diseases can be classified into three categories: (1) enhancement of general microbial biomass and diversity resulting in a masking of the rhizosphere effect, a reduction of the amplitude of wave-like oscillations and an increase in natural disease suppression, (2) removal of dormant propagules or pathogens in their saprotrophic phase from their food base by stimulating wave-like fluctuations in populations of potentially competitive microorganisms, for example by soil tillage, and (3) augmentation of microbial communities by biocontrol agents, which must be able to survive and grow in the rhizosphere. The first strategy is the main strategy used by organic farmers, while the second and third strategies are typical for conventional farms. In all cases, dynamic oscillations of microbial communities and individual species must be taken into account. This constitutes a new view of plant disease control.

Acknowledgements

We would like to thank Vladimir V. Zelenev for carrying out harmonic analyses and Gerbert Hiddink for sharing his unpublished data. This work was partially supported by NWO Russia collaborative grant (Dossier number 047.014.001) and NATO-Russia Collaborative Linkage Grant (RCLG) 'Risk analysis of pathogen spread in the vegetable production and processing industry'.

References

Aristovskaya TV (1980) Microbiological Processes of Soil Formation. Nauka, Leningrad (in Russian).

Berkelmans R, Ferris H, Tenuta M and van Bruggen AHC (2003) Long-term effects of crop management on trophic levels of nematodes other than plant parasitic nematodes disappear after one year of uniform treatment. Applied Soil Ecology 23: 223–235.

Blackburn N, Fenchel T and Mitchell J (1998) Microscale nutrient patches in planctonic habitats shown by chemotactic bacteria. Science 282: 2254–2256.

Blok WJ, Lamers JG, Termorshuizen AJ and Bollen GJ (2000) Control of soilborne plant pathogens by incorporating fresh organic amendments followed by tarping. Phytopathology 90: 253–259.

Boehm MJ, Madden LV and Hoitink HAJ (1993) Effect of organic matter decomposition level on bacterial species diversity and composition in relationship to Pythium damping-off severity. Applied and Environmental Microbiology 59: 4171–4179.

Boehm MJ, Wu T, Stone AG, Kraakman B, Iannotti DA, Wilson GE, Madden LV and Hoitink HAJ (1997) Cross-polarized magic-angle spinning 13C nuclear magnetic resonance spectroscopic characterization of soil organic matter relative to culturable bacterial species composition and sustained biological control of Pythium root rot. Applied and Environmental Microbiology 63: 162–168.

Bollen GJ (1974) Fungal recolonization of heat-treated glasshouse soils. Agro-Ecosystems 1: 139–155.

Brussaard L, Kuyper TW, Didden WAM, de Goede RGM and Bloem J (2004) Biological soil quality from biomass to biodiversity – importance and resilience to management stress and disturbance. In: Schjønning P, Elmholt S and Christensen BT (eds) Managing Soil Quality: Challenges in Modern Agriculture (pp. 139–161). CAB International.

Buckley DH and Schmidt TM (2001) The structure of microbial communities in soil and the lasting impact of cultivation. Microbial Ecology 42: 11–21.

Bulluck LR, Brosius N, Evanylo GK and Ristaino JB (2002) Organic and synthetic fertility amendments influence soil microbial, physical and chemical properties on organic and conventional farms. Applied Soil Ecology 19: 147–160.

Burdon JJ and Chilvers GA (1976) Epidemiology of Pythium-induced damping-off in mixed species seedling stands. Annuals of Applied Biology 82: 233–240.

Calderon FJ, Jackson LE, Scow KM and Rolston DE (2000) Microbial responses to simulated tillage in cultivated and uncultivated soils. Soil Biology and Biochemistry 32: 1547–1559.

Clarholm M (1981) Protozoan grazing of bacteria in soil – impact and importance. Microbial Ecology 7: 343–350.

Clark MS, Horwath WR, Shennan C, Scow KM, Lantni WT and Ferris H (1999) Nitrogen, weeds and water as yield-limiting factors in conventional, low-input, and organic tomato systems. Agriculture, Ecosystems & Environment 73: 257–270.

Copping LG (2001) The Biopesticide Manual, British Crop Protection Council, Surrey, UK.

De Deyn GB, Raaijmakers CE and Van der Putten WH (2004) Plant community development is affected by nutrients and soil biota. Journal of Ecology 92: 824–834.

De Vos OJ and van Bruggen AHC (2001) Soil microbial composition as affected by grass–clover mixture incorporation in the soil. ISME-9, Amsterdam, the Netherlands, P17.077.

Doebeli M and Ruxton GD (1997) Evolution of dispersal rates in metapopulation models: Branching and cyclic dynamics in phenotype space. Evolution 51: 1730–1741.

Doebeli M and Ruxton GD (1998) Stabilization through spatial pattern formation in metapopulations with long-range dispersal. Proceedings of the Royal Society of London Series B – Biological Sciences 265: 1325–1332.

Drinkwater LE, Letourneau DK, Workneh F, van Bruggen AHC and Shennan C (1995) Fundamental differences between conventional and organic tomato agroecosystems in California. Ecological Applications 5: 1098–1112.

El Titi A and Richter J (1987) Integrierter Pflanzenschutz im Ackerbau: Das Lautenbach Projekt. III. Schädlinge und Krankheiten 1979–1983. Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz 94: 1–13.

Finckh MR and Wolfe MS (1997). The use of biodiversity to restrict plant diseases and some consequences for farmers and society. In: Jackson LE) Ecology in Agriculture (pp 203–237) Academic Press, San Diego, CA, USA.

Franke-Snyder M, Douds DD, Galvez L, Phillips JG, Wagoner P, Drinkwater L and Morton JB (2001) Diversity of communities of arbuscular mycorrhizal (AM) fungi present in conventional versus low-input agricultural sites in eastern Pennsylvania, USA. Applied Soil Ecology 16: 35–48.

Fravel D (1999) Hurdles and bottlenecks on the road to biocontrol of plant pathogens. Australasian Plant Pathology 28: 53–56.

Garbeva P, van Veen JA and van Elsas JD (2004) Microbial diversity in soil: Selection of microbial populations by plant and soil type and implications for disease suppressiveness. Annual Review of Phytopathology 42: 243–270.

Gerlagh M (1968) Introduction of *Ophiobolus graminis* into new polders and its decline. Agricultural Research reports 713. Centre for Agricultural Publishing and Documentation, Wageningen.

Germida JJ (1996) Use of rhizobacteria as biofertilizers for enhancing growth and yield of crops. In: Scoles G and Rossnagel B (eds) Proceedings V Int. Oat Conf. & VII Int. Barley Genetics Symp. Univ. Extension Press.

Gilbert GS, Handelsman J and Parke JL (1994) Root camouflage and disease control. Phytopathology 84: 222–225.

Gilligan CA and Bailey DJ (1997) Components of pathozone behaviour. New Phytologist 136: 343–358.

Goud JC, Termorshuizen AJ, Blok WJ and van Bruggen AHC (2004) Long-term effect of biological soil disinfestation on *Verticillium* wilt. Plant Disease 88: 688–694.

Grayston SJ, Wang S, Campbell CD and Edwards AC (1998) Selective influence of plant species on microbial diversity in the rhizosphere. Soil Biology and Biochemistry 30: 369–378.

Grünwald NJ, Workneh F, Hu S and van Bruggen AHC (1997) Comparison of an *in vitro* and a damping-off assay to test soils for suppressiveness to *Pythium aphanidermatum*. European Journal of Plant Pathology 103: 55–63.

Grünwald NJ, Hu S and van Bruggen AHC (2000) Short-term cover crop decomposition in organic and conventional soils: Characterization of soil C, N, microbial and plant pathogen dynamics. European Journal of Plant Pathology 106: 37–50.

Guetsky R, Shtienberg D, Elad Y and Dinoor A (2001) Combining biocontrol agents to reduce the variability of biological control. Phytopathology 91: 621–627.

Guetsky R, Shtienberg D, Elad Y, Fisher E and Dinoor A (2002) Improving biological control by combining biocontrol agents each with several mechanisms of disease suppression. Phytopathology 92: 976–985.

Hoper H and Alabouvette C (1996) Importance of physical and chemical soil properties in the suppressiveness of soils to plant diseases. European Journal of Soil Biology 32: 41–58.

Kim DS, Weller DM and Cook RJ (1997) Population dynamics of *Bacillus* sp. L324-92R_{12} and *Pseudomonas fluorescens* 2-79RN_{10} in the rhizosphere of wheat. Phytopathology 87: 559–564.

Kloepper JW, Tuzun S, Zehnder GW, Wei G, Hoffland E, Bakker PAHM and vanLoon LC (1997) Multiple disease protection by rhizobacteria that induce systemic resistance-historical precedence. Phytopathology 87: 136–138.

Ko WH (1982) Biological control of Phytophthora root rot of papaya with virgin soil. Plant Disease 66: 446–448.

Koch E (1999) Evaluation of commercial products for microbial control of soil-borne plant diseases. Crop Protection 18: 119–125.

Kowalchuk GA, Buma DS, de Boer W, Klinkhamer PGL and vanVeen JA (2002) Effects of above-ground plant species composition and diversity on the diversity of soil-borne microorganisms. Antonie van Leeuwenhoek 81: 509–520.

Kozhevin PA (1989) Microbial Populations in Nature. Moscow State University Press 174 p. (in Russian).

Kreutzer WA (1965). The reinfestation of treated soil. In: Baker KF and Snyder WC (eds) Ecology of Soil-borne Plant Pathogens (pp 495–508) University of California Press, Berkeley.

Lawlor K, Knight BP, Barbosa-Jefferson VL, Lane PW, Lilley AK, Paton GI, McGrath SP, O'Flaherty SM and Hirsch PR (2000) Comparison of methods to investigate microbial populations in soils under different agricultural management. FEMS Microbiology & Ecology 33: 129–137.

Lazarovits G, Tenuta M and Conn KL (2001) Organic amendments as a disease control strategy for soil-borne diseases of high-value agricultural crops. Australasian Plant Pathology 30: 111–117.

Liiri M, Setala H, Haimi J, Pennanen T and Fritze H (2002) Soil processes are not influenced by the functional complexity of soil decomposer food webs under disturbance. Soil Biology and Biochemistry 34: 1009–1020.

Mäder P, Fließbach A, Dubois D, Gunst L, Fried P and Niggli U (2002) Soil fertility and biodiversity in organic farming. Science 296: 1694–1697.

Mas A, Houwen F and Verstraete W (1996) Agricultural factors affecting methane oxidation in arable soil. Biology and Fertility of Soils 21: 95–102.

McCully ME and Canny MJ (1985) Localization of translocated 14C in roots and root exudates of field grown maize. Physiologia Plantarum 65: 380–392.

Mol L and van Riessen HW (1995) Effect of plant roots on the germination of microsclerotia of *Verticillium dahliae*. I. Use of root observation boxes to assess differences among crops. European Journal of Plant Pathology 101: 673–678.

Mulder C, de Zwart D, van Wijnen HJ, Schouten AJ and Breure AM (2003) Observational and simulated evidence of ecological shifts within the soil nematode community of agroecosystems under conventional and organic farming. Functional Ecology 17: 516–525.

Nitta T (1991) Diversity of root fungal floras: Its implications for soil-borne diseases and crop growth. Japan Agricultural Research Quarterly 25: 6–11.

Oehl F, Sieverding E, Ineichen K, Mäder P, Boller T and Wiemken A (2003) Impact of land use intensity on the species diversity of arbuscular mycorrhizal fungi in agroecosystems of central Europe. Applied and Environmental Microbiology 69: 2816–2824.

