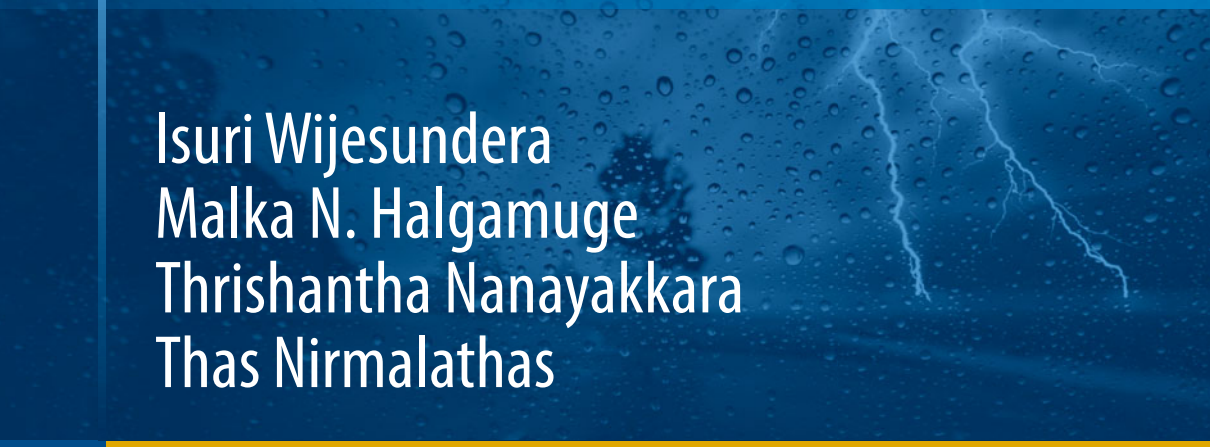
Springer Natural Hazards



Isuri Wijesundera
Malka N. Halgamuge
Thrishantha Nanayakkara
Thas Nirmalathas

Natural Disasters, When Will They Reach Me?

Springer

Springer Natural Hazards

The Springer Natural Hazards series seeks to publish a broad portfolio of scientific books, aiming at researchers, students, and everyone interested in Natural Hazard research. The series includes peer-reviewed monographs, edited volumes, textbooks, and conference proceedings. It covers all categories of hazards such as atmospheric/climatological/oceanographic hazards, storms, tsunamis, floods, avalanches, landslides, erosion, earthquakes, volcanoes, and welcomes book proposals on topics like risk assessment, risk management, and mitigation of hazards, and related subjects.

More information about this series at http://www.springer.com/series/10179

Isuri Wijesundera · Malka N. Halgamuge
Thrishantha Nanayakkara
Thas Nirmalathas

Natural Disasters, When Will They Reach Me?

Springer

Isuri Wijesundera
Melbourne School of Engineering
The University of Melbourne
Melbourne, VIC
Australia

Thrishantha Nanayakkara
Department of Informatics
King's College London
London
UK

Malka N. Halgamuge
Electrical and Electronic Engineering
The University of Melbourne
Melbourne, VIC
Australia

Thas Nirmalathas
Electrical and Electronic Engineering
The University of Melbourne
Melbourne, VIC
Australia

ISSN 2365-0656 ISSN 2365-0664 (electronic)
Springer Natural Hazards
ISBN 978-981-10-1111-5 ISBN 978-981-10-1113-9 (eBook)
DOI 10.1007/978-981-10-1113-9

Library of Congress Control Number: 2016938405

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media Singapore Pte Ltd.

The authors dedicate this book
to the University of Melbourne

Preface

Predicting natural disasters is extremely important in mitigating disaster related destructions. In most scenarios, it is equally difficult to predict such events due to the multitude of factors which affect the formation, propagation and spread of natural disasters.

This book specifically focuses on predicting the mean first passage time to reach a known trigger state for natural disasters modelled as dynamic random processes. An engineering perspective is followed to describe the prediction methods that particularly look at random processes which govern natural disaster dynamics that shows directional dependence and spatial inhomogeneity. Several applications are referred throughout the book including flood, cyclone and fire predictions.

Some of the material presented here can also be found in the following research publications:

1. I. Wijesundera, M.N. Halgamuge, T. Nirmalathas, and T. Nanayakkara (2012), "A geographic primitive based Bayesian framework to predict cyclone induced flooding". Journal of Hydrometeorology, (2012) [1].
2. I. Wijesundera, M.N. Halgamuge, T. Nirmalathas, and T. Nanayakkara (2015), "Bias modified *MFPT* prediction for minimal intervention in robotic walkers", [Under review].
3. I. Wijesundera, M.N. Halgamuge, T. Nirmalathas, and T. Nanayakkara (2015), "Mean first passage time for cyclone motion modelled as biased random walks", [Under review].
4. I. Wijesundera, M.N. Halgamuge, T. Nirmalathas, and T. Nanayakkara (2015), "*MFPT* calculation for random walks in inhomogeneous networks", [Under review].
5. I. Wijesundera (2015), "Estimation of mean first passage time (MFPT) in naturally biased inhomogeneous environments", [PhD Thesis, The University of Melbourne, Australia, 2015] [2].

Melbourne
February 2016

Isuri Wijesundera

References

1. Wijesundera I, Halgamuge MN, Nirmalathas T, Nanayakkara T (2013) Ageographic primitive-based bayesian framework to predict cyclone-inducedflooding*. J Hydrometeorol 14(2):505–523
2. Wijesundera I H (2015) Estimation of mean first passage time (MFPT) in naturally biased inhomogeneous environments. PhD Thesis

Contents

About the Authors

Isuri Wijesundera completed her PhD in engineering from the University of Melbourne in 2016 and obtained her BSc degree in Electronics and Telecommunication Engineering from the University of Moratuwa, Sri Lanka in 2007. She previously worked as a Telecommunications Engineer in Sri Lanka (2007–10) and is currently attached to the Big-Data unit at Telstra, Australia. Her recent work focused on a research concerning analytical methods to find mean first passage time (MFPT) on directionally-biased inhomogeneous media.

Dr Malka N. Halgamuge is a Researcher in the Department of Electrical and Electronic Engineering, University of Melbourne. She obtained her PhD from the same department in 2007. She was awarded the Next Step Initiative Fellowship (2015), Australia-China Young Scientist Fellowship (2014), Dyason Fellowship, University of California (UCLA), Los Angeles, USA (2013), Early Career Researcher (ECR) Award from Alexander von Humboldt Foundation (2013) and Solander Fellowships, Lund University (2007 and 2008). She was the recipient of the Vice-Chancellor's Engagement Award (2010) and Vice-Chancellor's Knowledge Transfer Award (2008) for her research at the University of Melbourne. Her current research interests include investigating biological effect of electromagnetic fields on humans, energy management in wireless sensor networks and prediction of natural disasters (flood and bushfires). She has written one textbook and more than 60 peer-reviewed technical articles.

Prof. Thas Nirmalathas is currently the Director of Melbourne Accelerator Programs (MAP) which supports entrepreneurial activities of the University Community through business acceleration models. He is also the Director of Melbourne Engineering Research Institute (MERIT) which brings the entire research activity of the Melbourne School of Engineering under one research institute. Professor Nirmalathas obtained his BEng and PhD in Electrical and Electronic Engineering from the University of Melbourne in 1993 and 1998 respectively. He has written more than 250 technical articles across areas including microwave photonics, optical wireless network integration, and broadband networks and services.

Thrishantha Nanayakkara received his BSc and MSc degrees in Electrical Engineering from the University of Moratuwa (UM), Sri Lanka (1996), and Saga University (SU), Japan (1998), and PhD in Robotics from SU (2001). He was a postdoctoral research fellow in the Department of Biomedical Engineering, Johns Hopkins University, USA, 2001–2003; a senior lecturer in the Faculty of Engineering at the UM; a Radcliffe Fellow at Harvard University, USA (2008/09), and a research affiliate at MIT (2008/09), USA. He is currently a senior lecturer in the Department of Informatics, King's College London. His research interests are in soft robotics, and robotic interaction with uncertain environments. He has published one textbook and more than 100 peer-reviewed papers.

Acronyms

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
FPT	First Passage Time
GFDL	Geophysical Fluid Dynamics Laboratory
GIS	Geographic Information System
GMM	Gaussian Mixture Model
GP	Geographic Primitive
GUI	Graphical User Interface
HB	Head-to-Back ratio
HP	Huygens principle
HPSO	Self-organising Hierarchical Particle Swarm Optimiser
JTWC	Joint Typhoon Warning Center
LB	Length-to-Breadth ratio
MAP	Maximum a Posteriori
MFPT	Mean First Passage Time
MTTF	Mean Time to Failure
NA	North Atlantic
NP	Network Primitive
PDF	Probability Density Function
PDW	Passive Dynamic Walker
PID	Proportional–Integral–Derivative
PSO	Particle Swarm Optimisation
RW	Rimless Wheel
SF	Scale-Free
SLEV	Second Largest Eigen Value
TC	Tropical Cyclone
TPM	Transition Probability Matrix

Mathematical Notations

Notation Common Throughout the Book

AIC	Akaike Information Criterion
$\hat{T}$	*MFPT*
Φ	Relative direction array
θ	Bias direction
dB	Box dimension
df	Fractal dimension
ds	Spectral dimension
dw	Walk dimension
dw_b	Bias modified walk dimension
$\mathcal{N}()$	Normal distribution
P_{DD}	Probability density distribution
P_r	Probability of reaching
r	Distance/radius
R	Rate of spread
S	Source
t	Time
T	Target
t_{exit}	Exit time from sphere of radius r
U	Bias intensity

Chapter 3

h	Height profile
v	Velocity
σ	Radius of rainfall distribution
$\mathbf{A}$	TPM

D	Water deposit distribution
R	Radius of the Earth (km)

Chapter 4

d	Number of dimensions
σ	Noise level
LB	Length-to-breadth ratio
HB	Head-to-back ratio
$MFPT^*$	*MFPT* with misses considered
$MFPT^r$	*MFPT* given the target is reached

Chapter 5

$\mathfrak{N}$	An NP
n	Number of hops
Y	Hypothetical 2D network
H_i	ith hypothesis
η	Number of NP types
n_d	Number of dimensions of a particle used in HPSO
M	Number of ellipses for optimisation
f_0	Cost function
N	Number of particles considered in HPSO
α	Direction of propagation
Ψ	Number of angle segments

Chapter 1
Introduction

A complex natural process that occurs on earth is referred to as *a natural disaster* when it results in catastrophic life, economic and structural losses. Ranging from cyclone activity, bushfire, and tsunami waves to thunderstorms, earthquake and floods, the severity of natural disasters keeps on increasing along with population growth. Health disasters such as epidemics also add to the above list of catastrophic events. The complexity of the natural processes result in inherent randomness in the formation, development, and propagation of natural disasters.

Forecasting natural disasters is indeed a formidable task. However, timely and accurate forecasts are essential for disaster management operations. Natural disaster management involves several techniques including mitigation, prevention and relief management. The forecasts required are dependent on the disaster type as well as the point of interest (i.e. application). Taking the example of cyclone propagation, desired predictions include severity and track forecasts, rainfall forecasts and flood predictions. While people living on slopes will be interested in the rainfall in terms of possible landslide threats while those living on valleys will be more concerned about the possibility of floods.

This book concentrates on situations where first encounters define the management process. For example, how much time left until evacuation is required before a wildfire reaches ones village? How long until a tsunami wave is expected to hit the local coast following an earthquake at a remote location? How long until the internal activity reaches a threshold forcing a volcanic eruption? These are some of the situations where natural disasters governed by random processes, or more specifically first encounters play a major role in deciding the future of ones actions. Although in some cases it might not be as obvious, a large class of complex processes leading to natural disasters can be modelled as random walks in their respective state spaces.

I. Wijesundera et al., *Natural Disasters, When Will They Reach Me?*,
Springer Natural Hazards, DOI 10.1007/978-981-10-1113-9_1

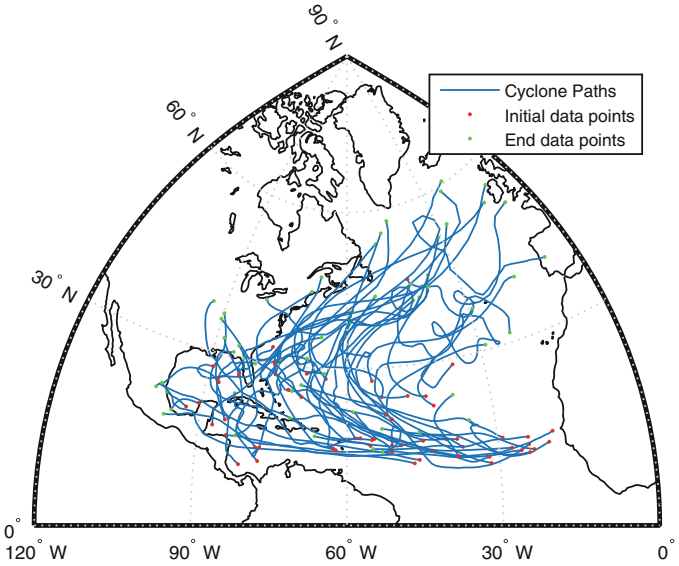


Fig. 1.1 Cyclone motion over the North Atlantic ocean (observed within the year of 2012) forming random walks

1.1 Natural Disasters as Random Walks

The state changes of a natural disaster process can be modelled as random walks owing to the high variability of dynamic environmental factors as well as feedback of the processes themselves. The spatial motion of the eye of a cyclone is an example that can be easily visualised as a random walk (Fig. 1.1).

Random Walk
A *random walk* is a fundamental dynamic process [1] which consists of a *path made up of a succession of random steps.* At each step the walker takes a decision on the next step either from a set of statically or dynamically connected nodes.

However, the state transitions of most other natural disasters can also be modelled as random walks but these might not be as obvious. Taking the example of volcanic eruptions; if the state of the system is expressed as a vector of temperature, earthquake magnitude and sulphur levels, the temporal state transitions result in a three-dimensional random walk.

As described earlier, the main interest of a given random walk changes with the walk itself and the point of interest. Two observers could relate to the same random walk with different interests. For example, while people living in a cyclone prone

coastal area could be more concerned about the estimated landfall time of a cyclone in view of the possible destruction brought by it, people living in drier inner lands would be more interested in the actual path taken by the cyclone in view of how much rain it could bring. The *transport properties* of a random walk is therefore defined with the interested result decided by it. These properties are generally obtained by the joint characteristics of the random walker and the impact of the topological properties of the network [2]. This topic will be discussed further in Chap. 2 under Sect. 2.1.5 and revisited in Chap. 4 under Sect. 4.2.

This book focuses on the large class of processes where the future of the systems depend greatly on the first time a particular target is reached by a random walker (i.e. a particular state is reached).

First passage time (FPT)
The FPT is defined as the time required for a random walker to first reach a predefined target state or set of states [3].

This is sometimes referred in terms of the *First passage probability* ($F(T, \hat{t})$), which is the probability of the random walker's FPT to some target T being some specific time $\hat{t}$. Then FPT is given as the time dependence of first passage probability [3].

The significant role played by first encounters has made the analysis of first passage properties of high importance for many diverse random walk applications. With examples ranging from natural disasters propagation [4–6], propagation of diseases [7–9], spread of gossip [10], stock price changes [11] to fluorescence quenching [3], first passage properties have been discussed widely in physical literature and biology. However, research has not been as extensive in the engineering context possibly due to the highly theoretical treatment of the subject [12]. Many natural processes, including those that govern natural disaster dynamics, can be modelled as random walks in an engineering perspective where the performance relies on first passage properties.

1.1.1 The Mean First Passage Time (MFPT)

For any random walk, the exact solution to finding FPT is almost impossible especially in unbounded domain [13]. Therefore, there are other transport properties that are used to get an indication of FPT. One such transport variable is the probability distribution of FPT being a certain value. The integral of the probabilities as time tends to infinity is 1. An important estimate that is commonly used to get an idea of time to failure is the value at the mean of this probability distribution. Although this value does not entirely represent the probability profile, this metric contains important information and the estimation of this value is much more feasible for

most random walks. Therefore, the Mean First Passage Time (*MFPT*) will be used as the main transport property of interest for the rest of the book.

Mean first passage time (*MFPT*)
The *MFPT* to a target node is the average time expected for a random walker to first find that target on a given network [14].

First passage properties in general and *MFPT* in particular has been used widely to indicate the efficiency of transport for a random walker in a network [2]. For each application, the random walker triggers some event at the first encounter at a specific target state and depending on the application one might want to minimise, maximise or just observe to make precautions. While for target search processes, the aim might be to minimise the *MFPT* [15], epidemic management would look at maximising *MFPT* [16] and for natural disaster management the objective might be to get the prediction for minimising damage [17].

Baring the high importance of *MFPT* estimation, first passage properties have been studied extensively during the last few decades. Powerful results have been published for networks with specific properties and will be referred to during the rest of this book.

1.1.2 Two Common Traits

Two major attributes present in most processes that govern natural disasters and which translated to random walks in their respective state spaces are the network inhomogeneity [7, 18] and network specific bias [19, 20]. These two generally increases the complexity of analysis. This book aims to discuss numerical and probabilistic models to capture the effects of these two characteristics on applications in an engineering perspective.

Biased random walk
A walk which is more likely to move in one direction than another. For such walks, there exists concentration gradients among exit points in a sphere of any radius r from the source node of the random walker. The degree of bias depends on the relative magnitude of the gradient in one direction relative to others.

The bias or potential fields that shape random walks are formed according to the applications, some examples being; thermal gradient for cyclone motion [21], wind distribution for fire propagation [22], topography for flood propagation [23, 24],

population density for disease propagation [25] and more generally state space attractors for any system modelled as random walks in its state space.

Current research on predicting *MFPT* for biased random walks mostly consider network bias in terms of node degree distribution [26] and node weight distribution [27]. These methods view the networks categorised by special characteristics such as scale-free property [28, 29] or the small world property [30]. These concepts will be discussed in detail in Chap. 2. A large class of real-world biased random walks do not possess such ideal conditions to classify by these characteristics. Some of them demonstrate temporal connectivity [31] where a node degree distribution is time dependent and does not represent the whole network. Therefore one major objective of this book is to discuss generic methods to analyse directionally biased random walks and calculate *MFPT* to reach a target state T.

In the context of this book, a network is defined to be inhomogeneous when a random walker's transition probabilities depend on its position.

Inhomogeneous network

A network on which a random walk shows different transport characteristics in different parts of that network.

Taking the example of fire propagation as a random spread, the speed, and direction of propagation is dependent on many factors such as wind, fuel and topography of an area [32]. Therefore, the propagation characteristics vary according to the locality of the walker in the network. Similarly, in the spread of diseases, differences in population density and interaction lead into inhomogeneity of the network [25].