Orwin KH and Wardle DA (2004) New indices for quantifying the resistance and resilience of soil biota to exogenous disturbances. Soil Biology and Biochemistry 36: 1907–1912.

Pierson EA and Weller DM (1994) Use of mixtures of fluorescent pseudomonads to suppress take-all and improve the growth of wheat. Phytopathology 84: 940–947.

Rapport (1995) Ecosystem health – more than a metaphor. Environmental Values 4: 287–309.

Roget DK (1995) Decline in root rot (*Rhizoctonia solani* AG-8) in wheat in a tillage and rotation experiment at Avon, South Australia. Australian Journal of Experimental Agriculture 35: 1009–1013.

Rovira AD (1973) Zones of exudation along plant roots and spatial distribution of microorganisms in the rhizosphere. Pesticide Science 4: 361–366.

Ryan MH, Chilvers GA and Dumaresq DC (1994) Colonization of wheat by VA-mycorrhizal fungi was found to be higher on a farm managed in an organic manner than on a conventional neighbour. Plant and Soil 160: 33–40.

Schippers B and van Vuurde JWL (1978). Studies of microbial colonization of wheat roots and the manipulation of the rhizosphere microflora. In: Loutit MW and Miles JAR (eds) Microbial Ecology (pp 295–298) Springer-Verlag, Berlin.

Schjønning P, Elmholdt S, Munkholm LJ and Debosz K (2002) Soil quality aspects of humid sandy loams as influenced by organic and conventional long-term management. Agriculture, Ecosystems & Environment 88: 195–214.

Scott EM, Rattray EAS, Prosser JI, Killham K, Glover LA, Lynch JM and Bazin MJ (1995) A mathematical model for dispersal of bacterial inoculants colonizing the wheat rhizosphere. Soil Biology and Biochemistry 27: 1307–1318.

Semenov AM (2001) Oscillations of microbial communities in soils. Works of All-Russia Conference 'To the 100-Anniversary from Birthday of Academician E.N. Mishustin', 22.02.2001, MSU, Moscow (pp. 57–72), M. MAKC Press (in Russian).

Semenov AM, van Bruggen AHC, van Diepeningen AD, Sayler RJ, Zelenev VV and de Vos O (2004) Wave-like development of bacterial communities and a biocontrol agent introduced in the wheat rhizosphere. Phytopathology 94: S94.

Semenov AM, van Bruggen AHC and Zelenev VV (1999) Moving waves of bacterial populations and total organic carbon along roots of wheat. Microbial Ecology 37: 116–128.

Sharma AK, Johri BN and Gianinazzi S (1992) Vesicular-arbuscular mycorrhizae in relation to plant-disease. World Journal of Microbiology and Biotechnology 8: 559–563.

Shumway RH (1988) Applied Statistical Time Series Analysis, Prentice-Hall, Englewood Cliffs, NJ, 379.

Sivapalan A, Morgan WC and Franz PR (1993) Monitoring populations of soil microorganisms during a conversion from a conventional to an organic system of vegetable growing. Biological Agriculture and Horticulture 10: 9–27.

Smith KP, Handelsman J and Goodman RM (1999) Genetic basis in plants for interactions with disease-suppressive bacteria. Proceedings of the National Academy of Sciences of the United States of America 96: 4786–4790.

Szczech M and Shoda M (2004) Biocontrol of Rhizoctonia damping-off of tomato by *Bacillus subtilis* combined with *Burkholderia cepacia*. Journal of Phytopathology 152: 549–556.

Tamis WLM and van den Brink WJ (1998) Inventarisatie van ziekten en plagen in wintertarwe in gangbare, geïntegreerde en ecologische teeltsystemen in Nederland in de periode 1993–1997. IPO-DLO Rapport nr. 98-01, Wageningen

Tamis WLM and van den Brink WJ (1999) Conventional, integrated and organic winter wheat production in the Netherlands in the period 1993–1997. Agriculture, Ecosystems & Environment 76: 47–59.

vanBruggen AHC (1995) Plant-disease severity in high-input compared to reduced-input and organic farming systems. Plant Disease 79: 976–984.

vanBruggen AHC and Semenov AM (1999) A new approach to the search for indicators of root disease suppression. Australasian Journal of Plant Pathology 28: 4–10.

vanBruggen AHC and Semenov AM (2000) In search of biological indicators for soil health and disease suppression. Applied Soil Ecology 15: 13–24.

vanBruggen AHC, Semenov AM and Zelenev VV (2000) Wave-like distributions of microbial populations along an artificial root moving through soil. Microbial Ecology 40: 250–259.

vanBruggen AHC, Semenov AM and Zelenev VV (2002) Wavelike distributions of infections by an introduced and naturally occurring root pathogen along wheat roots. Microbial Ecology 44: 30–38.

vanBruggen AHC and Termorshuizen AJ (2003) Integrated approaches to root disease management in organic farming systems. Australasian Plant Pathology 32: 141–156.

vanBruggen AHC, Hiddink GA, Semenov AV and Semenov AM (2004) Suppression of take-all disease in soils from organic versus conventional farms in relation to native and introduced *Pseudomonas fluorescens*. Phytopathology 94: S105.

van Diepeningen AD, de Vos OJ, Korthals GW and van Bruggen AHC (2005) Effects of organic versus conventional management on chemical and biological parameters in agricultural soils. Applied Soil Ecology, in press, available on line.

vanVuurde JWL and Schippers B (1980) Bacterial colonization of seminal wheat roots. Soil Biology and Biochemistry 12: 559–565.

Vilich V (1993) Crop rotation with pure stands and mixtures of barley and wheat to control stem and root rot diseases. Crop Protection 12: 373–379.

Weller DM (1988) Biological control of soilborne pathogens in the rhizosphere with bacteria. Annual Review of Phytopathology 26: 379–407.

Whipps JM (1997) Developments in the biological control of soilborne plant pathogens. Advances in Botanial Research 26: 1–134.

Whipps JM (2001) Microbial interactions and biocontrol in the rhizosphere. Journal of Experimental Botany 52: 487–511.

Workneh F, vanBruggen AHC, Drinkwater LE and Shennan C (1993) Variables associated with corky root and phytophthora root rot of tomatoes in organic and conventional farms. Phytopathology 83: 581–589.

Workneh F and van Bruggen AHC (1994) Microbial density, composition, and diversity in organically and conventionally managed rhizosphere soil in relation to suppression of corky root of tomatoes. Applied Soil Ecology 1: 219–230.

Zelenev VV (2004) Spatial and temporal fluctuations in bacteria, microfauna and mineral nitrogen in response to a nutrient impulse in soil. Ph.D. thesis. Wageningen University, Wageningen, the Netherlands.

Zelenev VV, Berkelmans R, van Bruggen AHC, Bongers T and Semenov AM (2004) Daily changes in bacterial-feeding nematode populations oscillate with similar periods as bacterial populations after a nutrient impulse in soil. Applied Soil Ecology 26: 93–106.

Zelenev VV, van Bruggen AHC and Semenov AM (2000) "BACWAVE", a spatial-temporal model for traveling waves of bacterial populations in response to a moving carbon source in soil. Microbial Ecology 40: 260–272.

Zelenev VV, van Bruggen AHC and Semenov AM (2005) Short-term wave-like dynamics of bacterial populations in response to nutrient input from fresh plant residues. Microbial Ecology 49: 83–93.

Zvjagintsev DG and Golimbet VE (1983) Dynamics of microbial number, biomass and productivity of microbial communities in soils. In: Nauka M (ed.) Successes of Microbiology, no. 18 (pp. 215–231) (In Russian).

European Journal of Plant Pathology (2006) 115:123–138
DOI 10.1007/s10658-005-0651-z

Patterns and management of crop multiple pathosystems

Serge Savary[1,2], B. Mille[3], B. Rolland[4] and P. Lucas[3]
[1]*IRD, Centre INRA, BiO3P, France (E-mail: ssavary@bordeaux.inra.fr)* [2]*UMR Santé des Plantes, Centre INRA de Bordeaux, 33883, B.P. 81, Villenave d'Ornon, France;* [3]*INRA, BiO3P, F-35653, Le Rheu cedex, France;* [4]*INRA, APBV, F-35653, Le Rheu cedex, France*

Accepted 1 July 2005

Key words: agro-ecosystem, epidemiological guilds, guilds of harmful agents, IPM, multiple epidemics, syndrome of disease, syndrome of production

Abstract

The study of multiple pathosystems has played a central role in the development of botanical epidemiology, leading to a number of approaches and concepts. Multiple pathosystems are facts, which are experienced by many non-cultivated, or cultivated, plant communities. The shapes and composition of multiple pathosystems vary in space and time because of their inherent structure of relationships, and also in response to management. Examples of variation in multiple pathosystems are given, of groundnut in Côte d'Ivoire, of wheat in Brittany, and of upland rice in northern Laos. Variation in the yield-reducing effects of multiple pathosystems is discussed, including interactions among disease elements, relationships with attainable performances, and linkages with production situations. Progress has been achieved in understanding the links between injury profiles, production situations, and attainable performances. Questions about the functioning and consequences of multiple pathosystems are central to defining the scientific bases for, the design of, and the implementing of IPM. The complexity of multiple pathosystems, however, remains a deterrent, not a challenge, to many plant pathologists. Progress achieved in designing production systems for hardy wheat in France, however, is very promising, because of the multidisciplinary science it involves, and because of the promise to deliver it carries. The concepts of epidemiological guilds and of guilds of harmful agents are offered as perspectives to address and manage syndromes of production and syndromes of disease.

Relevance of multiple pathosystems in botanical epidemiology

Multiple pathosystems as a research theme

The study of multiple pathosystems has played a central role in the development of botanical epidemiology. As a subject, it is the equivalent in botanical epidemiology of community ecology in general ecology. Research in the field has led to the development of a body of approaches, often statistical and multivariate, as the objective often has been mostly descriptive, rather than explanatory. Studies on multiple pathosystems (i) led to attempts to understand and manage them (e.g., Jörg et al., 1987; Daamen et al., 1989; Bastiaans and Daamen, 1994), (ii) resulted in analyses of case-studies, and efforts dealing with specific cases and contexts (e.g., Hamelink et al., 1988; Avelino, 1999), and (iii) often were perceived as practical endeavours only. Studies of multiple pathosystems, dealing with a complex subject, inviting complex analyses, and leading to complex interpretations, have arguably led to results that were difficult to share. Arguably, this type of research often addresses open-ended questions, not specific

hypotheses. Its value however is multifold: first, it provides a framework for other, hypothesis-specific, epidemiological research; second, it allows the description so furthering hypotheses for future research; and third, in some successful cases, it has allowed escape from idiosyncrasies and has generated some useful generalisations.

"*No one can be a good observer unless he is a good theorizer*" (Charles Darwin, quoted from Zadoks, 1972). There are very few plant populations that are exposed to one disease only. Multiple pathosystems are where initial observations are made, where incipient hypotheses are borne, where these are tested, and sometimes result in success in managing diseases. This paper is not intended as a review of so vast a topic. Its aim is only to highlight some important points, research issues, and research avenues. Our purpose is to touch upon a limited number, but very different, aspects of the subject. References therefore are used only to illustrate research themes and approaches, and are given with no intention of offering a comprehensive overview.

Multiple pathosystems as facts

Multiple pathosystems consist of a series of disease elements that are present in the same host stand. Over time, e.g., during the cycle of a field crop, a number of diseases may appear, spread, decline, and interact among themselves and the growing crop. Figure 1 shows a series of principal component analyses on three very different multiple pathosystems. Principal component analysis is used here as one convenient means to provide a preliminary overview of very complex structures. For instance, the multiple pathosystem of groundnut in Côte d'Ivoire involves a series of fungal pathogens (Savary, 1987a) affecting the foliage (*Cercospora arachidicola*, *Cercosporidium personatum*, *Puccinia arachidis*), shoots, and stems (*Corticium rolfsii*, *Aspergillus niger*), and pods (*Botryodiplodia* sp.). Another principal component analysis highlights a series of wheat diseases in Brittany: eyespot, brown rust, septoria blotch and yellow rust. A third example illustrates an analysis on upland rice injuries in northern Laos (IRRI, 1998), which involves an array of injuries by insects (stem borers causing dead hearts and white heads, root injuries by white grubs), foliage injuries (caused by several species), disease injuries caused by fungi (neck and leaf blast caused by *Magnaporthe grisea*, brown spot caused by *Cochliobolus miyabeanus*, sheath blight caused by *Rhizoctonia solani*, sheath rot caused by *Sarocladium oryzae*), and weed infestation by a number of species. Not

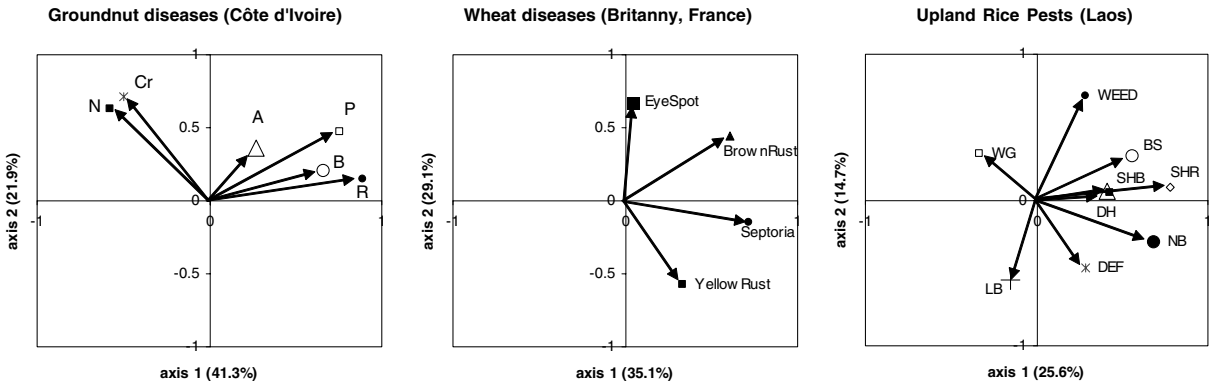


Figure 1. Three multiple pathosystems portrayed by principal component analyses: groundnut diseases in savanna and forest environments in Côte d'Ivoire, wheat diseases in Brittany, France, and upland rice in northern Laos. Left: Principal components analysis on 209 farmers' fields in several provinces (forest and savanna) of Côte d'Ivoire, 1982–1985; N: *Aspergillus niger* rot; Cr: *Corticium rolfsii* rot; A: *Cercospora arachidicola* leaf-spot; P: *Cercosporidium personatum* (*Phaeoisariopsis personata*) leaf-spot; B: *Botryodiplodia* pod rot; R: rust; Centre: wheat diseases in Britanny in a series of variety trials at varying levels of inputs, 2000, 2001, and 2003. Vectors indicate intensities of eyespot (*Tapesia yallundae*), of brown rust (*Puccinia recondita*), septoria (*Mycosphaerella graminicola*), and yellow rust (*Puccinia striiformis*). Data are from 180 individual plots (2.6×15 m) combining four crop management practices with five wheat cultivars in the replications over the 3 years. Right: upland rice pests in northern Laos, 1996 and 1997; LB: leaf blast; DEF: defoliating insects; NB: neck blast; DH: dead heart caused by stemborers; SHB: sheath blight, SHR: sheath rot; BS: brown spot; WEED: weed infestation; WG: white grub injury. Proportion of variances accounted for are indicated along each axis.

all the disease elements that actually were present in each of the three examples are represented. The three data sets used here correspond to different contexts for data acquisition (see legend of Figure 1). The groundnut data were collected during a multiyear survey in several provinces of Côte d'Ivoire, on 309 different farmers' fields (Savary, 1987a). The wheat data correspond to a series of varietal trials at different levels of inputs conducted over three different climatic years in Brittany (Rolland et al., 2003). The upland rice data were collected in a series of on-farm field experiments, where different fertiliser regimes were tested during two successive years (IRRI, 1998; Roder and Savary, unpublished data).

Some of the analyses involve more disease elements than others, and one of them involves more than just plant pathogens. Deriving conclusions on the overall importance of diseases in each pathosystem from the mere number of elements would of course be incorrect. These analyses only provide a view of possible associations, suggesting relationships, or absence of relationships, among disease elements. The relationships that seem to emerge from these summary analyses develop against the background of a large number of factors, including crop development stage, or crop management. For instance, a linkage appears in the groundnut multiple pathosystem between N (*A. niger*) and Cr (*C. rolfsii*); by contrast, there seems to be independence between Cr and A (*Cercospora arachidicola*) in the groundnut pathosystem, and independence between NB and LB (neck and leaf blast, respectively, both caused by *Magnaporthe grisea*) in the upland rice pathosystem. Collinearity or non-collinearity of (disease) vectors may lead to forwarding hypotheses, which in turn would require additional analyses.

Shapes of botanical pathosystems

Elements of multiple pathosystems

Multiple pathosystems have shapes, where individual diseases display a particular role. Several studies have shown that multiple pathosystems vary in shapes. Only two of the many reasons for change are illustrated here.

Change in age, i.e., development of the host stand, is one strong reason for change in the shape of multiple pathosystems. Figure 2 shows three separate principal component analyses on the groundnut–leaf-spot–rust pathosystem at three different ages of the groundnut stands. Although the analyses pertain to the same farmers' fields, comparison of analysis of Figure 2a (young stands), 2b (middle-age stands), and 2c (stands approaching or at harvest stage) indicate strong variation in relationships among variables. In young groundnut stands, a strong relationship between rust (R) and early leaf-spot (A) is apparent, both diseases being opposed to *A. niger* (N) wilt. In middle-age fields, a very strong association between rust (R) and late leaf-spot (P) is indicated, both diseases being opposed to early leaf-spot (A). In older fields, the relationship between rust and late leaf-spot has become loose, although both

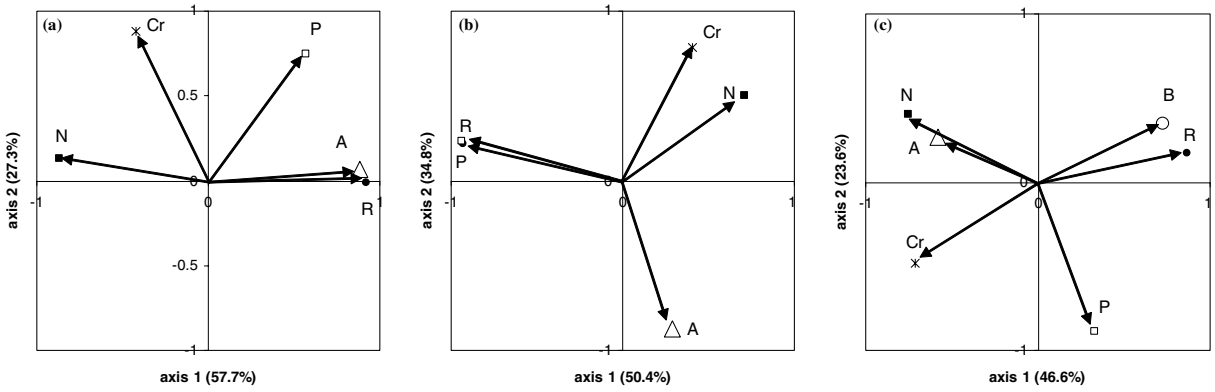


Figure 2. Patterns of change in multiple pathosystems over time and host development: principal component analyses of disease levels in farmers' fields in Côte d'Ivoire. (a) Groundnut field at early development stages (first trifoliate leaf – flowering); (b) groundnut field at medium development stages (flowering – pod filling); (c) groundnut field at final development stages (pod filling – harvest stage). Symbols for disease vectors are the same as in Figure 1.

remain opposed to early leaf-spot; and a linkage between *A. niger* and early leaf-spot is detected, which did not exist at earlier ages. These shifts in relationships are reflections of the respective dynamics of foliar diseases (rust and early leaf-spot usually establish earlier in a crop stand than late leaf-spot, while early leaf-spot generally declines as crop maturity approaches), and of environmental factors that favour certain diseases at certain stages of the crop (humid environment and contaminated seeds favour *A. niger* wilt and *C. rolfsii* in the early crop stages; dense, green canopies favour rust and late leaf-spot, whereas water stresses and poor soil fertility favour late leaf-spot in established stands; and more humid environments favour *Botryodiplodia* pod rot, late leaf-spot, and rust while drier environments favour *A. niger* wilt and early leaf-spot in older stands). One important factor that drives relationships among foliar diseases is disease-induced-defoliation. It will be addressed later in this discussion.

Crop management is another major reason for changes in shapes of multiple pathosystems. Figure 3 is an illustration of the effects of four crop management patterns in wheat experiments. Strong shifts in disease vectors are detected in Figure 3a–d, the transition from a to d corresponding to intensified wheat production. While in Figure 3a all diseases, except yellow rust, appear closely associated, an opposition between eyespot and both septoria leaf blotch and brown rust develops in Figure 3b, which persists in Figure 3c, but disappears in Figure 3d. Such sharp changes must be attributed to changes in fertiliser inputs,