Current literature addressing network inhomogeneity in terms of *MFPT* calculations commonly view the inhomogeneity in terms of node degree distribution [1, 33, 34] giving specific focus to scale free networks. The second objective of this book is developing methods to calculate the *MFPT* for networks that exhibit inhomogeneity in transport properties. This book takes a generic view when inhomogeneity is defined in terms of the variability of transport properties which are used to calculate the *MFPT*. This book tries to use the observation that many inhomogeneous networks consist of 'homogeneous patches' [25] into developing a generic approach to calculate the *MFPT*.

Continuing this discussion, it can be stated that this book will mainly focus on two major areas of *MFPT* calculations in natural disaster predictions; to address directional bias and inhomogeneity in transport properties. This book will focus on several diverse applications when describing random processes including cyclone propagation, and flood propagation. However, it is shown that by keeping the methods generic, the results obtained are not limited to these applications.

This book analyses random walks for obtaining the *MFPT* for a large class of processes governing natural disaster activity in an engineering perspective.

The main focus of this book can be stated as

The design of methods to predict the mean first passage time for random walks resulting from state transitions of processes leading to natural disasters, overcoming the complexities imposed by directional preference and behavioural inhomogeneity.

1.2 How to Read This Book

The main structure of the flow of this book is illustrated in Fig. 1.2 including the details of the case studies used to verify the presented models. The book starts with a discussion of the background for the remaining chapters and includes a literature survey. The novel developments presented in this book starts with a comprehensive case study for the application of predicting cyclone induced flood propagation which is analysed as a collection of interdependent and sequential random processes influenced by the effects of bias and network inhomogeneity. A modularised geographic primitive based Bayesian framework is presented generating efficient real-time predictions. This book then focuses on application independent generic models for *MFPT* prediction for processes translating into random walks in directionally biased media. A novel concept of using bias modified transport variables is introduced in the presented model and verified with archived cyclone track data. The book next focuses on processes translating onto random walks on networks with inhomogeneous transport property distributions. A heuristic approach is used to develop algorithms to partition the network using the novel concept of 'network primitives' and using a hop-wise calculation approach.

The remaining chapters of the book, shown in Fig. 1.2, are outlined below.

Chapter 2: Background guide to random walk analysis

This chapter outlines the background and the state of art for the generic problem of *MFPT* prediction for biased walks in inhomogeneous media in an engineering applications perspective. The material presented are divided into three major subsections which will be useful throughout the remaining chapters. The first part consists of the main notion in the context of book with an adaptation of the theoretical background. The second part presents a literature review of recent advancements in *MFPT* prediction methods and the last part includes an introduction to a few mathematical tools used in the next chapters.

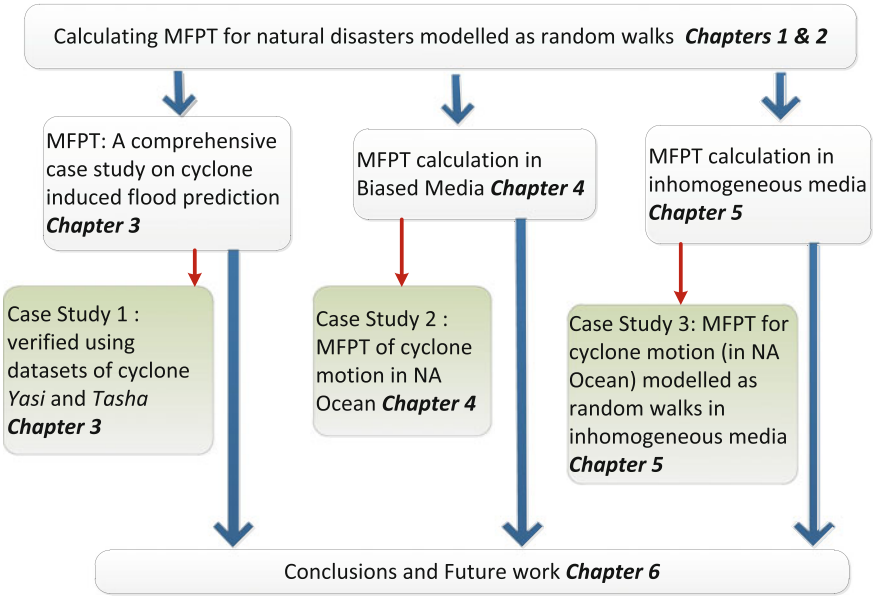


Fig. 1.2 Overview of the general research areas discussed and case studies in relation to the book structure

Chapter 3: Predicting cyclone induced flood: A comprehensive case study

Chapter 3 of this book comprises of a comprehensive study on developing methods for the specific application of predicting the *MFPT* for cyclone induced flood propagation. Introducing the novel concept of geographic primitives (GP), the terrain is divided using the flood propagation vector fields to account for network inhomogeneity. Easy and flexible data assimilation is obtained through a modularised approach which allows integrating outputs of other numerical models when they become available while being able to provide earliest predictions with minimal data. Transition probability matrices per GP are used to summarise simultaneous events distributed in the environment resulting in biased propagation, facilitating efficient real-time predictions using Markov chain analysis. Comparative studies are presented showing the improvements of efficiency and accuracy when compared to other models commonly used.

Chapter 4: First arrival time for natural disasters modelled as biased networks

This chapter aims on developing generic *MFPT* prediction methods for random walks subject to directional bias irrespective of application. Analysing an array of simulated random walks on a set of hypothetical networks, a case invariant modification of transport variables is presented as a viable solution to the problem of addressing directional bias. The empirical relationship obtained on calculating the bias modified walk dimension is shown to increase the prediction accuracy in the presence of

directional bias when compared to using conventional transport variables. Finally a detailed case study is presented using the real dataset of archived cyclone track data over the North Atlantic Ocean for the past 60 years. The comparative studies of prediction results and the generic nature of model development indicate wider applicability of the presented methods.

Chapter 5: Calculating *MFPT* for processes mapping into random walks in inhomogeneous networks

This chapter explores *MFPT* prediction methods for processes which lead to state changes behaving as random walks on a network inhomogeneous in transport properties (which are discussed in Chap. 4). For such networks, the application of *MFPT* calculation methods introduced in Chap. 4 along with many other methods described in literature are not straight forward. This chapter proposes a solution through using the concept of dividing the node distribution into patches/clusters known as network primitives (NPs) where all nodes within each primitive share common transport variables, and adopting a hop-wise approach to calculate *MFPT* between any source and target pair as an extension to the methods described in Chap. 4. This methodology's potential is demonstrated through simulated random walks and with a case study using the dataset of past cyclone tracks over the North Atlantic Ocean. The predictions using the presented method are compared to real data averages and predictions assuming homogeneous transport properties (discussed in Chap. 4).

Chapter 6: Conclusions and future research directions

The major conclusions of this book are summarised in Chap. 6, followed by possible directions for future research. A short discussion of initial simulations on *MFPT* calculation for random spreads (such as spreading of bush-fires) is also included.

1.3 Summary

This chapter provided an introduction and an entrance into the rest of the book which discusses methods of predicting time-to-fail for natural disasters by modelling the state changes as random processes. Specific attention is given to two common properties of such random processes (i.e. directional bias and spatial inhomogeneity). The rest of the book discusses prediction methods with several example disaster scenarios.

References

1. Noh JD, Rieger H (2004) Random walks on complex networks. Phys Rev Lett 92(11):118–701
2. Tejedor V, Benichou O, Voituriez R (2011) Close or connected: distance and connectivity effects on transport in networks. Physl Rev E 83(6):066–102

3. Redner S (2001) A guide to first-passage processes. Cambridge University Press, Cambridge
4. Gall JS, Ginis I, Lin SJ, Marchok TP, Chen JH (2011a) Experimental tropical cyclone prediction using the GFDL 25-km-resolution global atmospheric model. Weather Forecast 26(6):1008–1019
5. Ryan CJ (1993) Costs and benefits of tropical cyclones, severe thunderstorms and bushfires in australia. Clim Chang 25(3–4):353–367
6. Wijesundera I, Halgamuge MN, Nirmalathas T, Nanayakkara T (2013) A geographic primitive-based bayesian framework to predict cyclone-induced flooding*. J Hydrometeorol 14(2):505–523
7. Giuggioli L, Pérez-Becker S, Sanders DP (2012) Encounter times in spatially-heterogeneous populations and their role in epidemic spread
8. Eubank S, Guclu H, Kumar VA, Marathe MV, Srinivasan A, Toroczkai Z, Wang N (2004) Modelling disease outbreaks in realistic urban social networks. Nature 429(6988):180–184
9. Berger SI, Ma'ayan A, Iyengar R (2010) Systems pharmacology of arrhythmias. Sci Signal 3(118):ra30
10. Isham V, Harden S, Nekovee M (2010) Stochastic epidemics and rumours on finite random networks. Phys A: Stat Mech Appl 389(3):561–576
11. Fama EF (1995) Random walks in stock market prices. Financ Anal J 51(1):75–80
12. McKenzie HW, Lewis MA, Merrill EH (2009) First passage time analysis of animal movement and insights into the functional response. Bull Math Biol 71(1):107–129
13. Condamin S, Benichou O, Tejedor V, Voituriez R, Klafter J (2007b) First-passage times in complex scale-invariant media. Nature 450(7166):77–80
14. Tejedor V (2012) Random walks and first-passage properties. Thesis
15. Shlesinger MF (2006) Mathematical physics: search research. Nature 443(7109):281–282
16. Newman ME (2002) Spread of epidemic disease on networks. Phys Rev E 66(1):016–128
17. Torn RD, Whitaker JS, Pegion P, Hamill TM, Hakim GJ (2015) Diagnosis of the source of gfs medium-range track errors in hurricane sandy (2012). Mon Weather Rev 143(1):132–152
18. Godec A, Metzler R (2015) Optimization and universality of brownian search in quenched heterogeneous media. arXiv preprint arXiv:150300558
19. Azar Y, Broder AZ, Karlin AR, Linial N, Phillips S (1992) Biased random walks. In: Proceedings of the twenty-fourth annual ACM symposium on Theory of computing, ACM, pp 1–9
20. Fronczak A, Fronczak P (2009) Biased random walks in complex networks: the role of local navigation rules. Phys Rev E 80(1):016–107
21. Torn RD, Davis CA (2012) The influence of shallow convection on tropical cyclone track forecasts. Mon Weather Rev 140(7):2188–2197
22. Anderson HE (1983) Predicting wind-driven wild land fire size and shape. US Department of Agriculture Forest Service, Intermountain Forest and Range Experiment Station, Ogden
23. Asante KO, Arlan GA, Pervez S, Rowland J, Artan GA (2008b) A linear geospatial streamflow modeling system for data sparse environments. Int J River Basin Manage 6(3):1–9
24. Zerger A, Wealands S (2004a) Beyond modelling: linking models with GIS for flood risk management. Natural Hazard 33(2):191–208
25. Ferguson NM, Keeling MJ, Edmunds WJ, Gant R, Grenfell BT, Amderson RM, Leach S (2003) Planning for smallpox outbreaks. Nature 425(6959):681–685. doi:10.1038/nature02007
26. Skarpalezos L, Kittas A, Argyrakis P, Cohen R, Havlin S (2014) Efficiency of message transmission using biased random walks in complex networks in the presence of traps. arXiv preprint arXiv:14062437
27. Lee ZQ, Hsu WJ, Lin M (2014) Estimating mean first passage time of biased random walks with short relaxation time on complex networks. PLoS One 9(4):e93–348. doi:10.1371/journal.pone.0093348
28. Agliari E, Burioni R (2009) Random walks on deterministic scale-free networks: exact results. Phys Rev E 80(3):031–125
29. Barabasi AL (2009) Scale-free networks: a decade and beyond. Science 325(5939):412–3. doi:10.1126/science.1173299

30. Aparicio JP, Pascual M (2007) Building epidemiological models from r0: an implicit treatment of transmission in networks. Proc Biol Sci 274(1609):505–512
31. Starnini M, Baronchelli A, Barrat A, Pastor-Satorras R (2012) Random walks on temporal networks. Phys Rev E 85(5):056–115
32. Finney M, Station RMR (1998) FARSITE. Fire area simulator-model development and evaluation, US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden
33. Hwang S, Lee DS, Kahng B (2012) First passage time for random walks in heterogeneous networks. Phys Rev Lett 109(8):088–701
34. Gallos L, Song C, Havlin S, Makse H (2007) Scaling theory of transport in complex biological networks. Proc Natl Acad Sci 104(19):7746

Chapter 2
Background Guide to Random Walk Analysis

Abstract This chapter consists of three major subsections. Following a discussion of the ability to model a large class of environmental processes governing natural disasters as random walks, this chapter provides the theoretical background for this book which will be useful throughout the rest of the chapters. The first subsection will summarise the main notion used throughout the book and the adaptation of the theoretical background for natural disaster prediction in an engineering applications perspective. The second subsection will review recent research advancements in *MFPT* prediction methods. The final subsection will briefly describe some mathematical tools used in the proceeding chapters.

Random walks and their properties have been studied widely in literature due to the important role played by them in a range of disciplines. These methods can be adapted and improved for predicting natural disasters. This chapter discusses the literature background for the remaining chapters.

2.1 Introducing Some Basic Notions

Some common concepts that are used in the next chapters are presented in this section in the perspective of this book. Since the main focus of this book is on natural disasters, these definitions might be slightly different from traditional definitions in different disciplines. The more common definitions referred throughout this book are presented in the Glossary at the end of the book.

2.1.1 Random Processes

This book will use the term 'random' with the meaning of being unpredictable or non-deterministic. When one tosses a coin, it is commonly understood that one of the two sides would turn up with almost equal likelihood. Although it might be theoretically possible to calculate exactly which side turns up using the initial conditions such as coin weight distribution, initial velocity and point of contact, the complexity

I. Wijesundera et al., *Natural Disasters, When Will They Reach Me?*,
Springer Natural Hazards, DOI 10.1007/978-981-10-1113-9_2

of the dynamics and the interaction with the surrounding environment makes this calculation practically very difficult. For problems such as these, it is common to use probability distributions rather than absolute values to describe and analyse the system. An experiment tossing a coin a large N number of times and showing the 'head' up n_h number of times would describe the system with a probability of n_h/N of getting a 'head' and a $(1 - n_h/N)$ of getting a 'tail'.

This book suggests that considering complex dynamic systems like the motion of tropical cyclones as random processes is a feasible approach of getting better understanding of such systems.

- A **random process** is a probabilistic process characterised by a probability distribution(s) of one or more *random variables* [1].
- A **random variable** is one which can take any value out of a continuous or discrete range but the exact value is not deterministic.

There is a large class of complex real world random processes that govern the dynamics of natural disasters and of which the states map into random walks in their respective state spaces.

2.1.2 Random Walks

A *random walk* is a fundamental dynamic process [2] which consists of a *path made up of a succession of random steps.* At each step the walker takes a decision on the next step either from static or dynamically connected nodes. This book defines a random walk as *a stochastic process which results in a non-deterministic array of state changes*. The state space is application dependent. An example of a one dimensional random walk is the monthly mean temperature profiles. Figure 2.1 shows the monthly mean temperature profiles of Melbourne, Australia between 1971 and 2016.

A few diverse example random walks which result from real world dynamic systems are temporal variation of share prices and exchange rates [1], disaster propagation [4–6], spread of epidemics and pandemics [7], path of gossip spread [8], routing of information packets in large-scale infrastructures such as the Internet [9], and many other state transitions of physical non-perpetual systems.

Analysing random walks have also gained a lot of popularity due to them being one of the simplest mechanisms of exploring of and search on networks [10]. Random search processes are widespread in nature ranging from target searching molecules in living cells [11] to animals searching for food [12]. Search processes have been shown to be optimal if the walker follows the shortest path (i.e. one with the smallest number of links) between the two nodes under consideration [2]. But the shortest

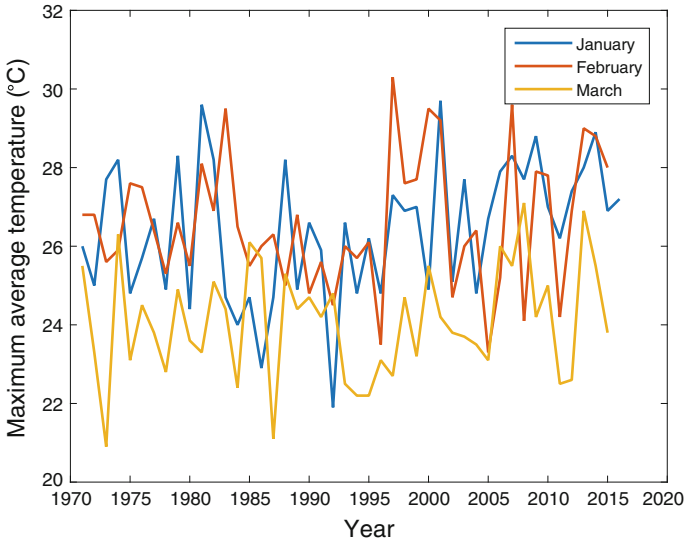


Fig. 2.1 One dimensional random walk of the monthly mean temperature of Melbourne, Australia between 1971 and 2016 (data from [3])

path can be only found theoretically when the global connectivity of the network is known, which is not always possible. Therefore, random walks have been used extensively as a mechanism to explore a network where only local connectivity is known at each node.

It has been observed that a large class of random walks resulting from natural processes are not purely Brownian [1]. This is true for many complex systems in natural disaster dynamics and therefore the methods described in the later chapters do not require the processes to be Brownian.

2.1.3 Networks

In the context of random walks of interest in this book, a network is defined as follows:

> A ***network*** is the complete set of states which the random walker can occupy.

Networks can either be *discrete* or *continuous*. A discrete network for a random walker is one where there are a limited number of nodes which are reachable by the walker on that network. A continuous network is one where the states which are achievable by the walker are continuous in nature and therefore *infinite*. For such

network, the probability of a random walker being at a location is expressed in terms of a subset on state space as opposed to a single point and a 'probability density distribution' (P_{DD}) is used to calculate this.

For a d dimensional state space, the probability that a walker is inside the subspace $\forall_{i=1:d}\ a_i < x_i < b_i$ is given as $\int P_{DD}(x_i, ..., x_d)$.

Another differentiation between continuous networks is that when the state space that is reachable by the walker under any circumstance is limited, the network is said to be *bounded* where boundaries can be *reflecting* or *absorbing*.

Random walks on a continuous state space or network space can be either continuous or discrete in nature. For engineering solutions, even continuous walks are usually discretised by measurement and therefore a random walk consists of a succession of *steps* in the time domain. Selecting the next step of the walk could either by solely dependent on the current step, where such random walk is said to possess the *Markovian property*, or might be shaped by the history of preceding steps (non-Markovian). Either way, the sum of the probabilities of the immediate neighbours being selected is equal to one.

Connected networks are ones where there would be *edges* that would connect the *nodes* such that there would be at least one path the walker can take to reach any target node from its current node. The *chemical distance* between two connected nodes is the shortest number of nodes between them [13]. The *degree* of a node is the number of edges connected to that node. There is another class of *dynamically connected* or *temporal* networks [14] where the connections or the edges of the network are dynamic and dependent on the current state of the network and walker. These types of networks, although have been given relatively less attention, are quite common in real world dynamic systems. For example, if there is a bushfire of which the spread relies on the fire at a single point reaching discrete fuel nodes (e.g. trees, bushes), the reachability of nodes depends on the spread rate which is defined by the wind and topography in the area. Therefore nodes that would otherwise be isolated could be connected given stronger winds. For a discrete *connected* or *dynamically connected* network, the probability of the walker being at any given node is a value dependent on the current state of the network and would always add up to one for the total node set. Most parts of this book would consider network connections to be temporal.

2.1.3.1 Random Walks on Complex Networks

A discrete network is said to be ***complex*** when the node structure becomes complicated and different from a normal lattice structure.

Table 2.1 Common features of complex networks

Scale-free networks	The degree distribution of these networks follow the power-low (at least asymptotically) where the probability that a given node has k neighbours is $P(k)/k$ [1]. Some example scale-free networks [24] are social networks [14], protein interaction networks [17] and metabolic network
Scale-invariant networks	The network properties would remain almost unchanged irrespective of the scale on which the network is looked at [18]. For such networks the number of nodes within a shell of size r is $N(r) \propto r^{df}$, where df is the fractal dimension [1] common at any node. Among these networks, deterministic fractal networks are when the same pattern is observed at all scales
Self-similar networks	The main properties (e.g. degree distribution) of these networks remain unchanged under length scale renormalisation. In [13] renormalisation is when a fractal network is covered with N_B number of non-overlapping boxes and when each box is replaced by a super-node and the links between boxes are transferred to boxes
Small-world networks	The network diameter D or the longer chemical distance between any two nodes, is proportional as $D \propto ln(N)$ for a network with N nodes [1]. A phenomenon observed as early as 1929 and is well known as *the six degrees of separation* [15]

A considerable portion of current literature on *MFPT* estimations looks at complex networks. The study on complex network structures began in the 1950s with Erdös-Renyi random graph which assumes that complex networks are wired randomly together [15]. This hypothesis was the basis of most initial work on complex networks in sociology, biology and computer sciences. Scientists have now identified several topological properties common in complex networks that are very different from usual Euclidean lattices [1]. Some common *complex features* are presented in Table 2.1.

Many real world networks have been shown to possess one or more of these features [2, 19]. An example such network that is commonly used to describe these properties is the yeast protein interaction network (PIN) which is scale-free and also has the small-world structure [1]. Some other such networks are the world wide web [20], social networks [14] and electric power transmission grids [21]. This book discusses the applicability of some of these features for networks on which random walks that describe the processes governing natural disasters occur.

Many theoretical methods have been researched on to predict the motion of a random walker in networks having a deterministic node structure (i.e. for networks which would fall under a predefined network classification whether simple or complex). An observation worth noting here is that there are many random walks describing processes that govern natural disaster dynamics that do not fall into such deterministic structures or are composed as a compound of several such structures having ideal temporal, directional, and spatial conditions.

2.1.4 Random Spread Versus Diffusion

Up to this point the considered networks consisted of only a single propagator in the network. Extending this concept, the cumulative behaviour of a large number of independent random walkers in the same network is defined as *random diffusion*. Diffusion is the continuum limit of a random-walk process [22]. As common as diffusions are in real world systems (e.g. Brownian motion), a contrasting phenomenon of *random spreading* exists where it is also a large collection of random walkers but which are interdependent. A similar concept is a *branching random walk* where at each step, several steps are branched out. The network is changed by each passing walker and therefore will have to be considered in calculations. An example random spread is the spread of bushfire over a network of fuel distribution biased by wind and topography (see Sect. 6.1 of Chap. 6). *Anomalous diffusion* is when the mean square displacement r^2 is no longer proportional to the time t [1].

2.1.5 Transport Characteristics Describing Random Walks

Understanding the properties of transport for random walks describing natural processes is crucial for a large class of applications [13]. Depending on the application and the network type, different transport properties are used to evaluate the performance of a system [19]. Some of the main such transport characteristics are summarised in Table 2.2.

In addition to the characteristics described in Table 2.2, there are several others which have been used to describe random walks such as 'transience', 'recurrence', 'number of distinct sites visited', 'average recurrence time', 'number of returns to origin', and 'occupancy of set by a random walk' [26, 27]. Other transport properties such as the 'walk dimension' (dw), 'fractal dimension' (df), 'spectral dimension' (ds), 'random walk centrality' (C) [2], etc. which are commonly used to measure certain characteristics of the network are also used in literature to derive methods to calculate the properties such as the $MFPT$. These will be discussed in detail in Chap. 4.

For networks of regular lattices, most of these transport properties can be determined analytically. This is however not the case for many real-world networks and therefore have been an objective of active research [10]. This book mainly focusses on calculating the $MFPT$ for real world dynamic systems behind natural disasters of which state transitions form random walks on Euclidean state spaces.

Table 2.2 Common transport properties

Cover time	The time for a random walk to visit all sites of the network [23]
Mean exit time from a sphere (t_{exit})	This is the first time a random walker reaches any point at a distance r from its starting point. This quantity is commonly used in analysing Brownian motion in Euclidean spaces [16]. In many cases, the length-scale-invariant properties of the random walk leads to the scaling form t_{exit}/r^{dw}, which defines the walk dimension dw [24]. This will be discussed in detail in Chap. 4. As opposed to *MFPT*, t_{exit} is not sensitive to the confining environment as only a sphere of r radius is explored
Mean first passage time (*MFPT*)	This is the probability that a random walker or a diffusing particle hits a specified target point T for the first time starting from a source point S. As discussed in the introduction, this metric is the most common transport feature used in applications where first encounters trigger important events. It has been shown in [24] that the *MFPT* crucially depends on the confining environment and the distance between S and T [23]
Mean return time	The average time the walker takes to come back to the starting node [10]
Mixing time	The time scale that determines how the probability distribution approaches its limiting behaviour [23]
Number of nodes visited	This is the number of nodes visited during a given time period t
Occupation probability	The probability that a diffusing particle is located at a known location T at time t [22]
Relaxation time	Asymptotic time of convergence to the *stationary distribution* [25]
Splitting probability	In the existence of two absorbing boundaries (B_1, B_2), this is the probability of absorbing in B_1 and B_2 as a function of S [22]
Stationary distribution	The probability density distribution of the random walker's position as $t \to \infty$
Survival probability	For an absorbing domain, this equals the integral of the first passage probability over all time and spatial extent of the boundary [22]

2.2 *MFPT*: A Literature Survey

Estimating *MFPT* for random walks has gained a lot of attention in various fields in the past decade due to the significant role played by first encounters in a large range of application fields. This book looks at two of the main approaches that have been used in literature to address this problem. One approach is through analytical expressions using transport properties of joint characteristics of the network and the random walker. The next is to use Eigen value analysis on transition probability matrices (TPM) for the state transitions modelled as random walks.

2.2.1 Estimating MFPT Using Random Walks' Transport Characteristics

Analysing transport properties of random walks on networks with known characteristics can be found in numerous studies. Among them are Brownian motion [16, 28], Lévy flights [24, 29], random walks on scale-invariant media [24], fractals [30], scale-free networks [2, 31] and many others. Studies have also been conducted on random walks which combines several types of motion [7]. Until recently, most methods of determining first passage properties in bounded domains have been mainly limited to one-dimensional geometries with higher spatial dimensions considered only in homogeneous media [24]. This book will discuss some of these methods in Chap. 4 with special consideration to methods presented in [24] that uses the pseudo Green function, that is referred throughout several chapters. However one limitation in most of the methods proposed in literature is that the results are generally applicable to deterministic networks with homogeneous node distributions [32, 33].

2.2.2 Analysing the Transition Probability Matrix to Estimate MFPT

The other method commonly discussed in many *MFPT* prediction methods is by Eigen value analysis on transition probability matrices (TPM) for dynamic systems modelled as random walks on their respective state spaces. For random walks possessing the 'Markovian' property (discussed in the next subsection), TPMs are a useful tool for reducing the complexity of *MFPT* computation.

Transition probability matrix (TPM)
For a system with n possible discrete states (network nodes), an $n \times n$ TPM gives the transition probabilities from a state i to j in the (i, j)th cell.

If the state transitions are Markovian, TPM for M state changes is simply obtained by the Markov chain with the result being the current state vector multiplied by $[TPM]^M$. The probability distribution of a regular Markov chain will converge to a unique stationary distribution regardless of the starting position [19].

Eigen value analysis gives the relaxation time of a Markov Chain in a global result as $1/(1 - \varepsilon_2)$, where ε_2 is the second largest eigenvalue [19].

Therefore if the state space can be digitised, TPM is a useful tool.

This is a more popular approach for many engineering applications such as that discussed in [34]. One advantage of this approach is that it is possible to derive the transition probability matrix from prior knowledge of the system either in the form of system modelling or past datasets. These will be discussed in detail in Chap. 3. One major drawback of this approach is that the global connectivity of the network generally needs to be known for generating the TPM and the analysis is only applicable to connected networks [19]. And since the computation involves multiple matrix multiplication, when applied to large scale networks with millions of nodes, the computational complexity makes the method infeasible for many applications [19]. Moreover, the solutions obtained from this fundamental matrix approach are known to be too generic to interpret when it is unclear on which factors (eigenvalue, node degree, connectivity, etc.) govern the *MFPT* from the solution expression [19]. Therefore, although the TPM method provides a very versatile approach for *MFPT* calculations, there still exists room for further improvements.

2.2.3 Random Walks in Natural Disaster Dynamics

Most *MFPT* prediction methods described in literature assume that the random walker has no memory. This is called the 'Markov property' (discussed under mathematical tools). However, most random walks governing natural disaster dynamics do not follow the strict Markov conditions. It has also been observed that a large class of natural processes are not always Brownian [1]. This book looks at systems which deviate from such ideal conditions especially in the existence of directional bias and network inhomogeneity. Current literature on these aspects are discussed in the following sections.

2.2.3.1 Bias on Natural Systems

An isotropic random walk that starts at the origin [22] is the basis on which many prediction methods depend on. However, in the past few years more literature has addressed the question of bias in *MFPT* calculations [2, 12, 19, 22]. In [22] it is shown that for a biased random walk, *MFPT* is dominated by the bias.

> For a walk on a d dimensional Euclidean space, where $d > 1$, biased random walks are mainly within a narrow cone of t length along the bias direction and having a width of $t^{1/2}$ where t is the time.

The most popular approach in current literature on the definition of bias in random walks is in the form of bias resulting from the node degree distribution [35]. Zhuo et al. [19] have shown that there are cases where the nodes share the same node

degree while having greatly different local topologies. Here they look at bias as being controlled by arbitrarily controlled node weights and show that the *MFPT* is a function of both node degree distribution and node weight distribution. The work in [36] shows that directed self-avoiding walks show anisotropic behaviour by decomposing the random walk as a forward walk along the preferred direction and a random walk perpendicular to that direction.

In one example, animal motion's central foraging behaviour is described as biased random walks in [12] describing *MFPT* through the effect of landscape on animal movement and search time. Directional bias has been introduced into the model with an advection term which describes the directed motion component. The relation between the magnitudes of the directed and random components of movement is shown to determine the shape of the *MFPT* curve.

When the random walks are biased

$$MFPT(S \rightarrow T) \neq MFPT(T \rightarrow S). \tag{2.1}$$

The asymmetry is characterised with the difference in the *MFPT*'s [22, 23]. In [2] it is shown that the difference can be determined by the *random walk centrality*.

The directional bias has also been addressed using TPM based methods in order to calculate *MFPT* for random walks [19] but is still computationally infeasible when networks topologies are not 100 % known or when networks have millions of nodes. This book discusses both methods of addressing directional bias on random walks in Chaps. 3, and 4.

Effect of Bias in Anomalous Diffusion

Biased anomalous diffusion has been shown in [37] as constant external fields giving rise to constant currents which increase with the strength of the biasing field. The drift velocity is shown to respond non-monotonically to the biasing field.

Correlated Walks and Biased Random Walks

Initial work analysing random walks considered the walkers to be 'uncorrelated'.

Uncorrelated walks are where each step taken by the walker is essentially independent from the previous step. Such motion sums up the concept of 'Brownian' motion.

It has long been realised that a large class of random walks are in fact correlated in the form of either *auto-correlation* where the walkers current step is shaped by its previous steps or *cross-correlation* where there is a relationship between two walkers

in that network. The effect of problem specific bias, is another factor shaping random walks in that network. The effect of bias may or may not lead to correlated random walks as will be explained with the following examples. Animal migrating paths which are shaped by biasing factors such as wind, temperature or water [12] as well as life history and morphological traits [38] often lead to correlated random walks as the vector field biasing the network remains more or less unchanged. But in the example of fire spread, the dynamicity of the vector field created by winds make the random walks biased but not correlated [39]. This book concentrates on the effect of bias on random walks irrespective of correlation.

A network without loops is generally called a 'tree' and it is impossible for a walker with the memory of the last step to go in a loop in a network with a tree structure. Loops in the network can lead to loops in the random walks. The vector field resulting from biasing factors in a network could be a deciding factor for the existence of loops in a random walk. This book considers random walks without loops for simplicity and limitation of computational resources. Moreover, the effect of biasing vector fields often shape the probability density distribution of selecting the next step at each step of a random walk. The probability of reaching nodes in the network is also shaped by the biasing vector field. Furthermore, biased networks lead to 'directed' links where the motion from one node to another is not the same when the nodes are reversed.

2.2.3.2 Inhomogeneous Networks

Network inhomogeneity is defined when a random walk has transition probabilities dependant on its position.

In most *MFPT* prediction methods found in literature, transport properties are assumed to be scale-invariant. Most use network renormalisation schemes such as those found in [13, 15, 24], that use self-similarity and scale-free properties. However, in recent work, network inhomogeneity has been addressed in several approaches. Structural inhomogeneity effects of nature of the diffusive and relaxation dynamics of a random walk has been studied in [2]. The majority of literature on network inhomogeneity looks at the inhomogeneity in node degree distribution [2, 13, 40] with some also looking at local weight distributions [19]. For animal motion, network inhomogeneity has been introduced in terms of landscape features in [12] where landscape heterogeneity on animal movement is considered in two parts: movement rate and movement direction. Seasonal hot-spots have been identified in motion of some predatory animals in [41]. In [2] it is shown that for homogeneous networks, random walks' centrality is the same for all nodes and any difference determines the difference of *MFPT*.

2.3 Some Mathematical Tools

In this section, a summary of some of the mathematical tools used throughout this book is included. The reader is encouraged to refer to this section for clarifications regarding the mathematical tools in the context of the methods discussed in the next chapters. The tools which are used only once are discussed within the chapter itself.

2.3.1 Bayesian Probability Theory

Contrasting to the Frequentist view of probability being a result of infinitely many attempts, in the Bayesian view, probability is a measure of belief of predicted outcome [42]. It uses prior data to refine prediction and hypothesis.

The main equation governing Bayesian probability is

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \tag{2.2}$$

where h is the hypothesis (prior probability) and D denotes the prior data (evidence). $P(h|D)$ gives the posterior probability and $P(D|h)$ gives the generative model or the likelihood.

In making an inference, several tools are available depending on the problem.

2.3.1.1 Maximum Likelihood (ML) Hypothesis

This is considered a good prediction option when no past data is available but not the ideal choice when there is available prior knowledge. The equation for this is

$$h_{ML} = argmax_{h \in H} P(D|h). \tag{2.3}$$

2.3.1.2 Maximum A Posteriori (MAP) Estimate

The MAP estimate gives the hypothesis with the highest probability given observed data. It is still a point estimate and obtained using

$$h_{MAP} = argmax_{h \in H} P(D|h)P(h). \tag{2.4}$$

2.3.1.3 The Bayesian Estimate

This method uses the full probability density distribution to infer on the system.

2.3.1.4 Bayes' Filtering Method

The method of Bayes' filtering is a practical approach which uses the posterior probability as the prior probability in the next step. This is used for rainfall prediction in Chap. 3. This method is used in many practical applications through many variations such as in *Kalman filter* which assumes linear dynamics and Gaussian noise, and a more recent version known as the *particle filter*. This gives the probability of being at a state X_t at time t given a sequence of observations $y = (y_1, y_2, \ldots, y_t)$ per [42] as

$$P(X_t|y_1, y_2, \ldots, y_t) \propto P(y_t|X_t)P(X_t|X_{t-1})P(X_{t-1}|y_1, y_2, \ldots, y_t). \tag{2.5}$$

2.3.2 Markov Matrices

A Markov Process is one where when given the current state of a process at time t, any information regarding the states of the process until time t does not effect on the prognosis of the system state after time t [43, 44].

In the context of a random walk, this is when the walker is at the ith node, the probability $P_{i,j}$ to jump to jth node does not depend on previous steps [1]. This memoryless property is called the *Markovian property*.

The Markovian property for a random walks is

$$P(s_{n+1} = x|s_1, s_2, \ldots, s_n) = P(s_{n+1} = x|s_n). \tag{2.6}$$

This probability condition defines the *Discrete Markov Chain*. A *Markov matrix*, also known as a stochastic matrix, describes the transitions of a Markov chain. The jth element of the ith row consists of the non-negative value $P_{i,j}$. A process with n possible states produces an $n \times n$ Markov matrix ($A_{n\times n}$). Two major properties of this matrix are,

- $\sum_{i=1}^{n} p_{ij} = 1, \ \forall j \in \{1, 2, \ldots, n\}$ and
- $0 \leq p_{ij} \leq 1, \quad \forall i, j \in \{1, 2, \ldots, n\}$.

The largest Eigen-value of a Markov matrix is 1. Finally, because when a random walk has no memory, it is "renewed" every time a specific node is reached. Thus recurrence also implies that every site is visited infinitely often for an isotropic random walk with the Markovian Property.

2.3.3 Linear Least Square Fitting Technique

When a data model is developed to explain a given system, and when such model is linear in the model coefficients, *linear regression* is used to fit the model to the dataset. The most common linear regression type is the *linear least square fit*. This method is used throughout this book to fit linear models.

If a linear model is given as

$$y = f(x) \tag{2.7}$$

the least squares method minimises the sum of the square of residuals. A residual for the kth data point (r_k), is the difference between the response observed (y_k) and the model response ($\hat{y}_k$) given as

$$r_k = y_k - \hat{y}_k \tag{2.8}$$

and the sum of square of residuals (S) is

$$S = \sum_{k=1}^{n} r_k^2 = \sum_{k=1}^{n} (y_k - \hat{y}_k)^2 \tag{2.9}$$

for n number of data points included in the fit.