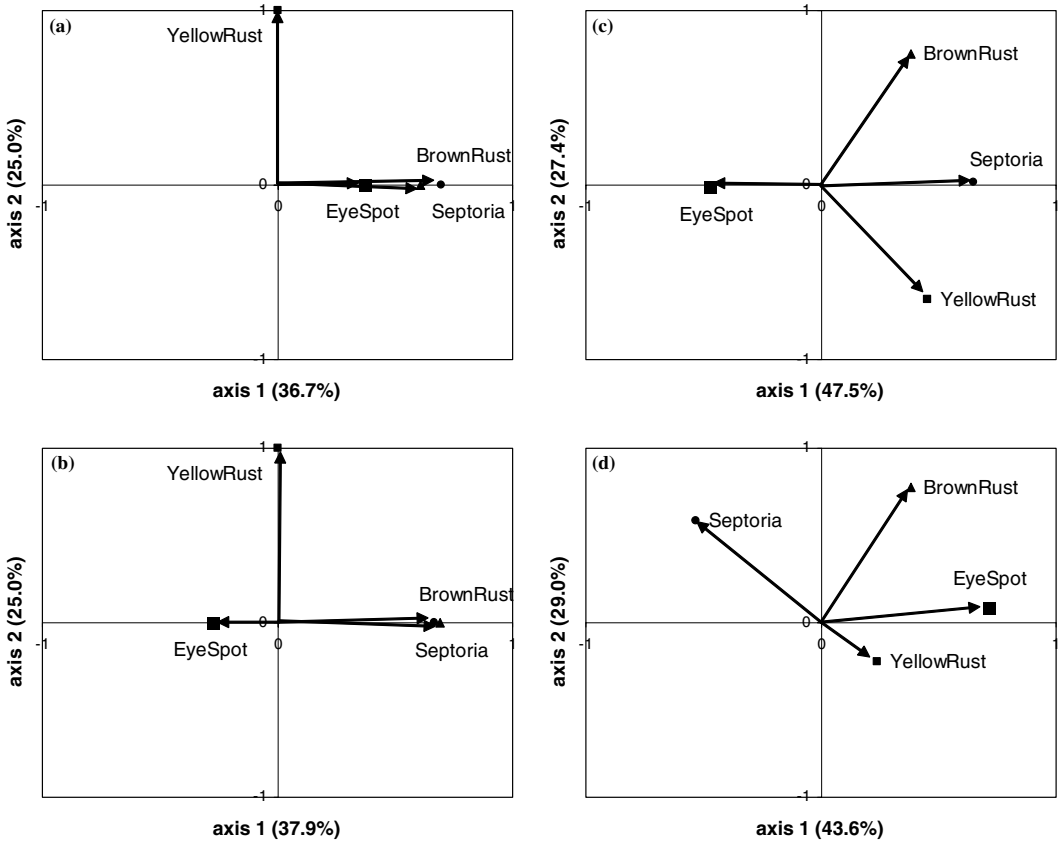


Figure 3. Patterns of change in multiple pathosystems over crop management practices: principal component analyses of disease levels in field trials in Britanny, France. (a) pattern of crop management A: fertiliser input with a target yield of 10 t ha^{-1}; seeding rate: 250 seeds m^{-2}; use of a crop growth regulator; three fungicide applications; (b) pattern of crop management B: fertiliser input with a target yield of 9 t ha^{-1}; seeding rate: 250 seeds m^{-2}; use of a crop growth regulator; two fungicide applications; (c) pattern of crop management C: fertiliser input with a target yield of 8 t ha^{-1}; seeding rate: 150 seeds m^{-2}; no growth regulator; one fungicide applications; (d): pattern of crop management D: fertiliser input with a target yield of 7 t ha^{-1}; seeding rate: 150 seeds m^{-2}; no crop growth regulator; no fungicide applications.

seeding rates, plant hormone use, and of course fungicide use. Crop management, involving pesticide use or not, has been found to be a major factor for changes in shapes of multiple pathosystems, in many diverse examples, including, e.g., wheat in Australia (Stynes, 1980) and the Netherlands (Daamen et al., 1989), lowland rice in Asia (Savary et al., 2000a), pea in Idaho (Wiese, 1982), groundnut in Côte d'Ivoire (Savary, 1987a), or coffee in Honduras (Avelino, 1999).

Shapes of multiple pathosystems in space

The spatial distributions of four different pathogens in the same crop stand are shown in Figure 4 (Lannou and Savary, 1991): rust (*Puccinia arachidis*), early leaf-spot (*Cercospora arachidicola*), late leaf-spot (*Cercosporidium personatum*), and web blight (*Rhizoctonia solani*) of groundnut. Several techniques, including geostatistical and multivariate, were used to show, as the maps strongly suggest, that (i) rust (Figure 2a) and web blight (Figure 2d) are spatially strongly associated, (ii) early leaf-spot (Figure 2b) is more severe where rust is less severe, and (iii) late leaf-spot (Figure 2c) does not intensify strongly where rust or early leaf-spot severities are extreme. The maps, which were drawn at the end of a cropping season, also show that two very different types of epidemics developed in the same stand, a typically focal epidemic (Figure 2d, web blight), and three general epidemics (Figures 2a, 2b and 2c, rust, early leaf-spot, and late leaf-spot, respectively) which did intensify locally.

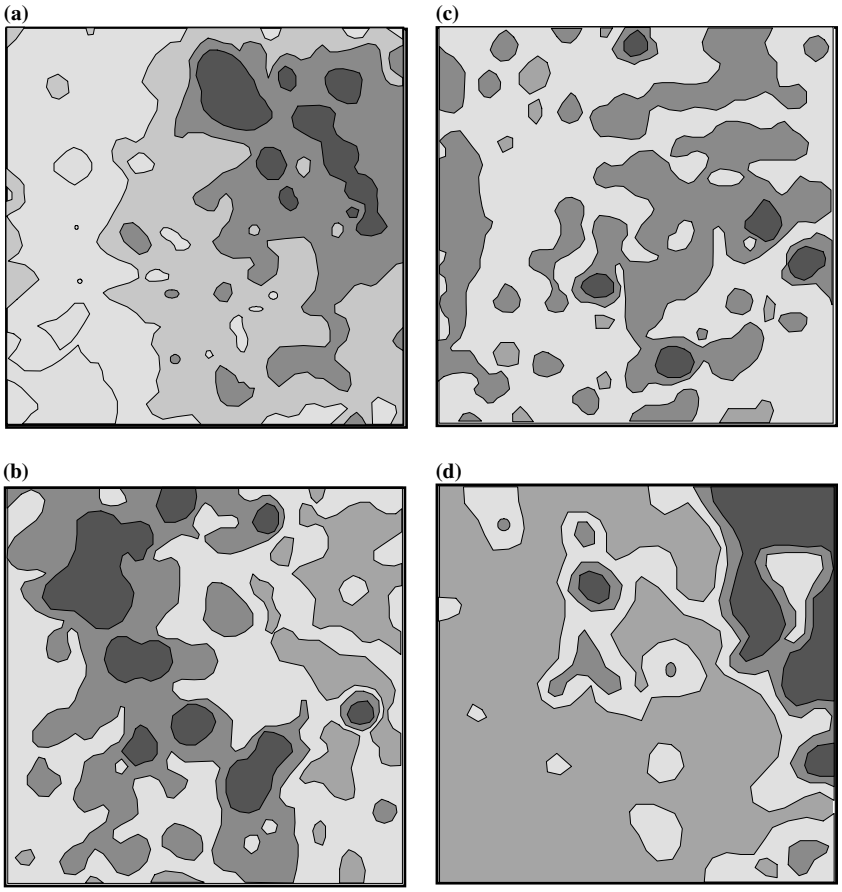


Figure 4. Patterns of change in multiple pathosystems over space: spatial distribution of four diseases a groundnut plot, Côte d'Ivoire (Lannou and Savary, 1991, modified). (a) Groundnut rust, *Puccinia arachidis*; (b) Early leaf-spot, *Cercospora arachidicola*; (c) Late leaf-spot, *Cercosporidium personatum* (*Phaeoisariopsis personata*), (d) Web blight, *Rhizoctonia solani*. Disease assessments were made at 90 days after sowing. Rust, early leaf-spot, and late leaf-spot: severity (% diseased leaf area) scales; web blight: incidence (% diseased plants) scale. From Lannou and Savary, 1991, modified.

Analysis of spatial patterns in multiple pathosystems may lead to a number of hypotheses, as in the case of multiple infection of hop stands by different viruses (Pethybridge and Turechek, 2003), which can lead to experimentally testing spatial co-occurrence and variation of infections in the host–vector–viruses system. In the case of groundnut diseases, much of the spatial co-variation of disease intensities may be attributed to competition towards vacant sites and defoliation of diseased tissues by some of the pathogens.

Nelson and Campbell (1993) studied a far more complex multiple pathosystem, which involves eight fungal pathogens of white clover (*Rhizoctonia solani*, *Pseudomonas andropogonis*, *Stagonospora meliloti*, *Cercospora zebrina*, *Curvularia trifolii*, *Colletotrichum trifolii*, *Polythrincium trifolii*, and *Uromyces* sp.), in presence or absence of three virus diseases (alfalfa mosaic virus, yellow vein virus, and peanut stunt virus). They detected disease aggregation of the fungal disease complex at several scales (leaf, plant, population), which fluctuated over time as foliation and defoliation occurred, and which varied spatially as well. Changes over time in the spatial aggregation of the foliar disease complex were associated with changes in disease severity itself and defoliation. The background of varying, multiple virus infection did not seem to affect either the dynamics of the foliar disease complex, or its spatial pattern.

Shapes of multiple pathosystems over time: dynamics of multiple pathosystems

One approach to addressing the dynamics of multiple pathosystems is by means of linked differential equations. The equations themselves can assume a number of shapes, but are in many respects fairly similar. One important difference however among seemingly analogous systems of equations is the nature of the modelled diseases variables, which often are proportions (severities or incidences, i.e., disease densities) or less frequently (as in the model used below), amounts of diseased tissues. This difference has important consequences on the meaning of parameters used. Use of linked differential equations derives in botanical epidemiology from the approach Van der Plank (1963) introduced to the field, which itself is related to earlier ecological models, including especially the Verhulst-Pearl logistic equation and systems of equations of the Lotka-Volterra type (Pianker, 1983, cited in Madden et al., 1987). Use of this approach has been extensive in botanical epidemiology, and has been based on both the development of simple equations that adequately describe disease progress over time (see, e.g., Madden, 1980; Campbell and Madden, 1990), and on tools to numerically integrate sets of differential equations that constitute simulation models (Zadoks, 1971; Zadoks and Rabbinge, 1985).

The groundnut–leaf-spot–rust pathosystem represents a good example to illustrate the approach. This system is fairly simple, having only two disease components, rust and leaf-spot, but leads to considering several processes and interactions: (i) a biotrophic pathogen which multiplies only on healthy, green tissues, (ii) a necrotrophic pathogen (only *Cercosporidium personatum* is considered here, but both *C. personatum* and *Cercospora arachidola* could be considered collectively) which causes extensive defoliation, which (iii) compounds physiological (senescence) defoliation, (iv) competition between the two pathogens in their access to growing crop tissues, and (v) the ability of one of the two diseases (leaf-spot) to multiply from defoliated, infectious tissues (Savary and Servat, 1991). Such characteristics are very similar to another legume-based multiple pathosystem, the bean–angular leaf-spot–anthracnose–rust system (Gomes Carneiro et al., 2000; Bassanezi et al., 2001; de Jesus et al., 2001), which has been extensively studied. A series of linked differential equations representing the groundnut–leaf-spot–rust pathosystem are given in Table 1, corresponding to the overall model structure of Figure 5. Parameters for the model (relative rates of crop growth, of increase of both diseases, of physiological defoliation, and of disease-induced defoliation) were estimated (Savary and Servat, 1991) using a numerical integration procedure coupled with a sequence of two optimisation procedures (Rosenbrock, 1960; Nelder and Mead, 1964), applied to a set of 15 epidemics where levels of the both diseases were artificially manipulated (Savary and Zadoks, 1992a).

Simulated outputs using optimised parameters are shown in Figure 6a. A regular, logistic-shaped increase of leaf-spot is combined with a faster increase of rust, which declines in the later stage of the epidemic; these are coupled with a regular

Table 1. Equations used in a rust–leaf-spot–groundnut multiple pathosystem simulation model

Equation	Hypotheses
(1) Rate of increase of leaf area: $dL/dt = RRL * L\{1 - [(L + Y + Z + totD)/L_{max}]\}$	The rate of increase of leaf area is proportional to a relative rate and the amount of (healthy) leaf, corrected for the fraction of leaf growth, relative to a maximum. Leaf growth includes defoliated tissues
(2) Rate of rust increase: $DY/dt = RRY * Y\{1 - [(Y + Z)/(L + Y + Z)]\}$	The rate of rust increase is proportional to a relative rate and the amount of rust-diseased tissues, corrected for the fraction of relative growth, relative to the current total leaf tissues
(3) Rate of leaf-spot increase: $DZ/dt = (RRZ * Z + RRZDZ * DZ) * \{1 - [(Z + Y)/(L + Y + Z)]\}$	The rate of leaf-spot increase is proportional to both (1) a relative rate for standing diseased tissues and the amount of leaf-spot-diseased tissues and (2) a relative rate for defoliated diseased tissues and the amount of defoliated infected tissues, corrected for fraction of relative growth, relative to the current total leaf tissues
(4) Rate of defoliation (healthy tissues) $dD/dt = RRDS * L + RRDZ * [Z/(L + Y + Z)] * L$	The rate of defoliation of healthy tissues is the sum of senescence-induced (relative rate and healthy tissues) and indirectly leaf-spot-induced (relative rate, proportion leaf-spot-diseased, and healthy tissues)
(5) Rate of defoliation (leaf-spot-diseased tissues) $dDZ/dt = (RRDS + RRDZ) * Z$	The rate of defoliation of leaf-spot-diseased tissues is proportional to a relative rate (senescence and leaf-spot accumulated) and to the amount of leaf-spot-diseased tissues
(6) Rate of defoliation (rust-diseased issues) $dDY/dt = (RRDS + RRDZ) * Y$	The rate of defoliation of rust-diseased tissues is proportional to a relative rate (senescence and leaf-spot accumulated) and to the amount of rust-diseased tissues

Variables (dimensions in brackets): L: healthy leaf tissues [L^2]; Y: rust-diseased tissues [L^2] ; Z: leaf-spot-diseased tissues [L^2]; D: healthy, defoliated tissues [L^2]; DY: rust-diseased, defoliated tissues [L^2]; DZ: leaf-spot-diseased, defoliated tissues [L^2]; RL: rate of leaf growth [$L^2\ T^{-1}$]; RY: rate of rust increase [$L^2\ T^{-1}$]; RZ: rate of leaf-spot increase [$L^2\ T^{-1}$]; RD: rate of defoliation of healthy tissues [$L^2\ T^{-1}$]; RDY: rate of defoliation of rust-diseased tissues [$L^2\ T^{-1}$]; RDZ: rate of defoliation of leaf-spot diseased tissues [$L^2\ T^{-1}$]; RRL: relative (intrinsic) rate of leaf growth [T^{-1}]; L_{max}: maximum leaf growth [L^2]; RRY: relative rate of rust increase [T^{-1}]; RRZZ: relative rate of leaf-spot increase from non-defoliated tissues [T^{-1}]; RRZDZ: relative rate of leaf-spot increase from infected defoliated tissues [T^{-1}]; RRDS: relative rate of defoliation (senescence) of healthy tissues [T^{-1}]; RRDZ: relative rate of defoliation of (leaf-spot) diseased tissues [T^{-1}]; totD: accumulated defoliation [L^2].

accumulation of defoliated tissues, and a bell-shaped green leaf area curve. When the relative rate of rust increase is reduced (Figure 6b) or increased (Figure 6c), a strong rust reduction, or increase, respectively, is simulated, coinciding with opposite behaviour of the leaf-spot epidemic. When the relative rates of leaf-spot increase from both non-defoliated and defoliated tissues are reduced (Figure 6d), the leaf-spot epidemic is strongly reduced, while the rust epidemic is strongly increased. When both these parameters are increased (Figure 6e), the leaf-spot epidemic is strongly increased, the rust epidemic is nearly halved, and the amount of defoliation is increased. A last set of tests, where the relative rate of defoliation induced by leaf-spot is varied is also shown. When the leaf-spot induced-defoliation rate is decreased (Figure 6f), the amount of total defoliation is barely altered, the leaf-spot epidemic remains essentially unaffected, whereas the rust epidemic is slightly enhanced. Increase of this parameter (Figure 6g), on the other hand, leads to only a slight increase in total defoliation, an unaffected leaf-spot epidemic, but a reduction by about 8% of the rust epidemic.

The overall picture this model gives is that the system is quite sensitive to comparatively small variation (10% or less) in parameter values. Leaf-spot is behaving as a very strong competitor for rust, especially through leaf-spot induced-defoliation. Rust on the other hand can rapidly take advantage of seemingly small increases in the amount of available healthy tissues. These interactions are strongly influenced by crop growth (and the effects of diseases on crop growth), and defoliation (especially through physiological senescence). One interesting response of the system is that a reduced retention (greater disease-induced

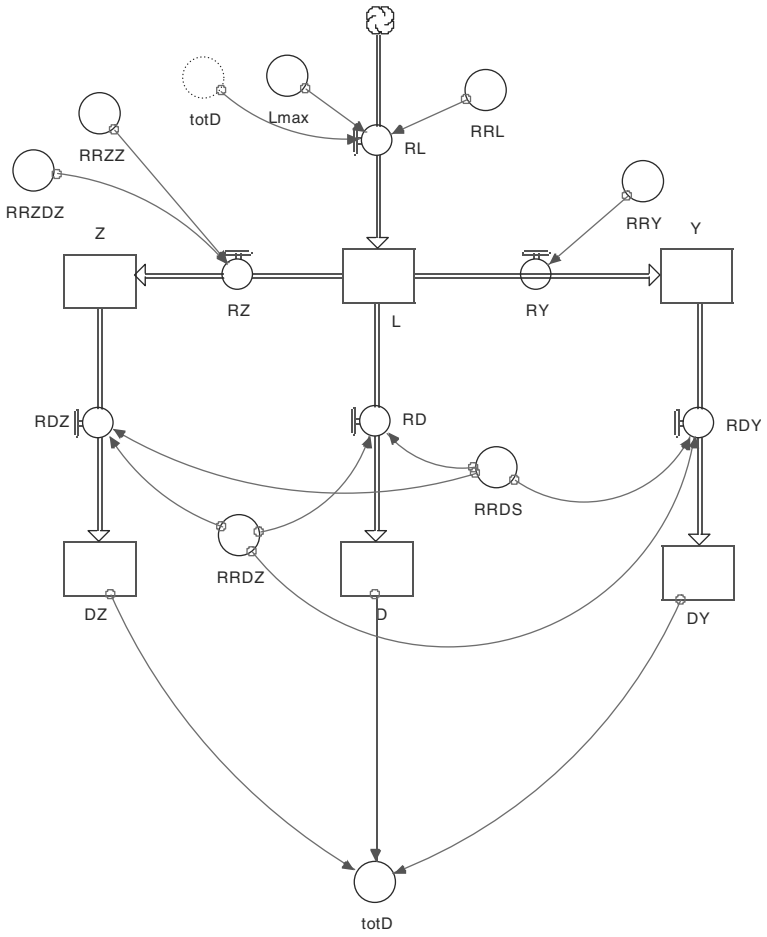


Figure 5. Overall structure of a mechanistic simulation model incorporating two foliar diseases, rust and leaf-spot on groundnut. State variables: L: healthy leaf tissues [L^2]; Y: rust-diseased tissues [L^2]; Z: leaf-spot-diseased tissues [L^2]; D: healthy, defoliated tissues [L^2]; DY: rust-diseased, defoliated tissues [L^2]; DZ: leaf-spot-diseased, defoliated tissues [L^2]; Rates: RL RL: rate of leaf growth [$L^2\ T^{-1}$]; RY: rate of rust increase [$L^2\ T^{-1}$]; RZ: rate of leaf-spot increase [$L^2\ T^{-1}$]; RD: rate of defoliation of healthy tissues [$L^2\ T^{-1}$]; RDY: rate of defoliation of rust-diseased tissues [$L^2\ T^{-1}$]; RDZ: rate of defoliation of leaf-spot diseased tissues [$L^2\ T^{-1}$]; Parameters: RRL: relative (intrinsic) rate of leaf growth [T^{-1}]; L_{max}: maximum leaf growth [L^2]; RRY: relative rate of rust increase [T^{-1}]; RRZZ: relative rate of leaf-spot increase from non-defoliated tissues [T^{-1}]; RRZDZ: relative rate of leaf-spot increase from infected defoliated tissues [T^{-1}]; RRDS: relative rate of defoliation (senescence) of healthy tissues [T^{-1}]; RRDZ: relative rate of defoliation of (leaf-spot) diseased tissues [T^{-1}]; totD: accumulated defoliation [L^2].

defoliation) of leaf-spot infected tissues (Figure 6a vs. 6g) leads to a reduced rust epidemic. Selecting varieties that shed their leaves at low leaf-spot severity might then be an efficient way of reducing rust epidemics.

The white clover–foliar fungal diseases–viruses system (Nelson and Campbell, 1993) renders the groundnut–rust–leaf-spot system a comparatively simple system to address. Nelson and Campbell question the relevance of the approach illustrated above, which in the case of the clover-based system would require a set of at least 10 equations. If this approach were to be taken, a complex model structure would have to be designed, a large number of parameters would have to be estimated, and numerical solutions would become difficult to interpret. The approach chosen by Nelson and Campbell (1993) in their field work, however, was not to consider each disease separately, but to quantify the leaf disease complex as a whole. This leads to the interesting avenue of perhaps considering groups of pathogens that share similar