The individual regression coefficients are found by solving the differentiation of S with respect to each parameter as

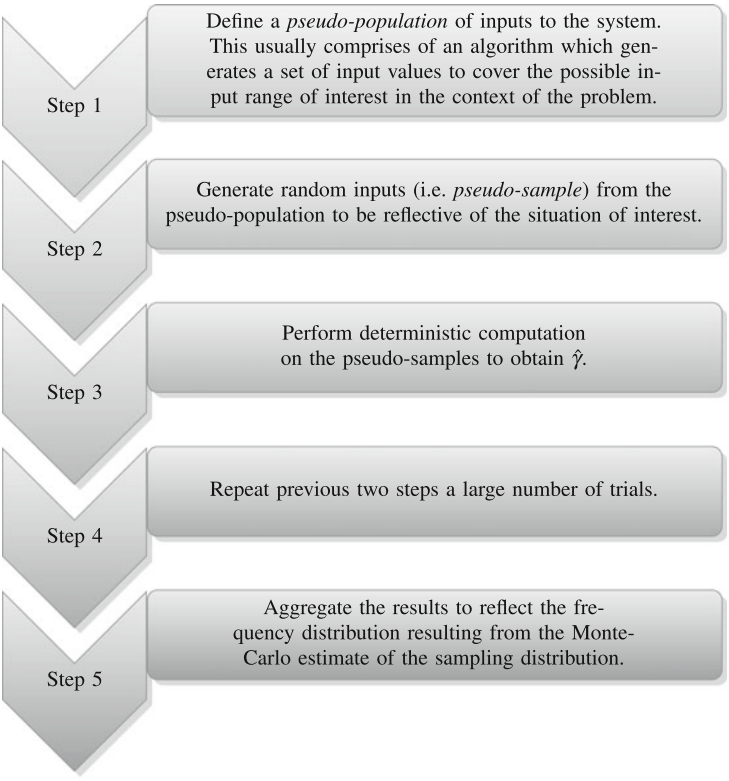
$$\underset{(i=1,2,\ldots,m)}{\forall} \frac{\partial [S]}{\partial p_i} = 0 \tag{2.10}$$

where $p_i; \ i = 1, 2, \ldots, m$ are the parameters.

2.3.4 Monte-Carlo Simulations

Monte Carlo simulations serve the objective of estimating a stochastic characteristic γ with an estimator $\hat{\gamma}$ computed from observed data [45]. The roots of modern Monte Carlo methods run back to the 1940s with the use of random numbers to examine problems from a stochastic perspective [46]. Monte Carlo methods have proven to be extremely useful in solving problems where analytical solutions do not exist or when conditions required for mathematical theory to be valid do not hold. Monte-Carlo simulations offer an alternative to analytical mathematics for understanding systems behaviour using random samples of known simulated data populations.

The basic steps of a Monte-Carlo simulation are as follows [45]:



The basic Monte-Carlo simulations are therefore very simple. The accuracy of a Monte-Carlo estimate depends upon creation of the pseudo-population in the thoroughness with which it is explored through pseudo-samples. Improvement of the result could be possible simply by increasing the number of trials. Therefore, a major limitation of Monte-Carlo simulations is the limitation of computational resources. This method is used extensively in the next several chapters of this book.

2.4 Summary

This chapter outlined the basic concepts and notions, a literature review of *MFPT* prediction methods and a few mathematical tools used throughout the rest of this book. The main objective of the rest of this book is on discussing new techniques to calculate the transport property introduced as *MFPT*, for random walks under different conditions and situations commonly imposed by processes governing natural disaster activity. The concepts and definitions introduced in this chapter will be used throughout the book. It is encouraged to revisit this chapter and the literature cited herein whenever clarifications are needed.

References

1. Tejedor V (2012) Random walks and first-passage properties. Thesis
2. Noh JD, Rieger H (2004) Random walks on complex networks. Physical review letters 92(11):118, 701
3. The Australian Bureau of Meteorology (2016) Climate data online, (available online at http://www.bom.gov.au/hydro/flood/qld/fldreports/)
4. Gall JS, Ginis I, Lin SJ, Marchok TP, Chen JH (2011) Experimental tropical cyclone prediction using the GFDL 25-km-resolution global atmospheric model. Weather Forecast 26(6):1008–1019
5. Ryan CJ (1993) Costs and benefits of tropical cyclones, severe thunderstorms and bushfires in australia. Clim Change 25(3–4):353–367
6. Wijesundera I, Halgamuge MN, Nirmalathas T, Nanayakkara T (2013) A geographic primitive-based bayesian framework to predict cyclone-induced flooding*. J Hydrometeorol 14(2):505–523
7. Godec A, Metzler R (2015) Optimization and universality of brownian search in quenched heterogeneous media. arXiv preprint arXiv:150300558
8. Isham V, Harden S, Nekovee M (2010) Stochastic epidemics and rumours on finite random networks. Physica A: Stat Mech Appl 389(3):561–576
9. Perra N, Baronchelli A, Mocanu D, Gonçalves B, Pastor-Satorras R, Vespignani A (2012) Random walks and search in time-varying networks. Phys Rev Lett 109(23):238, 701
10. Bonaventura M, Nicosia V, Latora V (2014) Characteristic times of biased random walks on complex networks. Phys Rev E 89(1):012, 803
11. Eggeling C, Ringemann C, Medda R, Schwarzmann G, Sandhoff K, Polyakova S, Belov VN, Hein B, von Middendorff C, Schönle A (2009) Direct observation of the nanoscale dynamics of membrane lipids in a living cell. Nature 457(7233):1159–1162
12. McKenzie HW, Lewis MA, Merrill EH (2009) First passage time analysis of animal movement and insights into the functional response. Bull Math Biol 71(1):107–129
13. Gallos L, Song C, Havlin S, Makse H (2007) Scaling theory of transport in complex biological networks. Proc Natl Acad Sci 104(19):7746
14. Starnini M, Baronchelli A, Barrat A, Pastor-Satorras R (2012) Random walks on temporal networks. Phys Rev E 85(5):056, 115
15. Nicolaides C (2011) Anomalous transport in complex networks. Thesis
16. Condamin S, Benichou O, Tejedor V, Voituriez R, Klafter J (2007b) First-passage times in complex scale-invariant media. Nature 450(7166):77–80

17. Han JDJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430(6995):88–93
18. Aldous D, Ganesan K (2013) True scale-invariant random spatial networks. Proc Natl Acad Sci 110(22):8782–8785
19. Lee ZQ, Hsu WJ, Lin M (2014) Estimating mean first passage time of biased random walks with short relaxation time on complex networks. PLoS One 9(4):e93–348, doi:10.1371/journal.pone.0093348
20. Albert R, Jeong H, Barabási AL (1999) Internet: diameter of the world-wide web. Nature 401(6749):130–131
21. Carreras BA, Lynch VE, Dobson I, Newman DE (2002) Critical points and transitions in an electric power transmission model for cascading failure blackouts. Chaos: Interdiscip J Nonlinear Sci 12(4):985–994
22. Redner S (2001) A guide to first-passage processes. Cambridge University Press, Cambridge
23. Sood V, Redner S, Ben-Avraham D (2005) First-passage properties of the Erdös-Renyi random graph. J Phys A: Math Gen 38(1):109
24. Condamin S, Benichou O, Moreau M (2007) Random walks and brownian motion: A method of computation for first-passage times and related quantities in confined geometries. Phys Rev E 75(2):021, 111
25. Aldous D, Fill J (2002) Reversible markov chains and random walks on graphs
26. Weiss GH (1994) Aspects and applications of the random walk. Random materials and processes, Amsterdam, The Netherlands. (New York: North-Holland, 1994)
27. Douglas JF (1995) Aspects and applications of the random walk. J Stat Phys 79(1):497–500
28. Mejía-Monasterio C, Oshanin G, Schehr G (2011) First passages for a search by a swarm of independent random searchers. J Stat Mech: Theory Exper 2011(06):P06, 022
29. de Jager M, Weissing FJ, Herman PM, Nolet BA, van de Koppel J (2011) Lévy walks evolve through interaction between movement and environmental complexity. Science 332(6037):1551–1553
30. Song C, Havlin S, Makse HA (2006) Origins of fractality in the growth of complex networks. Nat Phys 2(4):275–281
31. Song C, Havlin S, Makse H (2005) Self-similarity of complex networks. Nature 433(7024):392–395
32. Lau HW, Szeto KY (2010) Asymptotic analysis of first passage time in complex networks. EPL (Europhysics Letters) 90(4):40, 005
33. Haynes CP, Roberts AP (2009) Generalization of the fractal einstein law relating conduction and diffusion on networks. Phys Rev Lett 103(2):020, 601
34. Byl K, Tedrake R (2009) Metastable walking machines. Int J Robot Res 28(8):1040–1064. doi:10.1177/0278364909340446
35. Skarpalezos L, Kittas A, Argyrakis P, Cohen R, Havlin S (2014) Efficiency of message transmission using biased random walks in complex networks in the presence of traps. arXiv preprint arXiv:14062437
36. Redner S, Majid I (1983) Critical properties of directed self-avoiding walks. J Phys A: Math Gen 16(9):L307
37. Ben-Avraham D, Havlin S (2000) Diffusion and reactions in fractals and disordered systems. Cambridge University Press, Cambridge
38. Dingle H (2006) Animal migration: is there a common migratory syndrome? J Ornithol 147(2):212–220
39. Anderson HE (1983) Predicting wind-driven wild land fire size and shape. US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station
40. Hwang S, Lee DS, Kahng B (2012) First passage time for random walks in heterogeneous networks. Phys Rev Lett 109(8):088, 701
41. Shlesinger MF (2007) Mathematical physics: first encounters. Nature 450(7166):40–41
42. Doya K (2007) Bayesian brain: Probabilistic approaches to neural coding. MIT press, Cambridge

43. Dynkin EB, Kováry T (2014) Theory of markov processes. (electronic resource). (Burlington: Elsevier Science, 2014)
44. Ching WK, Ng MK (2006) Markov chains. (electronic resource): models, algorithms and applications. International series in operations research and management science: 83, New York: Springer, c2006
45. Mooney CZ (1997) Monte carlo simulation, vol 116. Sage Publications, New York
46. Landau DP, Binder K (2014) A guide to Monte Carlo simulations in statistical physics. Cambridge university press, Cambridge

Chapter 3
Predicting Cyclone Induced Flood: A Comprehensive Case Study

Abstract The prediction of *MFPT* in the propagation of cyclone induced flood is comprehensively discussed in this chapter as a special case study to gain an entrance to a more generic study in the proceeding chapters. The inhomogeneity in terrain is addressed using a novel concept of dividing the terrain into geographic primitives (GPs) identified through flood propagation vector fields. A modularised approach including a Bayesian framework is followed allowing easy and flexible assimilation of data and integration of outputs from other models when they become available. Simultaneous events distributed in the environment leading to biased propagation are encapsulated into transition probability matrices (TPM) allowing efficient real time predictions via computation through a Markov chain. Comparative studies using real datasets show the models ability of predicting up to 3 h ahead of official forecasts with a 33 % improvement of accuracy when compared to other models currently used.

3.1 Introduction

Flooding induced by cyclone activity is a result of a combination of several simultaneous and distributed random processes in the environment [1, 2]. The effectiveness of managing cyclone induced flood is highly dependent on how fast reasonably accurate predictions can be made, which is a particularly difficult task given the multitude of highly variable physical factors affecting. Even with supercomputers, collecting and processing vast amounts of data from numerous asynchronous sources makes it challenging to achieve high prediction efficiency.

The highly complex and distributed dynamic system that results in cyclone induced flooding events can be roughly divided into three major groups as processes leading to the spatial motion of cyclones [3], cyclone induced rainfall [4], and propagation of flood [5]. Both common network traits introduced in the preceding chapters; network specific bias and network inhomogeneity, are present in all three processes which show different propagation mechanisms in their respective domains. For this reason, this chapter discusses this specific application to gain an entrance

I. Wijesundera et al., *Natural Disasters, When Will They Reach Me?*,
Springer Natural Hazards, DOI 10.1007/978-981-10-1113-9_3

addressing the effects of bias and network inhomogeneity in a more generic approach of analysing MFPT for random walk processes governing the dynamics of natural disasters.

3.2 Cyclone Activity: One of the Major Forms of Natural Disasters

The past few decades has shown a great increase in the vulnerability to natural disasters all over the world, leading to social, economic and environmental tragedies. The tragic flooding in 2010–2011 in the state of Queensland, Australia, alone has resulted in over 200,000 people affected, with 35 confirmed deaths as stated by [6] (2011) and [7] (2011). The resulting damage was over AUD 1 billion with an estimated reduction of AUD 30 billion in Gross Domestic Product (GDP). According to the "Intergovernmental Panel on Climate Change", economic losses from weather and climate-related disasters have increased in the past few decades and heavy rainfalls associated with Tropical Cyclones (TC) are likely to increase with continued warming [8]."

To add to the increase in the importance of disaster monitoring, there has been some decline in monitoring infrastructure, as a result of disasters themselves [9]. Millions of dollars are being spent on disaster management sensor networks signifying the importance of efficient sensor deployment and management for which predictions of disaster dynamics are of high importance. With such knowledge, one method of extending the lifetime of a wireless sensor network is to optimise the on-board energy consumption of nodes. In order to conserve on-board energy, many design approaches have been researched on, such as network architecture, efficient sensing circuitry, algorithms and communication protocols [10, 11]. Various dynamic power management techniques have also been proposed which mainly address sleeping patterns and idle states dynamically [12]. In addition to power management in sensor nodes, efficiency and the effectiveness of a network can be improved drastically using Just-In-Time sensor deployment as described in [13].

An expected time-to-flood for a given area prone to cyclone induced flooding will immensely improve the effectiveness of disaster management operations. This includes for power management techniques and dynamic deployment of sensor networks, which would help obtaining a perfect balance between reducing network cost and capturing the most important data. It has been shown in [14] that the losses at a catastrophic event such as a natural disaster, decreases with the increase of predictability of the event.

3.2.1 Numeric Flood Prediction Models

There are several deterministic flood models in use today predicting likely inundations resulting from TC activity. Digital Elevation Model (DEM) of increasing

resolution is used to model water flow [9, 15, 16], taking into consideration that flood inundation extent is highly dependent on topography [15]. It has been shown in [15] that the predictive ability deteriorates with the decrease in resolution of DEM data used. This is especially true when levee structures are smoothed with lower resolution landscape. However, it has also been shown that the improvement of prediction accuracy with increased resolution is marginal for smoother landscape. Therefore, one drawback in this approach is the fact that it uses high resolution data throughout the landscape which results in increased computational cost. This raises the importance of using different resolutions based on the geography of the area.

3.2.2 The Concept of Geographic Primitives (GPs)

Dividing the landscape into primitives identified by water-flow characteristics has been presented as a viable solution to address the geography-optimised prediction of flood propagation [17].

Geographic Primitives (*GP*s)
A *GP* can be defined as a portion of landscape where the main driving force of a disaster shows statistically stereotypical behaviour (e.g. distinguishable water flow patterns can define *GP*s in the process of flood prediction, where a few identifiable *GP*s would be basins, mountain ranges, valley's, flat land, etc.).

Identifying *GP*s is done through visual inspection of contour patterns that would give low modal vector fields with patterns typical to the GP type. For flood prediction, it is the simplest unit that can be identified in terms of water flow patterns without being biased by any structural or agricultural aspect where the only focus is to identify simple low modal vector fields of water flow. The dominant factor forming the vector field is the contour patterns of the catchment. With the increasing availability of high resolution topography data, identifying *GP*s in this simplest form is feasible irrespective of availability of any other data. But if more data is available, such as of land use and Pedology, vector fields can be formed considering effects of all these factors. It's worth noting that complete and accurate datasets are needed if other factors are to be considered, as the accuracy of the prediction is decided by the quality of data used.

At this point a manual partitioning approach is discussed due to the complexity of the global vector field. Stereotypical behaviour of vector fields such as converging, diverging, directing, and scattering are visually identified on the flood vector field to demarcate *GP*s such as valleys, mountain crests, mountain ranges and flat land. While low-modal vector fields are recommended, uni-model vector fields would be the ideal case for a *GP*.

Literature gives many distribution models that aim to overcome the burden of spatial representation [18]. These can be divided mainly into two groups; one in which model elements are selected entirely upon topography data (e.g. gridded elements, contours and streamlines, triangulated irregular network facets and conceptual elements of hill-slopes and stream segments), and the other where layers of data are overlay in order to identify particular combinations of elevation, land cover, slope, aspect, soils, etc. (e.g. Hydrologic Response Units (HRU), Representative Elementary Columns (REC), and hydro-landscape units [19]). Model elements identified by any of these methods could pass as *GP*s if their flood vector fields show statistically tractable behaviour. In some cases there may be multiple identifiable *GP*s within one modal element or vice-versa in some others. The main difference between the definition of *GP*s and above mentioned hydrological model elements are that *GP*s only take vector field of surface water flow as the variable and not put any abstract label on the primitives.

3.2.3 Data Assimilation in Flood Prediction Models

Models have also been developed to incorporate data assimilation [16, 20]. In [16], the authors have addressed the problem of obtaining a topographically optimum model to improve the representation of "raw" topographic data so that its integration with lower-resolution numerical inundation models is optimal. The authors of [20] integrates flood model outputs with the Geographical Information System (GIS). A model currently in use for stream flow monitoring is the Geospatial Stream Flow Model (GeoSFM), which is a semi-distributed hydrological model developed as an extension of the ArcView GIS software [21]. GeoSFM software uses a wide range of inputs, including satellite rainfall estimates, soil data, land cover and elevation data, to predict stream flow. One shortcoming of this model is its inability to predict absolute flow magnitudes due to the absence of regional and seasonal bias correction [5] which is a difficult task with data limitations when trying to use a generic model for diverse regions. The GeoSFM model only finds abnormalities in water flow. One other limitation is that the output is numerical and the probability aspects are not present in the output.

The concept of *GP*s also address the large file size problem for high resolution topography. This division could be a regular grid or an irregular division as suggested in the deployment of disaster management wireless sensor networks in [22]. Figure 3.1 shows some identified geographic primitives on the landscape of Queensland, leading to some stereotypical flood distributions including scattering, converging and diverging. As topography will not always give clearly distinguishable primitives, catchments need not have strict boundaries and using overlapping boundaries is more appropriate. It should also be noted that the runoff between certain *GP*s would show stereotypical behaviour.