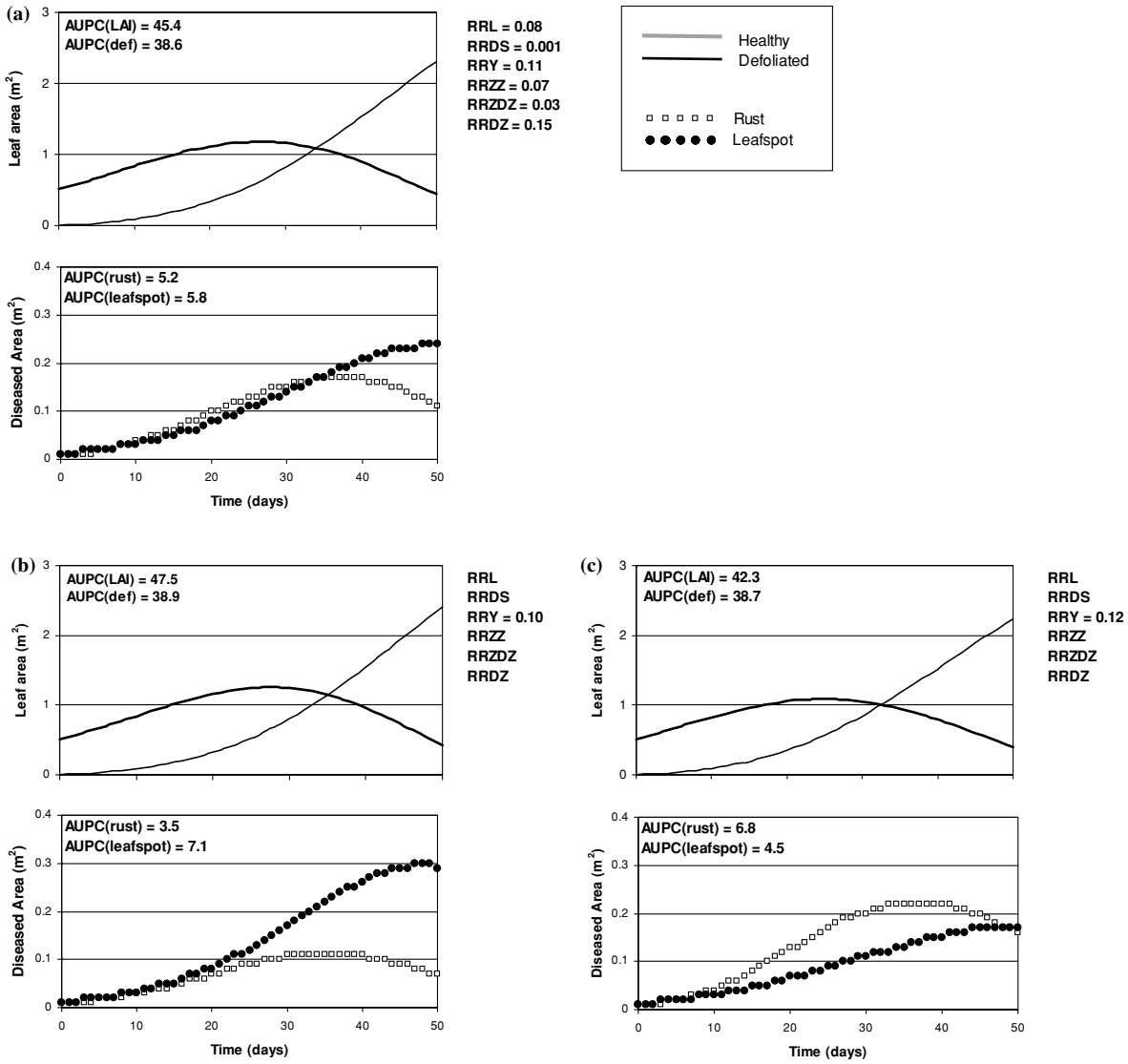


Figure 6. Two foliar diseases dynamically interacting, simulated rust and leaf-spot epidemics on groundnut. Each simulation is represented by two graphs of outputs, healthy and (total) defoliated area (upper half) and rust and leaf-spot severities (lower half). Numerical values of parameters are indicated in graphs b–g only when changes from reference (optimised) values (graph a) have been used. Calculated areas under progress curves (AUPC) of the rust epidemic, of the leaf-spot epidemic, of the leaf area index, and of the defoliated leaf area index are indicated. (a) simulated outputs with optimised parameter values for RRL (relative rate of leaf growth; [T^{-1}]), RRDS (relative rate of senescence defoliation of healthy tissues [T^{-1}]), RRY (relative rate of rust increase [T^{-1}]), RRZZ (relative rate of leaf-spot increase from non-defoliated tissues [T^{-1}]), RRZDZ (relative rate of leaf-spot increase from infected defoliated tissues [T^{-1}]), and RRDZ (relative rate of defoliation of leaf-spot diseased tissues [T^{-1}]). (b) simulated outputs for a reduced relative rate of rust increase (RRY). (c) simulated outputs for an increased relative rate of rust increase (RRY). (d) simulated outputs for reduced relative rates of leaf-spot increase (RRZZ and RRZDZ). (e) simulated outputs for increased relative rates of leaf-spot increase (RRZZ and RRZDZ). (f) simulated outputs for a reduced relative rate of defoliation induced by leaf-spot (RRDZ). (g) simulated outputs for an increased relative rate of defoliation induced by leaf-spot (RRDZ).

functional attributes in a community, and model the dynamics of guilds of pathogens, rather than of individual diseases.

One of the many criticisms of the linked differential equation approach (see, e.g., Nelson and Campbell, 1993) is its inability to account for

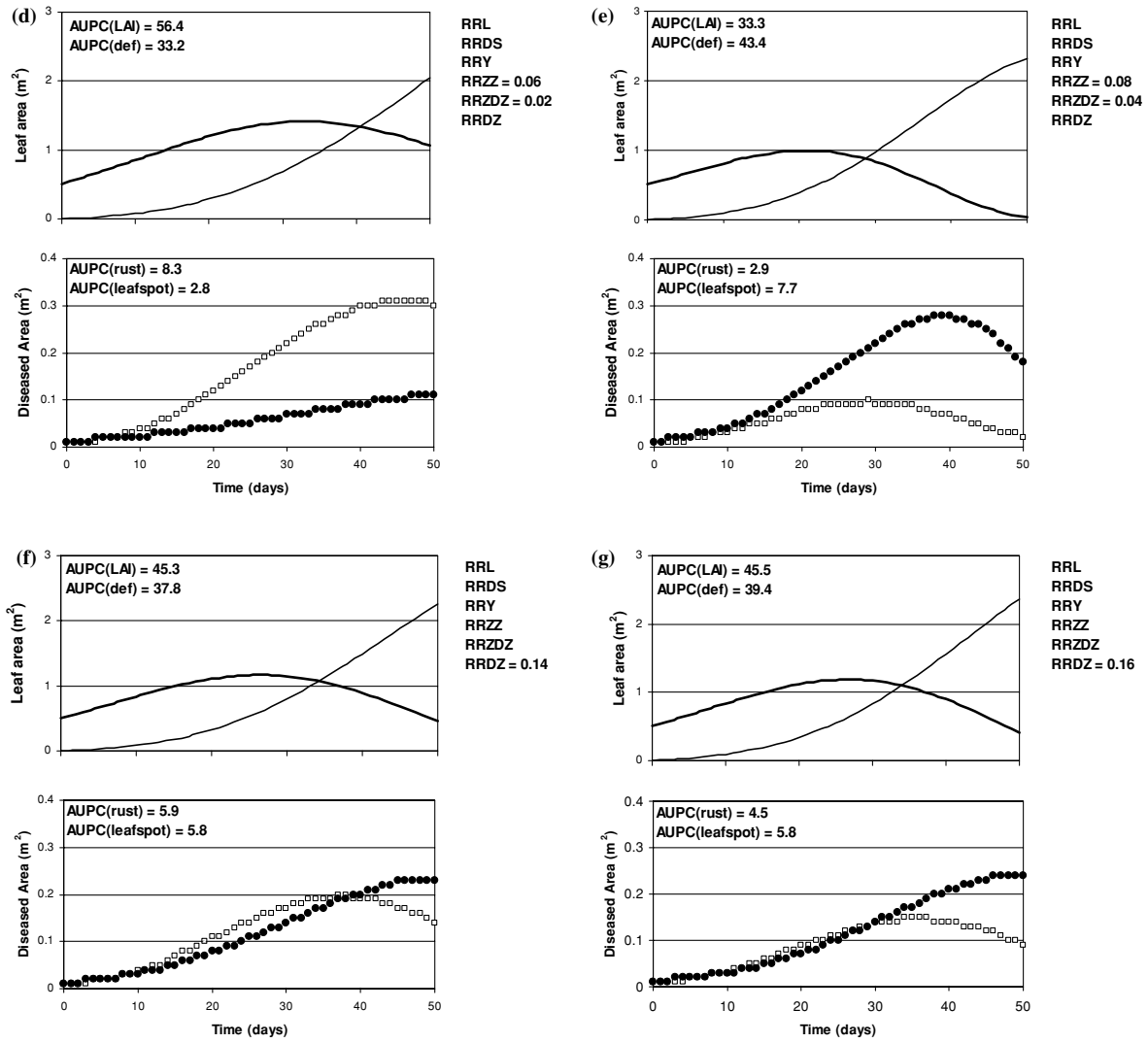


Figure 6. Continued

spatial patterns, and the essential effects of spatial patterns on disease dynamics. Figure 4 is sufficient evidence of the fact that the groundnut–leaf-spot–rust system is no exception. The model summarised here does however include implicit assumptions pertaining to the distribution of disease on host tissues. As in many pathosystems involving foliar diseases, a very strong vertical aggregation occurs in both the leaf-spot (Boote et al., 1980) and the rust (Savary, 1987b) pathosystems. Such aggregation of diseases (and defoliation) along the vertical dimension of a crop canopy must have very strong consequences on the behaviour of the multiple pathosystem. Considering the model outlined in equations of Table 1 and the flowchart of Figure 5, however, amounts to implicitly considering a growing canopy with two layers: (i) a (healthy) layer where defoliation is caused by senescence and distance effect of leaf-spot disease, and (ii) a (diseased) layer, where both rust and leaf-spot lesions occur, and where defoliation is caused by both leaf-spot and senescence effects.

The purpose of this type of model is to explore interactions within a framework of thinking defined, and limited by, a set of hypotheses. Expanding the model to address additional, albeit important, features of the considered system might prevent the interpretation of simulation results,

evaluation of the model, and its use. Further discussion on strategic models of this kind is given in McRoberts et al. (2003).

Multiple pathosystems and damage to crops

Multiple injuries and the resulting damage

Consideration of multiple pathosystem, i.e., pathogens dynamically interacting among themselves and a growing canopy, inevitably leads to questions regarding harmfulness of the multiple pathosystem. The extent of damage (*sensu* Zadoks, 1985, i.e., yield reduction) caused by a multiple pathosystem has scientific relevance of its own; it also has very practical implications with respect to the availability, efficiency, and deployment of management options. Plant pathologists did not study multiple pathosystems because of the interesting interactions among competing pathogens that may take place, and chose them because they were good examples for community ecology studies; they did not either address the issue with the prime objective of selecting plants that would be resistant against several diseases, or crops whose yields would be stable in the presence of several diseases. One simple, practical reason was the need to assess the damage caused by several injuries. Padwick (1956) made an early attempt to such quantification, with the formula:

$$\begin{aligned}&\text{Percent yield loss}\\&= \{100\\&\quad -[(100-P1)(100-P2)\\&\quad \times(100-Pn)]/100n-1\}\end{aligned}$$

where Pi is the percent yield loss caused by an individual injury i. Padwick's model assumes that the only interaction between diseases on yield is competition for the crop's resources. This amounts to forwarding the hypothesis that one disease cannot affect what other diseases have already injured (Johnson et al., 1986). Padwick's view strongly contrasted with several later analyses, whereby individual losses were merely accumulated in a 'loss profile' (see, e.g., Pinstrup-Andersen et al., 1976). Teng (1994) pointed to the fact that this latter reasoning may lead to the impossible result that diseases, or crop harmful agents in general, may cause losses exceeding 100%.

Quantification, analysis, and modelling damage is both the scientific-technical cornerstone for disease management (James, 1974; Chiarappa, 1980; Madden, 1983; Teng, 1983; Zadoks, 1985; Teng, 1987; Gaunt, 1995) and one of the very important entry points for disease management (see, e.g., Teng and Savary, 1992). This has been a very broad and active field of research for many decades. The subject has particular relevance when considering multiple pathosystems (Madden and Nutter, 1995), however, and a few points are discussed here.

Five directions

As opposed to an additive model for combined effects of injuries on damage, the model developed by Padwick was a useful starting point, from which several directions of thoughts were explored.

(i) A first direction concerns the nature of disease interactions in their yield-reducing effect: very often, a less than additive effect is observed, but some injuries may synergistically increase yield losses. Less than additivity has been demonstrated in one of the best documented studies on damage caused by a multiple pathosystem, the early blight–verticillium wilt–potato leafhopper of potato (Johnson, 1986; Johnson et al., 1986, 1987; Johnson and Teng, 1990). Injuries caused by diseases may, however, synergistically reduce yield. Synergies in yield-reducing effects are found in potato early dying caused by *Verticillium dahliae* and *Pratylenchus penetrans* (Francl et al., 1990), as well as in combinations of infections of wheat by *Septoria nodorum* and *Puccinia recondita* (Van der Wal et al., 1970). A first hypothesis therefore refers to the direction, positive or negative, of combined injuries on damage.

(ii) A second direction concerns the nature of competition, which may only be for 'resource', as Padwick's model refers, or may involve other mechanisms, resulting then in damage lower than expected from Padwick's model. Such is the case in the potato–early blight–verticillium wilt–leafhopper studied by Johnson (1986) and Johnson et al. (1986).

(iii) A third direction concerns the way damage measurements are expressed. Damage is commonly reported as percentage. The reported figures (percent losses) are therefore dependent on

the level of the uninjured yield reference, and their meaning will strongly depend on whether this yield reference is low or high. One of many alternatives to reporting damage as a proportion is expressing it as a biomass, and the choice will depend on the end-use of the information.