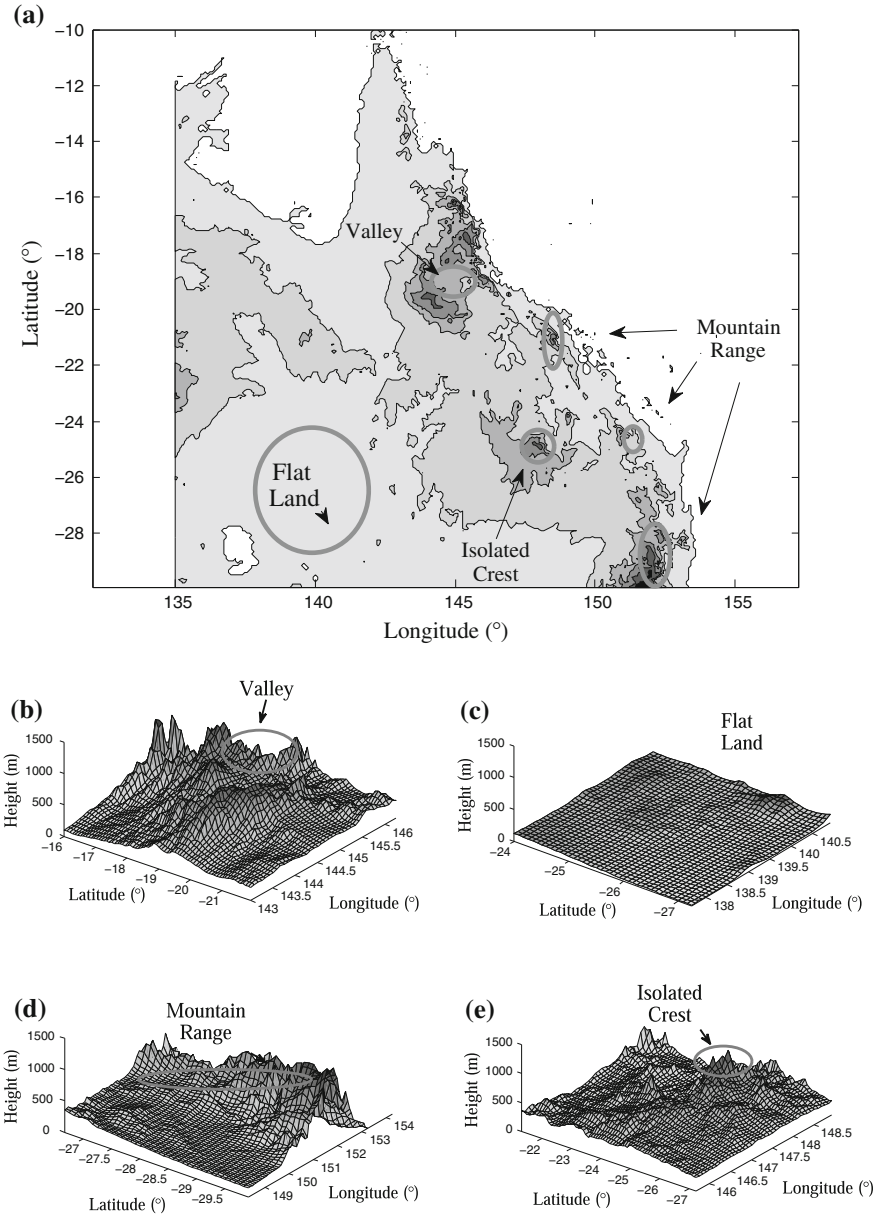


Fig. 3.1 Identifying geographic primitives which define the accuracy of prediction. **a** Some *GP*'s on the landscape of Queensland. **b** A Valley. **c** A flat land. **d** A mountaion range. **e** An isolated crest

3.2.4 Predicting Cyclone Induced Rainfall

The lead time for TC induced flood prediction can be increased by predicting TC induced rainfall which results in floods. Although there has been great advancements in numerical prediction models in the past decade [23–25], these models still have room for improvement due to the multitude of factors affecting TC induced rainfall and getting data on all these factors in real time is not practical while a probabilistic approach would be more viable.

3.2.5 Bayesian Theory for Natural Disaster Prediction

Bayesian probability approach has been increasingly used in natural disaster prediction (in [14, 26, 27]) due to the many advantages and flexibility associated with it. One main advantage in the approach is that it incorporates prior knowledge, pragmatically optimised by the user, which allows for probabilistic predictions as opposed to binary true/ false outcomes, which have a risk of misleading forecasts. In addition, this approach accounts for parameter uncertainty, reducing error from over-fitting of training data, and provides natural interpretation of regularisation [28]. Therefore, given the large amount of factors affecting cyclone induced rainfall characteristics [4], possibility of errors in observed data, and the ability to use prior knowledge in prediction [29], using a Bayesian framework can be identified as a suitable candidate for TC induced rainfall prediction.

3.2.6 Predicting the Spatial Motion of Cyclones

Including the prediction of the path of a cyclone could further increase the lead time of flood prediction by the model. Cyclone track prediction models have come a long way since the use of purely statistical models such as the CLIPER (CLImatology and PERsistence) model, proposed by [30], which is now used solely as a benchmark for assessing the skill of other models. Dynamic models, which use numerical weather prediction, are used most widely at present. These models generally require supercomputers to solve the mathematical equations governing the physics of the atmosphere and use numerical methods to solve these equations in order to generate forward-in-time forecasts of the track of the cyclone [31]. The Geophysical Fluid Dynamics Laboratory (GFDL) model [32] is one of the dynamical models used most widely. The dynamic models use a large range of data sources for assimilation including, satellite data, specialised aircraft data, and local area sensor networks. Gall et.al (2011) [33] states that GFDL is only a regional model as well a “Late” model where

the first prediction is only available 4–6 h after the initial track advisory is released, despite being high in accuracy. The output of this model could be used for the flood prediction model considering these limitations as well.

3.3 A Comprehensive Model for Cyclone Induced Flood Prediction

An ideal flood prediction would be available with high lead times using minimal available data and would be refined with any subsequently available data. This section presents a comprehensive and modularised prediction approach facilitating easy data assimilation.

Consider a geographical area prone to TCs, with a height profile $h(x_1, x_2)$, where x_1 and x_2 describe the location in terms of latitudinal and longitudinal coordinates, respectively. Prediction starts with available minimal data of the cyclone while it is refined with incoming data. Let $t = T, 2T, 3T, \ldots, kT$ be prediction times for k number of periodic predictions at T time intervals. Let the geographical area consist of many geographical primitives (GP) where the propagation of flood shows statistically stereotypical behaviour (e.g. basins, mountain ranges, valley's, flat land, etc.). Let Π be the set of node clusters with data currently available, and $\bar{\Pi}$ be those which are yet to sense any data. Prediction of failure time (i.e. time to flood) for node i with data, $FT_{i \in \Pi}$, as well as for nodes in $\bar{\Pi}$ start as soon as cyclone data is sensed anywhere in the total network. This data is used initially for cyclone path prediction, of which the output is used in the rainfall density distribution prediction phase and finally into flood density distribution prediction as explained in Fig. 3.2.

All data as well as predictions from more sophisticated models at each stage (cyclone path and rainfall) are used in refining the prediction as data becomes available. The failure time for the nodes in clusters ($i \in \bar{\Pi}$) can be inferred from available data ($i \in \Pi$). This estimate can be recalculated as more data is obtained, and the FT_i estimate at a given time can be used to deploy and manage the ith node cluster. This is possible even when there is no current available data at any of the nodes (i.e. $\Pi \in \oslash$) where the prior is mainly decided by historical data and mathematical simulations of water flow. The summary of the main symbols used in describing the model are listed in Table 3.1.

The modularised calculation flow is given in Fig. 3.2. Prediction starts with incoming cyclone best track data, which are used for cyclone path prediction to get a linear approximation of the path using a sliding window of samples where the window size and deviation depends on the speed of the eye of the cyclone. The output is used as an input to the rainfall prediction phase. Probability distribution of rainfall is predicted using Bayesian learning, with the initial prior taken from the R-CLIPER model and a likelihood function generated using the available rainfall data. The rainfall prediction is used for the next stage for flood prediction by considering the dynamics of water flow on a landscape approximated by a linear combination of Gaussian functions.

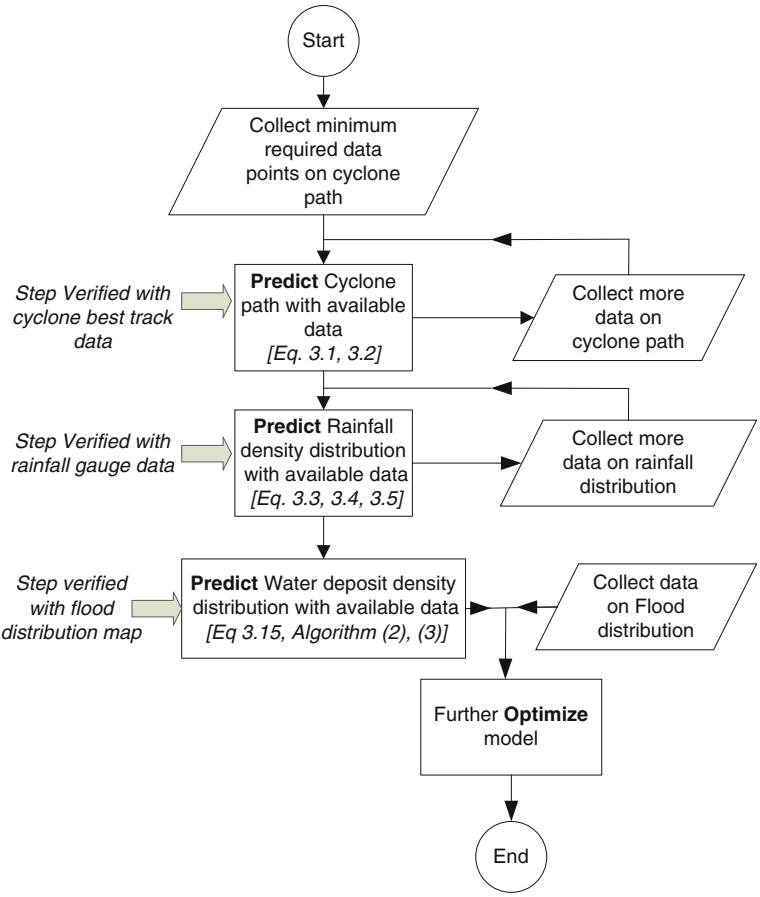


Fig. 3.2 Flowchart of the prediction model

The area is segmented into geographic primitives identified by water flow patterns during the learning phase and the calculation of time-to-failure is made more efficient with the use of probability transition matrices for each segment.

This model is developed on a few assumptions, mainly for simplicity and to focus more on the new concepts introduced herein. The main assumptions can be listed as (a) the terrain is frictionless (b) that soil absorption is negligible (c) evapotranspiration during prediction time is negligible (d) topography of the area remains unchanged during the time of prediction. Although the first three effects are out of the scope at this point, they can be readily integrated into the model in the motion equations and probability transition matrices, which are described in detail later in the chapter, and thereby improve on the accuracy. Assumption (d) is the basis of the proposed off-line calculations.

Table 3.1 Parameters used

Symbol	Definition	Units
T	Prediction time interval	s
h	Height profile	m
v	Velocity	m s^{-1}
x_1	Distance in latitudinal direction	°
x_2	Distance in longitudinal direction	°
GP	Geographical primitive	–
Π	Node clusters with data	–
$FT_{i\in\Pi}$	Failure time for ith cluster in Π	s
$FT_{i\in\Pi'}$	Failure time for ith cluster in $\bar{\Pi}$	s

Method assumptions

- The terrain (through which the water flows) is frictionless.
- Soil absorption is negligible during floods.
- Evapotranspiration during the calculation time is negligible.
- Topography of the area remains unchanged during the time of prediction. (This is the basis behind the concept of *geographic primitives* introduced in this chapter.)
- Rainfall during one prediction time period is limited to an area of 500 km radius. (This is true for most scenarios but neglects rainfall due to convergence along coasts during extratropical transition.)

3.3.1 Step 1: Predict Cyclone Path with Available Data

The future track of the cyclone, using any available data, is done in this step using a curve fitting technique. After comparative studies with many standard curves, including exponential, higher order polynomials and Gaussian, the best prediction has been identified with fitting a first order polynomial with track data available at each step using a windowing technique.

The aim of this step is to get the best prediction of the future track of the cyclone using best-track data available at the time of prediction. A curve fitting approach is taken using standard curves commonly used for such applications. The predicted track output from all these curves are compared to actual track taken after the time of prediction. The equations governing the four standard curves used for the comparison are given in Table 3.2.

Table 3.2 Standard curves used for curve fitting

Function	Equation
Linear polynomial	$f(x) = ax + b$
Exponential	$f(x) = ae^{(bx)}$
Gaussian	$f(x) = ae^{(-((x-b)/c)^2)}$
Higher order (cubic) polynomial	$f(x) = ax^3 + bx^2 + cx + d$

Figure 3.3 shows a graphical comparison of four curve fitting outputs using these four standard curves. The lines show the curve fitting using linear polynomial, exponential, Gaussian and higher order (3rd) polynomial functions respectively. The three columns show comparison at three sample prediction points. Initially the higher order polynomials (last row) are ruled out from the comparison since although it provides a good fit for existing points, the prediction was deviated away from the track with the effect of the higher order terms. For Gaussian curves (third row), even though the error is less than the earlier case, high bias towards the origin of coordinates made prediction less accurate for most cases (Fig. 3.3h, i). Deciding between the linear polynomial and exponential curves are a tough choice as both curves produce good results in most cases. But when there is a slight change of direction of a curved track, using a linear polynomial proved to be a better predictor as shown in Fig. 3.3c compared to Fig. 3.3f. Therefore, considering the fact that using linear polynomial would eliminate any errors arising from bias, it is chosen for use in the cyclone track prediction step.

This selection is mainly due to non existence of any bias to the origin of the fit. The first order polynomial used can be expressed as

$$x_{1t} = a + bx_{2t}, \tag{3.1}$$

where x_{1t} is the latitude and x_{2t} is the longitude at time t, b is the slope and a is the intercept of the line. The unknown coefficients a and b are computed through linear regression using least squares fitting technique on past track data. The sum of the squares of the deviations of the linear approximation (R^2) of a set of n data points can be expressed as

$$R^2(a, b) = \sum_{i=1}^{n} [x_{1i} - (a + bx_{2i})]^2$$

and the regression coefficients can be found by solving

$$\begin{aligned} \frac{\partial[R^2]}{\partial a} &= -2 \times \sum_{i=1}^{n} [x_{1i} - (a + bx_{2i})] = 0 \text{ and} \\ \frac{\partial[R^2]}{\partial b} &= -2 \times \sum_{i=1}^{n} [x_{1i} - (a + bx_{2i})] \times x_{2i} = 0 \end{aligned} \tag{3.2}$$

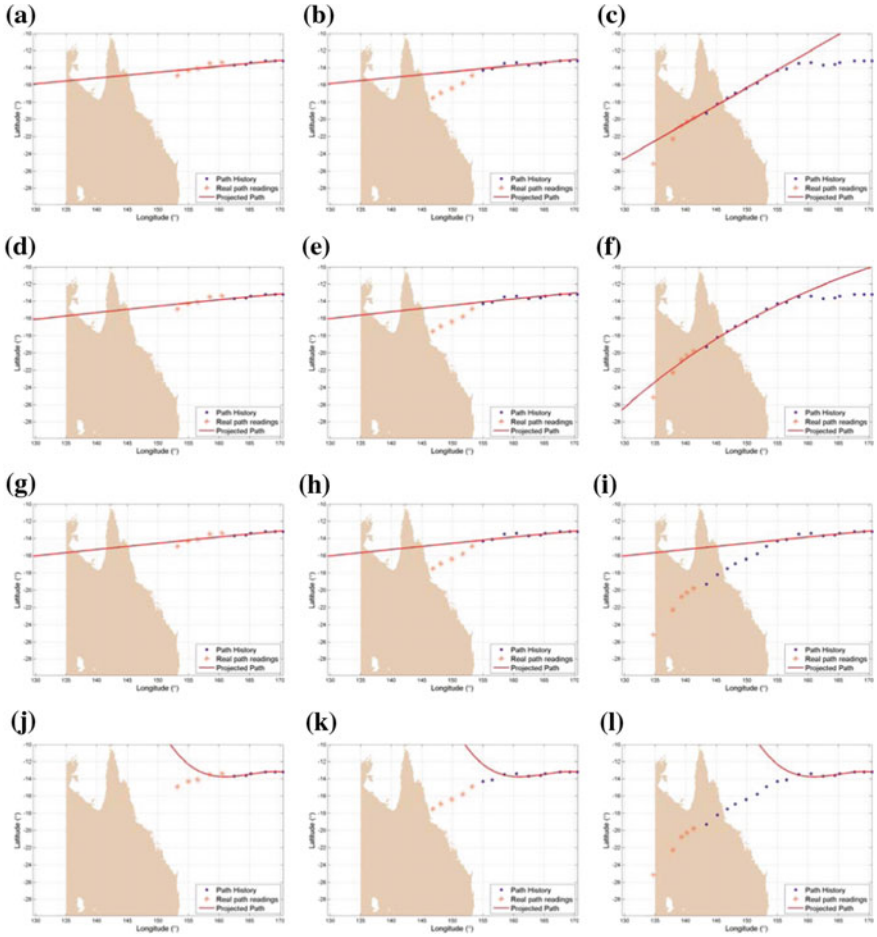


Fig. 3.3 Four common standard functions compared that are used for the prediction of future track followed by a cyclone through curve fitting technique. Three prediction steps are compared in the three Columns. The rows show predictions using linear polynomial, exponential, Gaussian and higher order polynomial functions respectively

The central track of the predicted motion of the cyclone (before adding the allowable deviation) is calculated using Eq. (3.1) and $x_{1t} = (x_{1p} - x_{1(p-1)})t/(\delta t)$ where p is the position of the cyclone according to latest data, and δt is the time interval between readings. The probability distribution of the cyclone path is iteratively predicted with incoming data using a sliding window technique using a window size n, which

reduces as speed of the travel of the eye of the cyclone decreases. This is because a cyclone slows down when it turns. The calculation algorithm is as described in Algorithm 1.

Algorithm 1 Algorithm used for cyclone path prediction

Step 1: Set $i = 0$;
Step 2: Collect ith location data input until $i = n$, where n is the default size of the moving window used for calculations;
Step 3: Calculate the current speed of the cyclone and adjust window size for **Step 4** accordingly (a default step size of 5 is used in the calculation which reduces to 3 when two consecutive readings show 30 % below average speed and further reduces to 2 when three or more closest readings show 30 % below average speed;
Step 4: Use linear regression from Eq. (3.2) to get a and b;
Step 5: Calculate upper and lower margins of the distribution using the standard deviation weighted by speed of the travel of the eye of the cyclone and the R^2 value of fit;
Step 6: Continue at **Step 3** for each received location data point.

The predicted probability distribution of cyclone path, refined as data arrives, is used as an input to the calculation of rainfall probability distribution.

3.3.2 Step 2: Predict Rainfall Density Distribution with Available Data

Using the knowledge of the current and (predicted) future track of the eye of a cyclone, a prediction of rainfall density distribution can be made. As the cyclone path prediction gets refined with incoming data, the rainfall prediction also changes accordingly. Probability density distributions of locality of rainfall are calculated through Bayes filtering technique, where the prediction (i.e. posterior) for one time step is used as a prior for the next prediction step. The rainfall probability density distribution is calculated as

$$prediction = likelihood * priori. \tag{3.3}$$

Division by the probability of data, as used in conventional Bayes theorem is not used here as it is just a scaling factor. The predicted distribution is normalised to get the probability density distribution of rainfall locality where the integral of the distribution over the considered area equals one. In the initial prediction, where there is no available data of rainfall, the priori estimate is taken as the output from the R-CLIPER model [34].

The R-CLIPER model (Rainfall CLImatology and PERsistence Model)
For tropical cyclone rainfall, the R-CLIPER model was developed for predictions based upon climatology and persistence. As explained in [34], in the R-CLIPER model, a climatological rainfall rate is determined and then integrated along the storm track. Because the primary interest in tropical cyclone rainfall is over land, the variation in rainfall rate after landfall has been taken into account. Asymmetries are not taken into account in the R-CLIPER model because most significant factor for this asymmetry appears to be the storm response to the environmental vertical wind shear which depends on the particular synoptic environment. This is a major limitation of the model.

The equation governing the model is given as

$$R(r,t) = [a\exp(-\alpha t) + b]\exp(-(r - r_m)/r_e), \tag{3.4}$$

where parameters a and b are defined from the fit to the gauge data by radius, and r_m is the radius of maximum rainfall (which is at the origin) and r_e is the radial distance to the edge of the distribution which is usually taken as 500 km.

The major advantage to the R-CLIPER is the simplicity afforded by its use of the forecast storm intensity as an additional factor. One issue with this method is that the algorithm slightly underestimated the rain estimates because it assigned the values derived for the three intensity ranges to the middle of each intensity range. The storm intensity statistics suggest that the rain estimates are more representative of the lower edge of each intensity range.

The rainfall climatology and persistence (R-CLIPER) model is widely used as a benchmark for cyclone induced rainfall prediction.

3.3.2.1 Finding Likelihood Function for Rainfall Prediction

Several approaches are compared to get a suitable likelihood function from available rainfall data at time of prediction. Looking at the rainfall distribution, first it was attempted to fit the rainfall distribution with distance to an inverted Mexican hat distribution (Fig. 3.4) and curve fitting for several positions; are plotted in Fig. 3.5. However, this did not prove to be a very effective likelihood function as there exists a lot of asymmetries in the rainfall distribution resulting from cyclone activity. The equations governing the surface fitting functions used in this section are summarised in Table 3.3.

The curve fitting by combinations of Gaussians for the same 8 cyclone eye positions are shown in Fig. 3.15. The fitting is done first by smoothing the readings with a moving average and then by curve fitting. The comparison of rainfall predictions of the presented model with actual rain and the R-CLIPER model are plotted in Figs. 3.6 and 3.7 and a comparison of the results are given in Table 3.4.

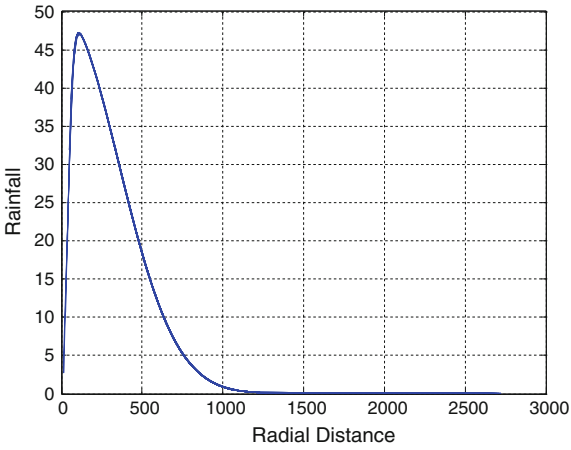


Fig. 3.4 Initial approximation of rainfall using an inverted Mexican hat distribution

As soon as data is received, a likelihood function is computed, which is used in prediction of rainfall after a T time period. Many models are attempted in obtaining a suitable likelihood function which would produce a more accurate prediction as the posterior using Eq. (3.3). A surface-fitting technique is used on available rainfall data with common 3D functions. First attempt is to approximate the available rain data to a 3D Gaussian function without orientation. Then a sum of two Gaussians is used to replicate an inverted Mexican hat surface which captures the low rainfall usually occurring within the eye of the cyclone (i.e. within 50 km radius from the centre of the cyclone). Finally, comparative studies on resulting posterior distributions showed that using a bivariate Gaussian function with orientation used as the likelihood function produced the best predictions. This function can be expressed as

$$\begin{aligned} f_l &= f(x_1, x_2, r, \theta, \sigma) \\ &= r \exp\left[-\Big(a(x_1 - x_{1o})^2 + 2b(x_1 - x_{1o})(x_2 - x_{2o}) + c(x_2 - x_{2o})^2\Big)\right] \end{aligned} \tag{3.5}$$

where f_l = likelihood function, $a = cos^2\theta/2\sigma_x^2 + sin^2\theta/2\sigma_y^2$, $b = -sin2\theta/4\sigma_x^2 + sin2\theta/4\sigma_y^2$, $c = sin^2\theta/2\sigma_x^2 + cos^2\theta/2\sigma_y^2$, θ = angle, σ = width and r = maximum rainfall. The distributions on all cases are limited to 500 km (i.e. $\sigma = 500$ km) around the centre of the cyclone (x_{1o}, x_{2o}) . The final rainfall probability distribution obtained under this stage is used as an input to the flood prediction stage.

3.3.3 *Step 3: Predict Water Deposit Density Distribution with Available Data*

The next step in the presented model is to calculate flood distribution through water flow calculations using the predicted rainfall distribution as input.

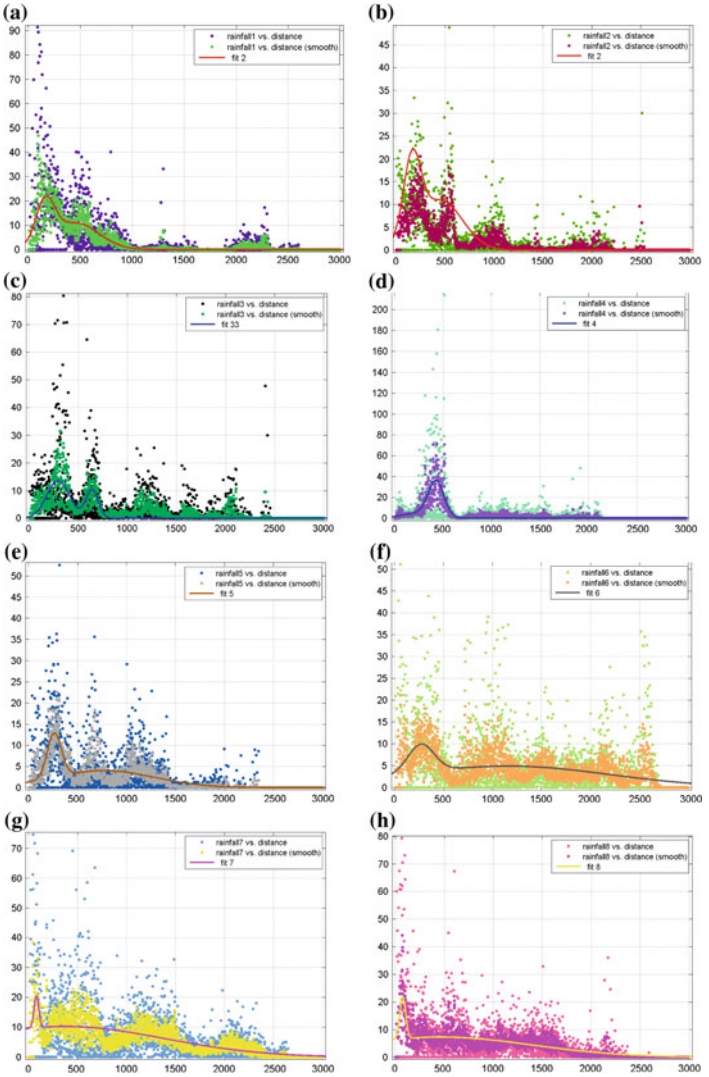


Fig. 3.5 Curve fitting for real rainfall data with a 'Inverted Mexican Hat' type function with respect to distance for Cyclone Yasi. **a** Cyclone eye position 1 at (Latitude $= -19.3$, Longitude $= 123.4$) on 2011-02-03-0000. **b** Cyclone eye position 2 at (Latitude $= -19.8$, Longitude $= 121.3$) on 2011-02-03-0600. **c** Cyclone eye position 3 at (Latitude $= -20.3$, Longitude $= 120.2$) on 2011-02-03-1200. **d** Cyclone eye position 4 at (Latitude $= -20.8$, Longitude $= 139.3$) on 2011-02-03-1800. **e** Cyclone eye position 5 at (Latitude $= -22.3$, Longitude $= 137.9$) on 2011-02-04-0600. **f** Cyclone eye position 6 at (Latitude $= -25.2$, Longitude $= 134.7$) on 2011-02-05-0600. **g** Cyclone eye position 7 at (Latitude $= -25.4$, Longitude $= 135.4$) on 2011-02-05-1200. **h** Cyclone eye position 8 at (Latitude $= -25.6$, Longitude $= 135.3$) on 2011-02-05-1800

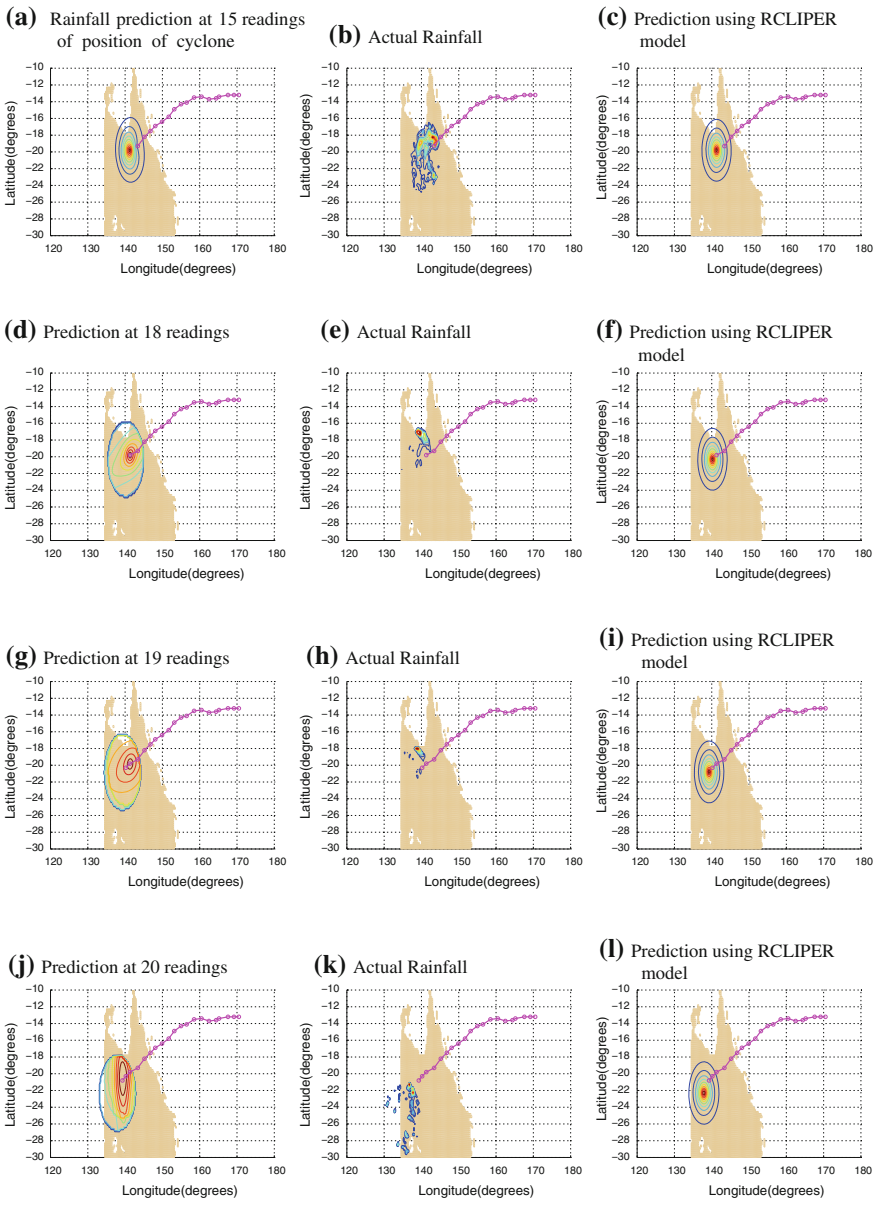


Fig. 3.6 Comparison of rainfall prediction using the presented model and prediction using the RCLIPER model compared to actual rainfall of cyclone 'Yasi' [6]

Table 3.3 Standard functions used for surface fitting

Function	Equation	Notes
Inverted mexican hat	$f(x) = a_1 e^{(-((x-b_1)/c_1)^2)} + a_2 e^{(-((x-b_2)/c_2)^2}$	
3D Gaussian	$f(x, y) = ae^{(-((x-b1)/c1)^2)} e^{(-((y-b2)/c2)^2)}$	
R-CLIPER	$f(r, t) = [ae^{-\alpha t} + b]e^{-(r-r_m)/r_e}$	From Eq. (3.1)
3D Gaussian with orientation	$f(x_1, x_2) = re^{-\left(a(x_1-x_{1o})^2+2b(x_1-x_{1o})(x_2-x_{2o})+c(x_2-x_{2o})^2\right)}$	Equation (3.5) of main text

Table 3.4 Comparison of Prediction with R-CLIPER model: The distance of the centroid of the mean contours of the prediction to that of the actual rainfall is compared. (The predictions are in km)

Cyclone position	Prediction	RCLIPER prediction
1	293	297
2	340	332
3	456	447
4	485	491
5	185	363
6	262	288
7	299	370

3.3.3.1 The Dynamics of Water-Flow

In the presented model, water flow calculations are carried out in latitudinal and longitudinal directions. In this model, it is assumed that the rainwater would go down a slope at a constant velocity between two sample times. The algorithm used for calculating water flow for one direction in one sub portion of a time period is given in Algorithm 2.

A hypothetical simulation is presented with an area of a hilly geographical primitive considered. The objective of the simulation is estimating the mean-first-passage-time for flooding after a particular hurricane, with given rain probability distribution function, hits the area. The rainfall distribution prediction is used as an input for this and is refined with actual rainfall data as they become available.

In the first stage of simulation, a 2D vertical cross section of an imaginary landscape is shown in Fig. 3.8a. The hills are simulated using the addition of two Gaussian functions.

$$h(x) = h_1 \times \frac{1}{\sqrt{\pi\sigma_1^2}} \exp\left\{-\frac{x-\mu_1^2}{2\sigma_1^2}\right\} + h_2 \times \frac{1}{\sqrt{\pi\sigma_2^2}} \exp\left\{-\frac{x-\mu_2^2}{2\sigma_2^2}\right\} \tag{3.6}$$

where h_i is the height, σ_i is the standard deviation, μ_i is the mean and $i = 1, 2$.

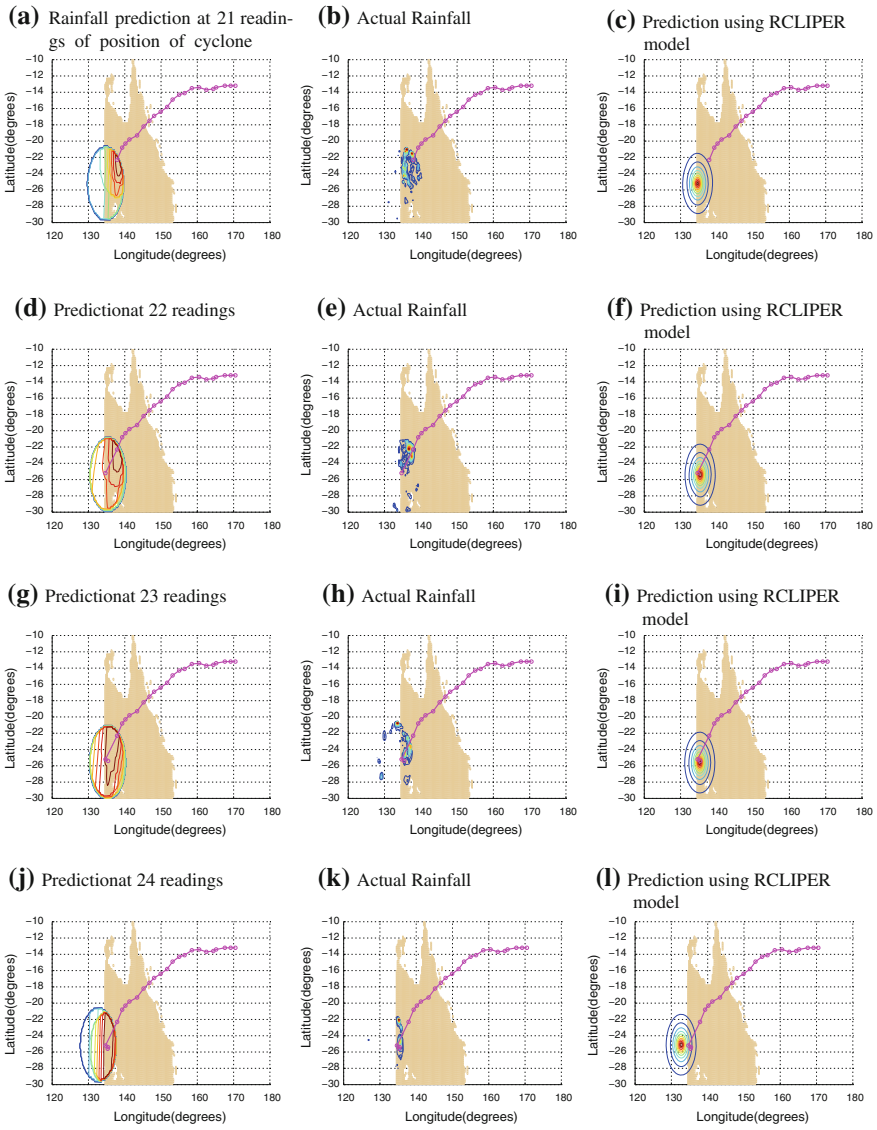


Fig. 3.7 Cont. Fig. 3.6; Comparison of rainfall prediction using the presented model and prediction using the RCLIPER model compared to actual rainfall of cyclone 'Yasi' [6]

The first derivative of this function with respect to horizontal axis is then calculated to get the slope at each point of the simulated landscape. This is later used to calculate water flow. The probability distribution of rain is taken as a having a particular Gaussian distribution. Here, it is assumed that this is prior knowledge (i.e. we take the probability distribution of rain as our prior). Figure 3.8a shows the artificially

Algorithm 2 Algorithm used in calculating water flow. Here, the T time period is broken down to m number of iterations for increased accuracy.

for time $\in \{t_1, \ldots, t_m\}$ **do**

  if |Current Slope - Last Slope| > |Current Slope| &
    |Current Slope - Last Slope| > |Last Slope| **then**
    Accumulate
  end if

  if Slope at x <0 **then**
    $x = x$ + velocity $\times \frac{T}{m} \times$ cos (arctan ($slope$))
  end if

  if Slope at x >0 **then**
    $x = x$ - velocity $\times \frac{T}{m} \times$ cos (arctan ($slope$))
  end if
end for

simulated 2D landscape with the simulated rainfall distribution superimposed. Then the time-to-flood is calculated at the valley using the rainfall probability distribution and dynamics of water flow in the given landscape. Figure 3.8b shows the water flow and accumulation with respect to time. It can be observed that as time goes to infinity all the rainwater will flow into a valley in the given scenario.