(iv) A fourth direction concerns the relationships that link damage to the various injuries (the damage function; Zadoks, 1985). The damage function may involve the yield pertaining to a given production situation as an explanatory variable (the attainable yield; Rabbinge and De Wit, 1989). As is the case with the damage caused by single harmful agents (see, e.g., Rossing, 1991, for the grain aphid on wheat, and Rabbinge et al., 1985, for powdery mildew in wheat), the amount of damage caused by a multiple pathosystem may depend on the level of attainable yield. This has been exemplified in the case of the multiple pathosystem of groundnut in Côte d'Ivoire (Savary and Zadoks, 1992b), and in the more complex and diverse multiple pathosystem of lowland rice in Asia (Savary et al., 2000b).

(v) A fifth direction of thought follows the realisation that diseases, and any harmful agent in general, belong to one or a few categories, based on the type of injury mechanisms they trigger. Rabbinge and Rijsdijk (1981) and Boote et al. (1983) defined the limited number of ways for a harmful organism to hamper the physiological performances of a growing canopy. This has particular relevance when considering multiple pathosystems for two reasons. First, it provides a basis for designing experiments, developing field survey procedures, and defining field measurements that refer not to specific diseases (or harmful agents) but to specific injuries. Diseases then are not measured with respect to how fast they intensify, but rather to how much they may affect the performances of a crop. Further, they need not necessarily be measured individually, but collectively, as Nelson and Campbell (1993) did. Second, this categorization provides a framework for modelling mechanistically the physiological interactions between a crop stand and a multiple pathosystem. This direction of thought has been underpinning research involving simulation models as tools for understanding damage caused by diseases and means to reduce them (Rouse, 1988; Rabbinge et al., 1989; Gaunt, 1995), including work conducted on multiple pathosystems such as potato–early blight–verticillium wilt (Johnson, 1986; Johnson et al., 1986, 1987; Johnson and Teng, 1990).

New developments have taken place, where these five points are considered in the case of lowland rice in Asia (Pinnschmidt et al., 1995; Willocquet et al., 2000, 2002, 2004). Simulation models have been developed that make use of the concept of guilds of injuries (Pinnschmidt et al., 1995; Willocquet et al., 2000, 2002) which have been used to analyse and understand the yield-reducing effects of several pathogens, insects, and weeds in the same crop. A modelling structure has been designed so that it can simultaneously handle production situations (as drivers of attainable crop performances) and injury profiles (as drivers of multiple injuries) in the very combinations where field characterisation had shown these (production situation) $\times$ (injury profile) associations occur (Willocquet et al., 2000, 2002)). Production situations and their associated injury profiles were then used as the modelling context where disease and pest management tools could be most efficiently deployed, and where progress should be expected, and so expressed in yield gains, instead of yield losses (Willocquet et al., 2004).

Multiple pathosystems and integrated pest management

A negative view

In his article on the functioning and performances of tropical ecosystems, Janzen (1973) was expressing his frustration at science not achieving its goals in vital fields of application: "*Nearly all research in tropical agriculture is highly reductionist, parochial, and discipline-oriented*". At the time when Janzen wrote, much of the synthesis tools that now are available to plant pathologists did not yet exist. A negative view, similar to Janzen's, could be expressed considering the very slow pace of progress that has been achieved in understanding, analysing, and managing multiple pathosystems. In spite of the availability of tools to address it, the complexity of these systems remains a deterrent, not a challenge, to many plant pathologists. But the primary reason why progress has been so slow is the weakness of communication among disciplinary fields (McRoberts et al., 2003).

Questions about multiple pathosystems are central to defining the scientific bases for, the designing of, and the implementing of IPM. Reductionism, in its many forms, and discipline-oriented science, are the very same reasons that hamper progress in IPM (Jeger, 2000), and it would seem that Joni Mitchell's song ("*Hey farmer farmer. Put that DDT now. Give me spots on my apples. But leave me the birds and the bees. Please.*"; (McRoberts et al., 2003)) does not seem to be fading away anytime soon.

A positive view

There has nevertheless been a change in the way disease management has been addressed, scientifically and technically, over the past 50 years – the time-span covering the cycle of the International Epidemiology Workshops. One good example appears to be the multiple pathosystem of wheat in wheat-based systems of western Europe.

Initial steps were taken in the Netherlands with the EPIPRE project which saw scientists sharing experience with farmers, adapting theories to practice, and farmers empowered in their disease management decisions from epidemiological and systems science (Zadoks, 1989).

This early farmers-driven project had a setting different from that of today. Much work has been accomplished since, and already was on its way then, to show through long-term experiments (Jordan et al., 1985; Webster, 1985; McRoberts et al., 2000) that integrated production systems for field crops tend to perform better financially than high-input systems when commodity prices are low (McRoberts et al., 2003). The notion that disease management depends on production situations – not only because multiple pathosystems are so sensitive in their composition to crop management, but simply because disease management is only part of crop management, and so, necessarily depends on the socio-economic dimensions of what a production situation is – was no longer a theory but a concept put into practice and a way to conduct research.

The old, simple, idea that stable yield, and stable yield characteristics, including multiple, incomplete, host plant resistance are criteria for selecting varieties was revisited by G. Doussinault (Doussinault, 1998; Doussinault et al., 2001). Selecting for maximum yield and maximum grain protein content under intensive production conditions, i.e., high (nitrogen) fertiliser inputs and a pesticide umbrella leads to ignoring a large fraction of genetic resources available in a germplasm, and to restricting progress within unrealistically favourable production situations. Hardy wheat varieties, which yield-wise may perform somewhat below, or nearly as well as, conventional high yielding cultivars, but exhibit a number of incomplete resistances and have fair performances with respect to protein contents of the grain, are being tested in a number of sites in France. Incomplete resistances concern eyespot (*Tapesia yallundae*) and leaf blotch (*Mycosphaerella graminicola*). Experiments do not involve varieties only, but patterns of crop management; they also lead to assessment of economic performances. So far (over 3 years, 1999–2002), reported results are very encouraging (Rolland et al., 2003): under a low input regime, hardy wheat cultivars yields are reasonably stable (within 10 to 8 t ha^{-1}), have acceptable grain protein contents (10.5–11.5%), and are produced at costs reduced from 400 to 150 euros ha^{-1} compared to an intensive production system. In 46 (i.e., 73%) of the 63 tested combinations of (year × crop management pattern × cultivar), low-input crop management with hardy varieties generated the highest net returns. Thus in about 73% of the cases, host plant resistance to the wheat multiple pathosystem is mobilised as a tool for system stabilisation – instead of pesticides for system perturbation, and yield-driven control. Interestingly, the effort appears to be led by breeders; that economists, agronomists, and pathologists contribute to designing production systems that suit a seed-based technology is making an experience of this kind very promising, not only because of the science it involves, but because of the promise to deliver it entails.

Perspectives

Epidemiological guilds

The example of approach used above to model the dynamics of simultaneous disease epidemics epidemics invites the question of how to best address the temporal and spatial structure of multiple pathosystems whose disease components may be

numerous, or variable. Nelson and Campbell (1993) addressed a very diverse multiple pathosystem. Aside from rendering their study doable, their approach to quantify and analyse several foliar diseases collectively is an important choice, which carries the simplifying, and new, assumption that several diseases may be considered as an aggregate, from an epidemiological standpoint. One approach to analysing complex, multiple pathosystems may thus be grounded on the consideration of epidemiological guilds, rather than of individual pathogens. Definition of these guilds and clustering individual disease components in them could make use of results achieved from comparative epidemiological work (Kranz, 1974, 1980). Simplification of this kind, if successful, would have the merit of accommodating quantitative epidemiological knowledge on individual disease components, while also generating a common framework for understanding and management of multiple pathosystems that differ in their biological components, but share common epidemiologically functional traits.

Guilds of harmful agents

This concept has proven useful. The notion that harmful organisms share common injury mechanisms (Rabbinge and Rijsdijk, 1981; Boote et al., 1983) enables us to bring together organisms that otherwise profoundly differ. This has found several applications, in designing field quantification methods, field experiments to measure damage, and of course, simulation models. These in turn have found strategic applications for research prioritisation towards multiple pathosystems, for assessing the impact of new crop management and changes in production situations, and of disease management tools. It also has allowed a shift of thinking, from assessment of damage, to projection of yield gains (Willocquet et al., 2004), that is, a shift from what has been lost because of current practices, to what could be gained from future options.

Syndromes of production, syndromes of disease

The term syndrome has two definitions (Babcock Grove, 1961): (i) a group of symptoms or signs typical of a disease, disturbance, condition, or lesion in animals or plants, and (ii) a set of concurrent things. Andow and Hikada (1989) used the term to both describe different patterns of management of rice in Japan, and the plant health consequences this difference in production situations has on rice diseases. Physicians do not differ in their use of the word (e.g., Peto, 2001; Zimmet et al., 2001). The detection of linkages between production situations and injury profiles is analogous to considering corresponding syndromes of production and syndromes of diseases. This represents an avenue towards improving plant health, i.e., better management of multiple pathosystems, via improvements of management. Considering epidemiological guilds and guilds of harmful agents might be an interesting direction to take towards that aim.

References

Andow DA and Hikada K (1989) Experimental natural history of sustainable agriculture: Syndrome of production. Agriculture, Ecosystems and Environment 27: 447–462.

Avelino J (1999) Contribution à la gestion du pathosystème *Hemileia vastatrix* et *Coffea arabica* au Honduras. Influence des principaux facteurs physiques, phytotechniques et biologiques. PhD Dissertation, Paris XI Orsay Univ, 127 pp.

Babcock Grove P (ed.) (1961) Webster's Third New International Dictionary. Merriam-Webster Inc. Springfield, Massachusetts.

Bastiaans L and Daamen RA (1994) The role of crop husbandry practices in wheat production for disease control. Brighton Crop Protection Conference, Pests and Diseases – 1994. Vol. 3 (pp. 1169–1174).

Bassanezi RB, Amorim L, Bergamin Filho A, Hau B and Berger RD (2001) Accounting for photosynthetic efficiency of bean leaves with rust, angular leaf-spot and anthracnose to assess crop damage. Plant Pathology 50: 443–452.

Boote KS, Jones JW, Smerage GH, Barfield CS and Berger RD (1980) Photosynthesis of peanut canopies as affected by leaf-spot and artificial defoliation. Agronomy Journal 72: 247–252.

Boote KJ, Jones JW, Mishoe JW and Berger D (1983) Coupling pests to crop growth simulators to predict yield reductions. 73: 1581–1587.

Campbell LC and Madden LV (1990) Introduction to Plant Disease Epidemiology, John Wiley & Sons, New York, 532.

Chiarappa L (ed.) (1980) Crop Loss Assessment Methods. Supplement No. 3 FAO/Comm. Agric., Bureaux, Farnham, England, 123 pp.

Daamen RA, Wijnands FG and Van der Vliet G (1989) Epidemics of diseases and pests of winter wheat at different levels of agrochemical inputs. A study on the possibilities for designing an integrated crop system. Journal of Phytopathology 125: 305–319.

deJesus Junior WC, Vale FXR, Coelho RR, Zambolim L, Costa LC and Bergamin Filho A (2001) Effects of angular

leaf-spot and rust on yield loss of *Phaseolus vulgaris*. Phytopathology 91: 1045–1053.

Doussinault G (1998) La sélection pour les variétés adaptées à une agriculture durable. Dossiers de l'Environnement de l'INRA 16: 39–44.

Doussinault G, Judeau B and Jahier J (2001) Evolution de la variabilité génétique chez le blé. Dossiers de l'Environnement de l'INRA 21: 91–104.

Francl LJ, Madden LV, Rowe RC and Riedel RM (1990) Correlation of growing season environmental variables and the effect of early dying on potato yield. Phytopathology 80: 425–432.