As the next step, the water flow equations for 3D landscape are presented with synthetic data before moving on to real datasets. The simulation results are illustrated in Fig. 3.9 for a virtual 3D landscape having a virtual Gaussian rainfall distribution. It is observed that a part of rainfall would accumulate into the valley towards the centre of the landscape and the rest would flow out of the considered portion of land through three river valleys. Therefore, a similar calculation on a real landscape would assist in the identification of locations in the area prone to flooding and landslides.

In order to calculate the two dimensional displacement, partial derivatives of the landscape height profile h, on two perpendicular horizontal axes, $\partial h/\partial x_1$ and $\partial h/\partial x_2$, are used to get the slopes for calculating the water flow. The volume of water initially at (x_{1s}, x_{2s}) would end up at (x_1, x_2) after a time period of T. The displacement equations in these directions are given by,

$$\begin{aligned} x_1 &= x_{1s} - vT \cos\left(\arctan\left\{\frac{\partial h}{\partial x_1}\right\}\right) \\ x_2 &= x_{2s} - vT \cos\left(\arctan\left\{\frac{\partial h}{\partial x_2}\right\}\right) \end{aligned} \tag{3.7}$$

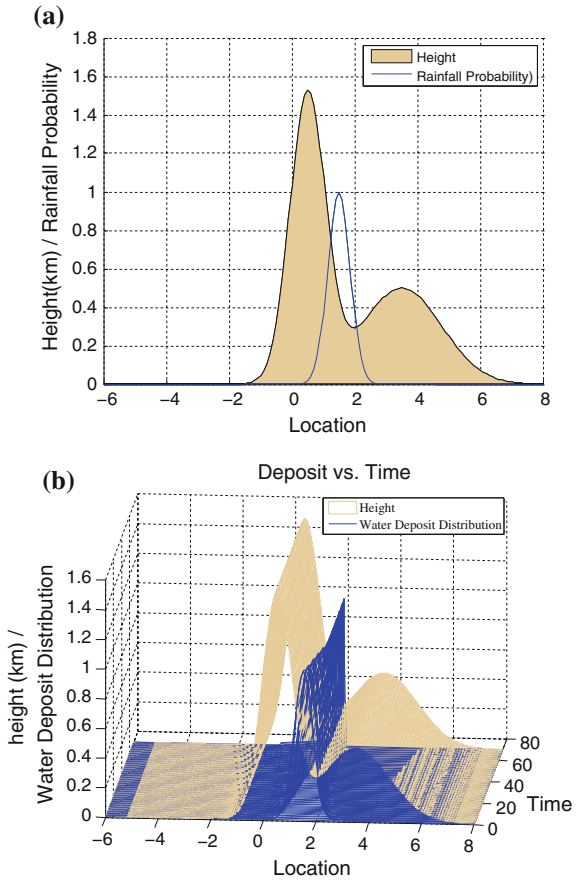


Fig. 3.8 Water flow for a 2D landscape. **a** Imaginary rainfall distribution over 2D landscape. **b** Water accumulation for a 2D landscape as a function of location

where v is the velocity of water flow. For the above calculation, the landscape needs to be approximated with some differentiable function. Alternatively, numerical gradients could have been used at the cost of requiring a large amount of memory to store the gradient matrix. This would also have the disadvantage of a roughness in landscape height profile resulting from the limitation of resolution of available topology data. Therefore, in the presented model, the landscape is approximated using a sum of bivariate Gaussian functions.

3.3.3.2 Gaussian Approximation of Topography Data

Here, the presented model is discussed using the landscape of Queensland, Australia. The area is simulated using 250 m data from Shuttle Radar Topography Mission (SRTM) [35] and resolution is coarsened to 25 km by taking the simple arithmetic mean. It should be noted that the model would give better results with the original

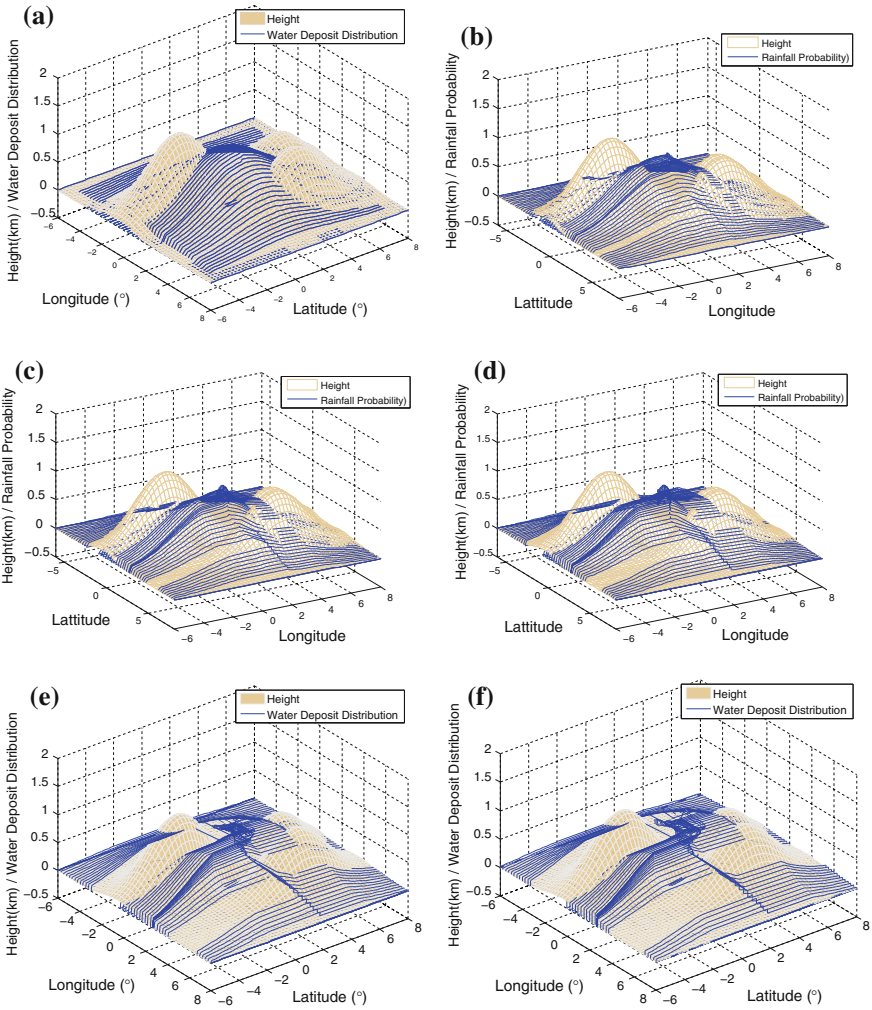


Fig. 3.9 Dynamics of water flow in a 3D landscape with t = T, 10T, 20T, 25T [T = Time interval = 0.02 s]: Probability distributions of water accumulation superimposed on the landscape. **a** After 1 time interval (t = T). **b** After 5 time intervals (t = 5T). **c** After 10 time intervals (t = 10T). **d** After 15 time intervals (t = 15T). **e** After 20 time intervals (t = 20T). **f** After 25 time intervals (t = 25T)

higher resolution data if computational resources are not a limitation. The landscape of Queensland was reconstructed using a bootstrapping technique to approximate with a linear combination of bivariate Gaussian functions with orientation. The reasons for this step are that the landscape was required to be approximated by a differentiable function in order to be used in the Eq. (3.7), to generate unlimited off-line data from limited available topography data, to remove errors in calculation resulting from the roughness of raw data, and to reduce computational costs associated

with large datasets when dealing with numerical gradients. The approximation is done with an initial equally spaced matrix of Gaussian functions, iteratively minimising a cost function dependent on error, and adjusting height and angle variables as explained in the steps below.

A bivariate elliptical Gaussian function centred at location (x_{1o}, x_{2o}) can be expressed as

$$\begin{aligned} h &= f(x_1, x_2, w, \theta, \sigma) \\ &= we^{\left[-\left(a(x_1-x_{1_o})^2+2b(x-x_{1_o})(x_2-x_{2_o})+c(x_2-x_{2_o})^2\right)\right]} \end{aligned} \tag{3.8}$$

where $a = cos^2\theta/2\sigma^2 + sin^2\theta/2\sigma^2$, $b = -sin2\theta/4\sigma^2 + sin2\theta/4\sigma^2$, $c = sin^2\theta/2\sigma^2 + cos^2\theta/2\sigma^2$, θ = angle, σ = standard deviation and w = weight of Gaussian.

Let h_c be the initial combination of Gaussians forming the height equation for the landscape. It can be expressed as a linear summation of elliptical Gaussian functions (using Eq. (3.8)):

$$h_c = \sum_{\forall i,j} f(x_{1i}, x_{2j}, w_{ij}, \theta_{ij}, \sigma_{ij}) \tag{3.9}$$

Error at each coordinate point (i, j) could be obtained by subtracting the current height (h_c) obtained from Eq. (3.9) from the actual height for each coordinate point as $e_{ij} = h_{ij_r} - h_{ij}$. Since error is a scalar, the cost function is $J = (1/2)e^2$ and the new weights can be obtained minimising the cost function

$$w_{ij_{new}} = w_{ij} - \eta \frac{\partial J}{\partial w_{ij}}.$$

which can be calculated using the chain rule for $\partial J/\partial w$ as

$$\frac{\partial J}{\partial w_{ij}} = e_{ij}.e^{\left[-\left(a(x_1-x_{1_o})^2+2b(x-x_{1_o})(x_2-x_{2_o})+c(x_2-x_{2_o})^2\right)\right]} \tag{3.10}$$

The error is reduced until stable below a threshold value by iterating the above steps. Same procedure was followed with angle θ as the variable. The above steps are summarised in Algorithm 3. This method reduces the amount of raw data that is needed for the calculations and thereby reducing the computational cost. If the cost function does not go below the recommended value, the number of Gaussians used is increased. But the amount of data needed to reconstruct the landscape would always remain below the amount of raw data by a large margin. The method has

been validated with data for one sub-area of landscape in simulations [17] where the model accurately predicts the area of flooding with respect to the flood distribution map downloaded from the Australian Bureau of Meteorology website [36].

Algorithm 3 Algorithm used in approximating the landscape with a summation of bivariate Gaussian functions with orientations

$i_t = 0$
$h_c = \Sigma_{\forall i,j} f(x_{1i}, x_{2j}, w_{ijo}, \theta_{ijo}, \sigma_{ij})$
while $i_t \leq i_{t_{max}}$ **do** {max no. of iterations = $i_{t_{max}}$}
  $i_t = i_t + 1$
  for each coordinate point i, j **do**
    $error = h_{actual} - h_c$
    if $error \leq \epsilon$ **then** {check stopping criterion}
      break;
    end if
    $J = (1/2)e^2$
    $w_{ij_{new}} = w_{ij} - \eta \frac{\partial J}{\partial w_{ij}}$
    $\theta_{ij_{new}} = \theta_{ij} - \eta \frac{\partial J}{\partial \theta_{ij}}$
  end for
  $h_c = \Sigma_{\forall i,j} f(x_{1i}, x_{2j}, w_{ij_{new}}, \theta_{ij_{new}}, \sigma_{ij})$
end while

While the limitations arising from the use of numerical gradients are reduced with the above approximation, the water flow calculations for any given landscape take up a considerable amount of processing since the water displacements need to be calculated for each sampling period and for each coordinate point on the landscape. This is especially true when the area of the landscape considered is large, and would not be efficient on a real time system. Using the fact that dynamics of water would not change with time for a given landscape, the products of the calculations can be stored in a matrix which is used to get the probability distribution of water accumulation at any future referencing time, given the initial rainfall distribution probability.

3.3.3.3 Transition Probability Matrix to Improve Calculation Efficiency

The static nature of topography is used to increase the efficiency of water flow calculations using a Transition Probability Matrix (TPM). This removes the need to recalculate the water flow using the gradient method for every different rainfall distribution and prediction time. When there is no further precipitation on the land, the water deposit distribution at some future time would only depend on that at the earlier time step. In this calculation only the surface run-off is considered and the soil moisture or evapotranspiration are not taken to account for reasons stated earlier.

The probability transition matrix for a landscape with n states (i.e. grid locations) can be expressed in a matrix $\mathbf{A} = [p_{ij}]_{n\times n}$ and is given by

$$\mathbf{A} = \begin{array}{c} \\ \text{Current} \\ \text{States} \end{array} \begin{array}{c} \begin{array}{cccc} \text{Next} & \text{States} & & \end{array} \\ \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \end{array},$$

where p_{ij} denotes the probability of the water deposit at location i to be moved to location j after a time interval of T.

The matrix is obtained by considering the dynamics of a unit rainfall distribution at each grid location of the considered landscape and calculating water flow resulting from it. The columns in the matrix are normalised.

The TPM possessed the following properties:

- $\sum_{i=1}^{n} p_{ij} = 1, \ \forall j \in \{1, 2, \ldots, n\}$
- $0 \leqq p_{ij} \leqq 1, \quad \forall i, j \in \{1, 2, \ldots, n\}$

Therefore the TPM is a Markov matrix and the water deposit distribution for any future time can be obtained by multiplying the water deposit distribution at the earlier time-step by the TPM when there is no further precipitation and no inflow from outside the *GP*. If the time step is defined as T and the current state is k, the probability distribution at time $(k + 1)$ can be obtained by

$$P(k + 1) = \mathbf{A}P(k) \tag{3.11}$$

where $P(k)$ is the probability distribution at time k. The states at future times (for intervals of T) can be obtained as a Markov chain. Therefore, the probability distribution after m time steps is

$$P(k + m) = \mathbf{A}P(k + m - 1). \tag{3.12}$$

From Eqs. (3.11) and (3.12), water deposit distribution at any future time m is given as

$$P(k+m) = \mathbf{A}^m P(k). \tag{3.13}$$

As time-to-failure is calculated for each *GP*, TPMs are calculated for each of them. Dividing the landscape into *GP*s, in addition, reduces the sizes of TPMs thus increasing efficiency in calculations. The characteristics of the *GP* are encapsulated in the properties of its TPM. For example, if the *GP* is a valley, there would be an absorbing state and the TPM would be an absorbing matrix where water deposit at any initial location would eventually end up at one of the absorbing states. But if the *GP* is a flat land, although there would be absorbing states the TPM would not be an absorbing matrix. In this case, although some water deposit will be stagnated at those points, another portion would flow away. The importance is that the properties of the topography in a given area can be identified mathematically through the properties of its TPM.

The absorbing rate of the Markov matrix is defined by its second-largest-eigenvalue ($SLEV$) which bounds the time needed to reach the steady state. Therefore, for the TPM of each geographic primitive, the time-to-failure can be defined through its $1/(1 - SLEV)$.

In a more complex setting where where water is being added to the current distribution through precipitation or through boundaries of a *GP*, Eq. (3.13) cannot directly be applied. For this scenario, the water distribution calculations are done for each time step using a slightly different form of Eq. (3.11) as

$$P(k+1) = \mathbf{A}(P(k) + D(k)) \tag{3.14}$$

where $D(k)$ is the water deposit distribution introduced at time k through run-off and inflow. The water deposit distribution prediction after m time periods (at time $(k+m)$) can be generalised as

$$P(k+m) = \mathbf{A}^m P(k) + \sum_{i=0}^{m-1} \mathbf{A}^{m-i} D(k+i). \tag{3.15}$$

Equation 3.15 reduces to Eq. (3.13) when there is no water introduced during m time periods and the second portion of Eq. (3.15) reduces to a Geometric series when the only amount of water introduced into the *GP* is from inflow. In all three scenarios, maximum complexity is $O(n^m)$. In all calculations, conversion of distance between positions to metric lengths and vice-versa are done as described in the Appendix 1 at the end of the chapter.

3.4 Two Case Studies

A comparison of prediction results is discussed here with two data sets; those of cyclone "Yasi" and cyclone "Tasha", which made landfall on Queensland, Australia, in, January–February, 2011, and December 2010 respectively. The simulations of cyclone "Yasi" is presented in detail while a few simulation steps with data of cyclone "Tasha" is included. The landscape of Queensland was taken as the considered land with x_1 (Latitude) ranging from 30 to 10S and the x_2 (Longitude) between 135E and 155E. The velocity of water flow was taken to be $v = 0.25\,\mathrm{m\,s^{-1}}$ and computing time interval as $T = 15\,\mathrm{min} = 900\,\mathrm{s}$. The radius of the earth was taken as 6378100 m. The topography of Queensland was modelled using the coarsened SRTM terrain data as previously described. 48×48 number of Gaussian functions are linearly combined to approximate the landscape of Queensland as described under Sect. 3.3.3.2 and setting $i = 1, 2, \ldots, 48$ and $j = 1, 2, \ldots, 48$ in Eq. (3.9). The landscape plotted from SRTM data is compared to the landscape simulated using the above method in Fig. 3.10.

The data source GES-DISC Interactive On line Visualisation ANd aNalysis Infrastructure ([37] (2011)) as part of the NASA's Goddard Earth Sciences (GES) Data and Information Services Centre (DISC) was used to obtain 'Yasi's' rainfall data. The best-track-data of cyclone "Yasi", which entered Queensland, Australia in January–February 2011, was obtained from the Regional and Mesoscale Meteorology Branch ([38] (2011)) and is plotted in Fig. 3.11. The same data was obtained for cyclone "Tasha" which entered Queensland on the 24th of December, 2010.