Gaunt RE (1995) The relationship between plant disease severity and yield. Annual Review of Phytopathology 33: 119–144.

Gomes Carneiro SMTP, Amorim L, Bergamin Filho A, Hau B and Bianchini A (2000) Dinâmica de area foliar, desfolha e variaveis de area foliar sadia em feijoeiros com infecçoes isoladas e conjuntas de *Phaeoisariopsis griseola* e *Colletotrichum lindemuthianum*. Summa Phytopathologica 26: 406–412.

Hamelink J, Hau B and Kranz J (1988) Untersuchungen zur Entwicklung von Befalls/Verlust-Relationen bei Schadereger-Komplexen an Einzelpflanzen des Maises. Zeitschrift Fur Pflanzenkrankheiten Und Pflanzenschutz 95: 258–269.

IRRI (1998) Program Report for 1998. A Characterization of Upland Rice Biotic Constraints in Laos. International Rice Research Institute, Los Banos, Philippines.

James CW (1974) Assessment of plant diseases and losses. Annual Review of Phytopathology 12: 27–48.

Janzen DH (1973) Tropical agroecosystems. Science 182: 1212–1219.

Jeger MJ (2000) Bottlenecks in IPM. Crop Protection 19: 787–792.

Johnson KB (1986) Systems analysis and modeling of yield loss in potato caused by interacting populations of early blight, *Verticillium* wilt, and the potato leafhopper. PhD Dissertation, Univ. Minnesota 201 pp.

Johnson KB, Radcliffe EB and Teng PS (1986) Effects of interacting populations of *Alternaria solani*, *Verticillium dahliae*, and the potato leafhopper (*Empoasca fabae*) on potato yield. Phytopathology 76: 1046–1052.

Johnson KB, Teng PS and Radcliffe EB (1987) Analysis of potato foliage losses caused by interacting infestations of early blight, *Verticillium* wilt, and potato leafhopper; and the relationship to yield. Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz 94: 22–33.

Johnson KB and Teng PS (1990) Coupling a disease progress model for early blight to a model of potato growth. Phytopathology 80: 416–425.

Jörg E, Weissenfels D and Kranz J (1987) Krankheits- und Schadlingsbefall an Haupt- und Nebenhalmen des Weizens. Zeitschrift Fur Pflanzenkrankheiten Und Pflanzenschutz. 94: 509–519.

Jordan VWL, Hurcheon JA, Donaldson GV, Farmer DP (1985) Research into and development of integrated farming systems for less intensive arable crop production. In: Atkinson D and McKinlay RG (eds) Integrated Crop Protection: Towards Sustainability? Vol. 63 (pp. 287–296) BCPC Symposium.

Kranz J (1974) Comparison of epidemics. Annual Review of Phytopathology 12: 355–374.

Kranz J (1980) Comparative epidemiology: An evaluation of scope, concepts, and methods. In: Palti J and Kranz J (eds) Comparative Epidemiology. A Tool for Better Disease Management. Pudoc Wageningen. (pp. 18–28).

Lannou C and Savary S (1991) The spatial structure of spontaneous epidemics of different diseases in a groundnut plot. Netherlands Journal of Plant Pathology 97: 355–367.

Madden LV (1980) Quantification of disease progression. Protection Ecology 2: 159–176.

Madden LV (1983) Measuring and modeling crop loss at the field level. Phytopathology 73: 1591–1596.

Madden LV, Pirone TP and Raccah B (1987) Temporal analysis of two viruses increasing in the same tobacco fields. Phytopathology 77: 974–980.

Madden LV and Nutter FW Jr (1995) Modeling crop losses at the field scale. Canadian Journal of Plant Pathology 17: 124–137.

McRoberts N, Hughes G and Savary S (2003) Integrated approaches to understanding and control of diseases and pests in field crops. Australasian Journal of Phytopathology 32: 167–180.

Mc Roberts N, Foster GN, Sutherland A, Willock J and Edwards-Jones G (2000) An evaluation of returns from crop management schemes based on farmer practices. Aspects of Applied Biology 62: 277–284.

Nelder JA and Mead R (1964) A simple method for function minimization. Computer Journal 7: 308–313.

Nelson SC and Campbell LC (1993) Disease progress, defoliation, and spatial pattern in a multiple-pathogen disease complex on white clover. Phytopathology 83: 419–429.

Padwick GW (1956) Losses caused by Plant Diseases in the Tropics. Commonwealth Mycological Inst. Kew, Surrey, Phytopath. Papers No.1. 60 pp.

Peto J (2001) Cancer epidemiology in the last century and the next decade. Nature 411: 390–395.

Pethybridge SJ and Turechek WW (2003) Analysis of the association among three virus infecting hop in Australia. Plant Pathology 52: 158–167.

Pinnschmidt HO, Batchelor WD and Teng PS (1995) Simulation of multiple species pest damage in rice using CERES-rice. Agricultural Systems 48: 193–222.

Pinstrup-Andersen P., de Londono N. and Infante M. (1976) A suggested procedure for estimating yield and production losses in crops. PANS 22: 359–365.

Rabbinge R and Rijsdijk FH (1981) Disease and crop physiology: A modeler's point of view (pp. 201–220). In: Ayres PG (ed.) Effects of Disease on the Physiology of the Growing Plant. Cambridge University Press, Cambridge 258 pp.

Rabbinge R, Jorristma ITM and Schans J (1985) Damage components of powdery mildew in winter wheat. Netherlands Journal of Plant Pathology 91: 235–247.

Rabbinge R and De Wit CT (1989) Systems, models and simulation. In: Rabbinge R, Ward SA and Van Laar HH (eds) Simulation and Systems Management in Crop Protection (pp. 3–15). Pudoc, Wageningen, 420 pp

Rabbinge R, Ward SA and Van Laar HH (eds) (1989) Simulation and Systems Management in Crop Protection. Pudoc, Wageningen 420 pp.

Rolland B, Bouchard C, Loyce C, Meynard JM, Guyomard H, Lonnet P and Doussinault G (2003) Des itinéraires techniques à bas intrants pour des variétés rustiques de blé tender: une alternative pour concilier économie et environnement. Courrier de l'Environnement de l'INRA 49: 47–62.

Rosenbrock HH (1960) An automatic method for finding the greatest or the least value of a function. Computer Journal 3: 175.

Rossing WAH (1991) Simulation of damage in winter wheat caused by the grain aphid *Sitobion avenae*. 3. Calculation of damage at various attainable yield levels. Netherlands Journal of Plant Pathology 97: 87–103.

Rouse DI (1988) Use of crop-growth models to predict the effects of diseases. Annual Review of Phytopathology 26: 183–201.

Savary S (1987a) Enquête sur les maladies fongiques de l'arachide (*Arachis hypogaea*) en Côte d'Ivoire (I). Méthodes d'enquête et étude descriptive: Les conditions culturales et les principales maladies. Netherlands Journal of Plant Pathology 93: 167–188.

Savary S (1987b) The effect of age of the groundnut crop on the development of primary gradients of *Puccinia arachidis* foci. Netherlands Journal of Plant Pathology 93: 15–24.

Savary S and Servat E (1991) Un modèle simplifié de simulation des épidémies de rouille et de cercosporioses de l'arachide. In: Savary S (ed.) Approches de la Pathologie des Cultures Tropicales. Exemple de l'Arachide en Afrique de l'Ouest (pp. 119–145). Karthala/ORSTOM, Paris 288 pp.

Savary S and Zadoks JC (1992a) Analysis of crop loss in the multiple pathosystem groundnut-rust-leaf-spot. I. Six experiments. Crop Protection 11: 99–109.

Savary S and Zadoks JC (1992b) Analysis of crop loss in the multiple pathosystem groundnut-rust-leaf-spot. II. Study of the interaction between diseases and crop intensification in factorial designs. Crop Protection 11: 110–120.

Savary S, Willocquet L, Elazegui FA, Teng PS, Du PV, Zhu D, Tang Q, Lin X, Singh HM and Srivastava RK (2000a) Rice pest constraints in tropical Asia: Characterization of injury profiles in relation to production situations. Plant Disease 84: 341–356.

Savary S, Willocquet W, Elazegui FA, Castilla NP and Teng PS (2000b) Rice pest constraints in tropical Asia: Quantification of yield losses due to rice pests in a range of production situations. Plant Disease 84: 357–369.

Stynes BA (1980) Synoptic methodologies for crop loss assessment. In: Teng PS (ed.) Crop Loss Assessment (pp. 166–175). Misc. Publ. Univ., Minn. Agric. Exp. Stn. 7

Teng PS (1983) Estimating and interpreting disease intensity and loss in commercial fields. Phytopathology 73: 1587–1590.

Teng PS (ed.) (1987) Crop Loss Assessment and Pest Management. APS Press, St. Paul, MN 270 pp.

Teng PS (1994) Integrated pest management in rice. Experimental Agriculture 30: 115–137.

Teng PS and Savary S (1992) Implementing the systems approach in pest management. Agricultural Systems 40: 237–264.

Vander Wal AF, Shearer BL and Zadoks JC (1970) Interaction between *Puccinia recondita* and *Septoria nodorum* on wheat, and its effect on yield. Netherlands Journal of Phytopathology 76: 261–263.

Van der Plank JE (1963) Plant Diseases. Epidemics and Control. Academic Press, New York 349 pp.

Webster JPG (1985) Analysing the economic consequences of sustainability in farming systems. In: Atkinson D and McKinlay RG (eds) Integrated Crop Protection: Towards Sustainability? Vol. 63 (pp. 287–296) BCPC Symposium.

Wiese MV (1982) Crop management by comprehensive appraisal of yield determining variables. Annual Review of Phytopathology 20: 419–432.

Willocquet L, Savary S, Fernandez L, Elazegui FA and Teng PS (2000) Development and evaluation of a multiple-pest, production situation specific model to simulate yield losses of rice in tropical Asia. Ecological Modeling 131: 133–159.

Willocquet L, Savary S, Fernandez L, Elazegui FA, Castilla N, Zhu D, Tang Q, Huang S, Lin X, Singh HM and Srivastava RK (2002) Structure and validation of RICEPEST, a production situation-driven, crop growth model simulating rice yield response to multiple pest injuries for tropical Asia. Ecological Modelling 153: 247–268.

Willocquet L, Elazegui FA, Castilla N, Fernandez L, Fischer KS, Peng S, Teng PS, Srivastava RK, Singh HM, Zhu D and Savary S (2004) Research priorities for rice disease and pest management in tropical Asia: A simulation analysis of yield losses and management efficiencies. Phytopathology 94: 672–682.

Zadoks JC (1971) Systems analysis and the dynamics of epidemics. Phytopathology 61: 600–610.

Zadoks JC (1972) Methodology in epidemiological research. Annual Review of Phytopathology 10: 253–276.

Zadoks JC (1985) On the conceptual basis of crop loss assessment: The threshold theory. Annual Review of Phytopathology 23: 455–473.

Zadok JC and Rabbinge R (1985) Modelling to a purpose. In: Gilligan CA (ed.) Advances in Plant Pathology. Mathematical Modelling of Crop Diseases. Vol. 3 (pp. 231–244) Academic Press, London.

Zadoks JC (1989) EPIPRE, a computer-based decision support system for pest and disease control in wheat: Its development and implementation in Europe. In: Leonard KJ and Fry WE (eds) Plant Disease Epidemiology. Vol. 2 (pp. 3–29) McMillan, New York.

Zimmet P, Alberti KGMM and Shaw J (2001) Global and societal implications of the diabetes epidemic. Nature 411: 782–787.