Cyclone 'Yasi's' path prediction with incoming data, with 95 % confidence bounds using the prediction steps, with a moving window, is depicted in Fig. 3.12. The four graphs show how the prediction changes with incoming data and the use of the moving window, and compares it with the actual track of the cyclone. The allowable deviation is calculated taking the speed of cyclone into consideration, and the window size for prediction dependent on the R^2 value of linear fit and the speed as discussed under

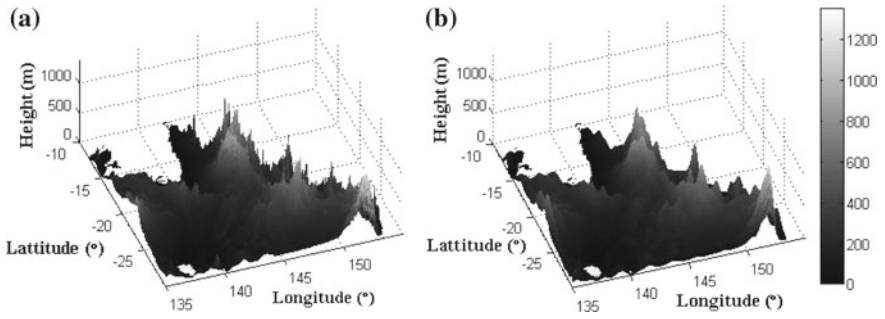


Fig. 3.10 **a** Topography of Queensland generated using SRTM topology data, compared to **b** Topography approximated through bootstrapping technique using a linear combination of Gaussian functions with orientation

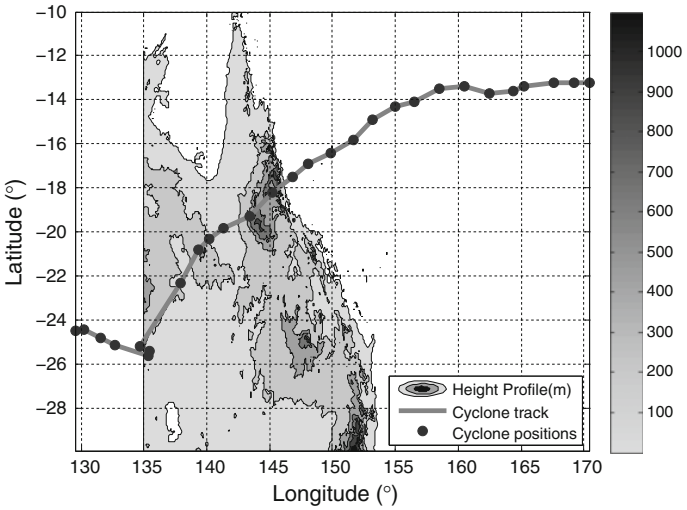


Fig. 3.11 Track of cyclone "Yasi" in January–February, 2011. The cyclone track is shown, plotted over the landscape of Queensland (The surrounding *white space* denotes sea and land which does not belong to Queensland). Topography is according to colour-bar

the presented model. Table 3.5 compares the official 12 h track forecast from JTWC (Joint Typhoon Warning Center) advisories to the proposed model's track forecast. A similar simulation for cyclone "Tasha" is shown in Fig. 3.13.

Verifying the predicting step for rainfall distribution is shown next. In Fig. 3.14, the output of the presented model is compared with actual rain gauge data for Queensland in January–February 2011. Comparisons are with the predictions using the R-CLIPER model and using the same Bayesian inference model proposed, with a bivariate Gaussian likelihood function without orientation (Figs. 3.6 and 3.7). The error for each step of rainfall prediction stage, as shown in Fig. 3.14, is calculated as the distance of the prediction's mean contours centroid from that of the mean rainfall. The same comparison for cyclone "Tasha" in shown in Fig. 3.15, which made landfall in Queensland, Australia in late December 2010.

Table 3.5 Comparison of errors (to the nearest km) of official 12 h cyclone track forecasts with presented model's forecasts

Prediction time	Official forecast error (km)	Model forecast error (km)
2011-01-31 12:00	78	44
2011-02-1 00:00	31	53
2011-02-1 12:00	24	76
2011-02-2 00:00	60	54
2011-02-2 12:00	31	9
2011-02-3 00:00	81	49

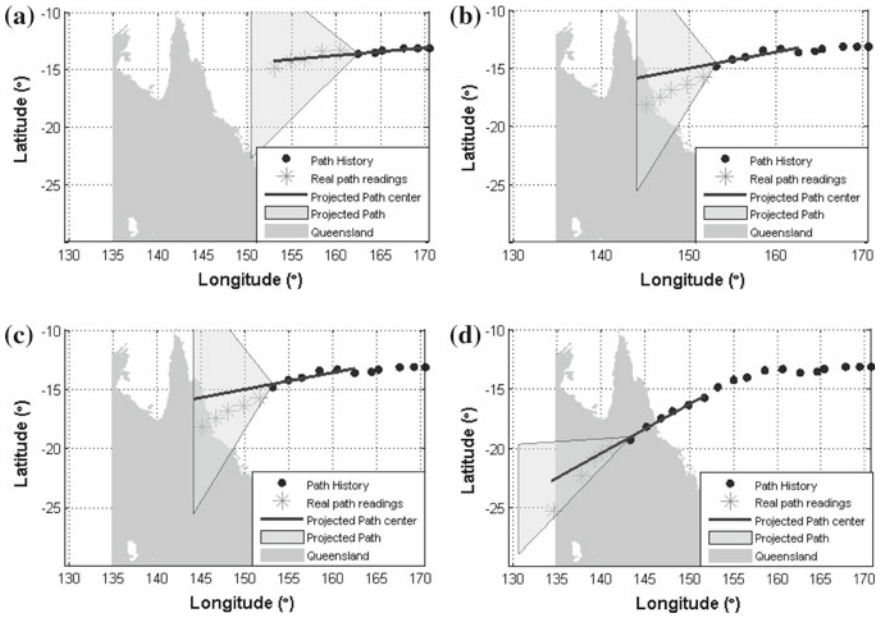


Fig. 3.12 Cyclone track prediction using a moving window (size dependent on speed of cyclone) of past data of Cyclone Yasi at 6, 11, 14 and 17 best-track data points available. *Dots* show the past track, *bold line* shows the centre of predicted path up to 30 h ahead, the *shaded area* is the allowed deviation (changing depending on R^2 value of linear fit) and the stars (*) show the actual path taken by cyclone after prediction time. **a** Path projection with 6 received readings. **b** Path projection with 11 received readings. **c** Path projection with 15 received readings. **d** Path projection with 17 received readings

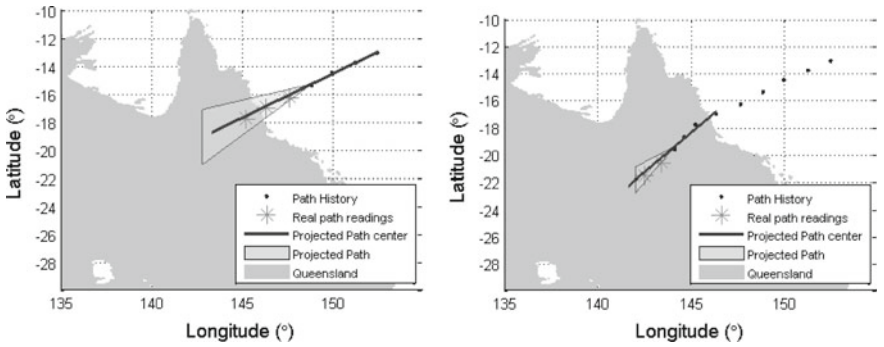


Fig. 3.13 Cyclone track prediction using a moving window of past data of Cyclone Tasha at 4 and 9 data points of past cyclone track available. *Dots* show the past track, *bold line* shows the centre of predicted path up to 30 h ahead, the *shaded area* is the allowed deviation (changing depending on R^2 value of linear fit) and the stars (*) show the actual path taken by cyclone after prediction time

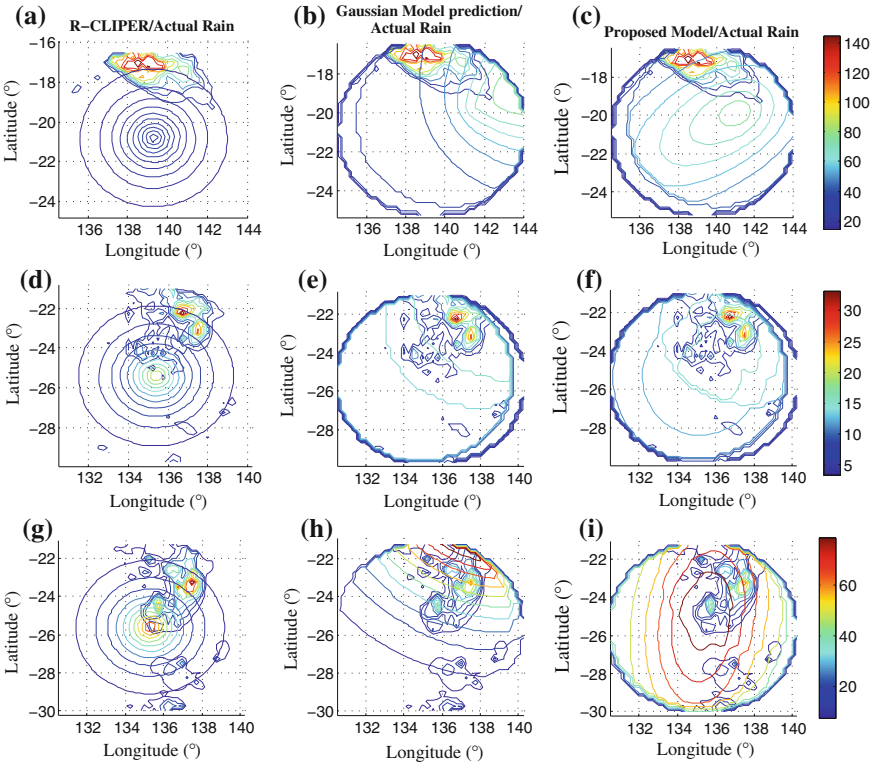


Fig. 3.14 Comparison of real rainfall gauge data with predicted rainfall over Queensland Australia, as a result of the activity of cyclone 'Yasi'. Three prediction steps are compared in the three rows of sub-figures. The columns show prediction using the R-CLIPER model, using a bivariate Gaussian likelihood function with Bayesian Inference, and using a bivariate Gaussian likelihood function with orientation with Bayesian Inference, respectively. In all three cases the predictions (sets of contours which are unique for each figure) are plotted on *top* of actual rain data (the irregular shaped contour sets that repeat in all three figures in a row). Prediction error is taken as the distance of the centroid of contour of the prediction's mean from contour centroid of actual rain mean. **a** Error = 50.9 km. **b** Error = 43.3 km. **c** Error = 43.0 km. **d** Error = 202.2 km. **e** Error = 100.9 km. **f** Error = 100.9 km. **g** Error = 163.5 km. **h** Error = 70.0 km. **i** Error = 70.2 km

Initial simulations obtained *GP*s by dividing the landscape into overlapping sub areas of dimensions 3° latitude by 3° longitude and the overlapping in 0.833° in each direction. This was done for simplicity in simulations. Flood distribution with time was simulated using a TPM for each of these sub-areas. The fact that accuracy of prediction depends on the *GP* type is elaborated in Fig. 3.16. The *GP* is a valley in the top two figures, where the flood prediction is less affected by a small error in the rainfall distribution approximation. In contrast, the error in flood prediction stage is magnified for the terrain in the bottom two figures where the *GP* is composed of a mountain range dividing the area into two portions. Therefore, in the output

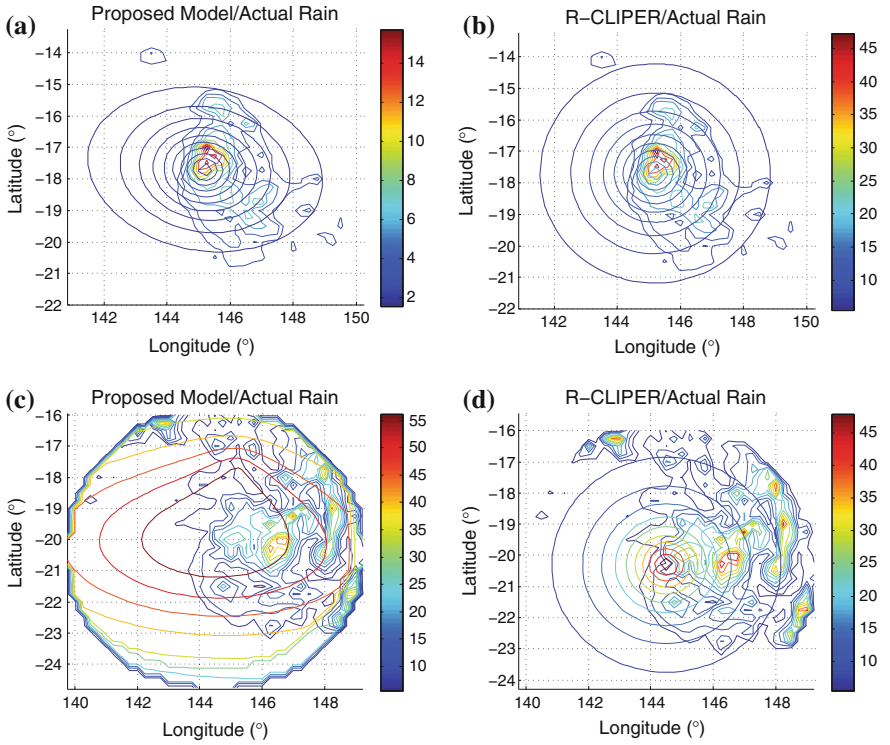


Fig. 3.15 Rainfall prediction versus actual rainfall for cyclone "Tasha" for two prediction steps (i.e. positions of cyclone eye) shown in two lines of sub-figures. In both cases the predictions (sets of contours which are unique for each figure) are plotted on *top* of actual rain data (the irregular shaped contour sets that repeat in both figures in a row). Prediction error is taken as the distance between the centroid of contour of the prediction mean from contour centroid of actual rain mean. **a** Error =139.5 km. **b** Error = 142.6 km. **c** Error = 54.5 km. **d** Error = 55.9 km

prediction, the variance of the output is highly dependent on the *GP* type. The topographic contour plots of three sub-areas with the contour plots of their TPMs are given in Fig. 3.17 illustrating the TPM's dependency on landscape. The absorbing states of the TPM of the landscape in Fig. 3.17c shows that the valley is an absorbing state for water deposit distributions. Therefore, it is evident that the properties of the topography in a given area can be identified mathematically through its Probability Transition Matrix and its properties. This is especially important when the landscape is complex where the TPM can be used to identify *GP*s.

Further simulations are run on identified *GP*s of irregular shapes. Two such *GP*s are plotted in Fig. 3.18. First *GP* is a valley that can be treated as an independent portion of calculations where water would neither enter nor leave the *GP*. The absorbing states plotted as dots show that the bottom of the valley is an absorbing state and

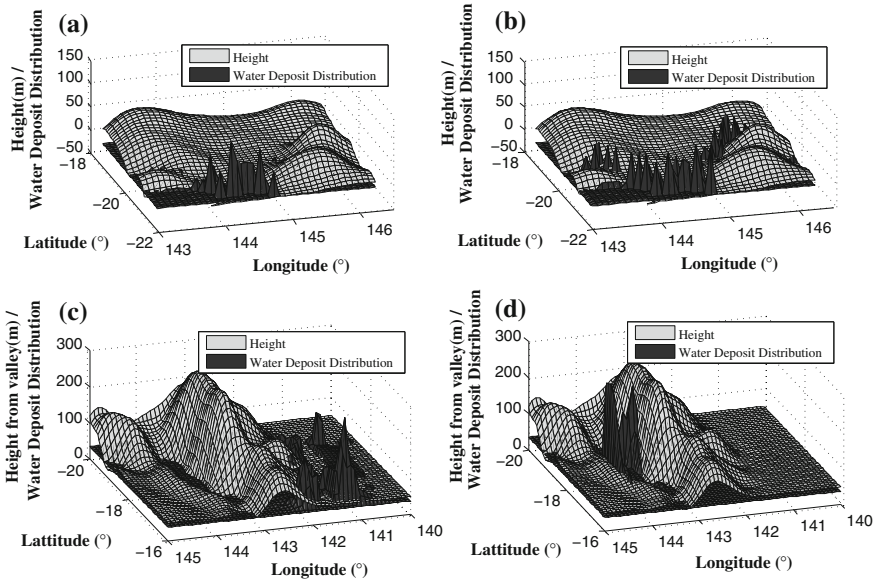


Fig. 3.16 Flood prediction **a** for a valley, using real rainfall gauge data **b** for the same valley, using rainfall prediction with added error **c** for an area containing a mountain range, using rainfall gauge data **d** for the same area containing a mountain range, using rainfall prediction with added error. Flood distributions for all cases are predictions after three time steps

there are two such minimum points in this sub-area. The bottom two figures show another *GP* of one side of a mountain range. There are no absorbing states in this TPM and water will only flow out of it and never into it. Therefore in both these *GP*s water flow calculations can be run ignoring the inflow term in Eq. (3.15).

3.4.1 Model Summary

A distributed Bayesian framework, defined across a set of predefined geographical primitives discussed that predicts the time to flood resulting from complex processes leading to the formation, development, and propagation of cyclones. Moreover, it allows starting flood prediction as soon as sufficient data points of cyclone path are received. This would be before the cyclone makes landfall allowing preparation well in advance. The prediction gets more and more refined with any incoming data. Furthermore, the partitioned approach allows asynchronous data to be input at any stage of the computation.

The flexibility in data assimilation resulting from the model's modularised structure is an important attribute of the presented model. In the rainfall prediction stage

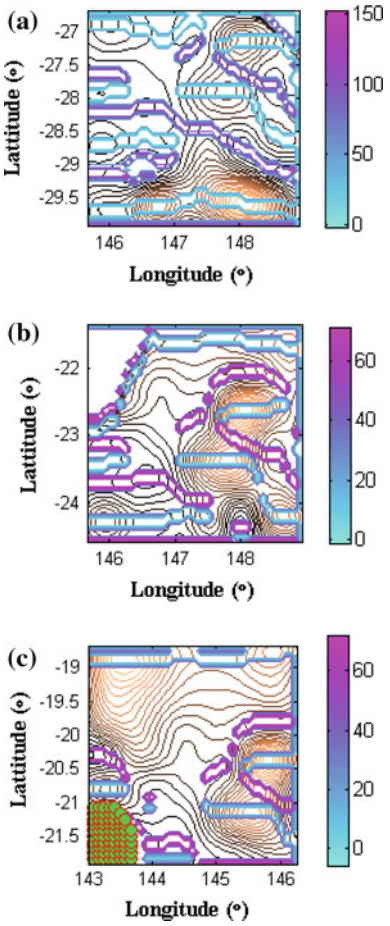


Fig. 3.17 Dependence of the Transition Probability Matrix (TPM) on landscape for three sub-areas if Queensland. The *brown* contours are of the landscape (*shade* gets lighter with increasing height) and the TPM contours are according to the colour-bar. **c** Shows the absorbing states of the TPM as *green dots* which can be observed located at the *bottom* of the valley in this case

of the model, all available rainfall data was used but not any other features such as wind shear or topology. The results summarised on Table 3.6 shows a 33 % increase of accuracy for rainfall density distribution prediction suggesting that by using Bayesian inference technique with recurrent learning, the prediction could be better than with methods using more sophisticated data assimilation. The main reason for the increase in accuracy is the ability of the proposed Bayesian learning based method to combine the advantages of data driven and heuristic modelling by amalgamating past knowledge through the prior distribution and new data through the likelihood function.

The comparison results in Table 3.5 makes evident the effectiveness of using a simple approach with minimal features in short term cyclone path prediction stage. The model forecasts are better for most cases, while noting that official forecasts are released three hours later than forecast of model due to the need of collecting a large amount of data from many sources.

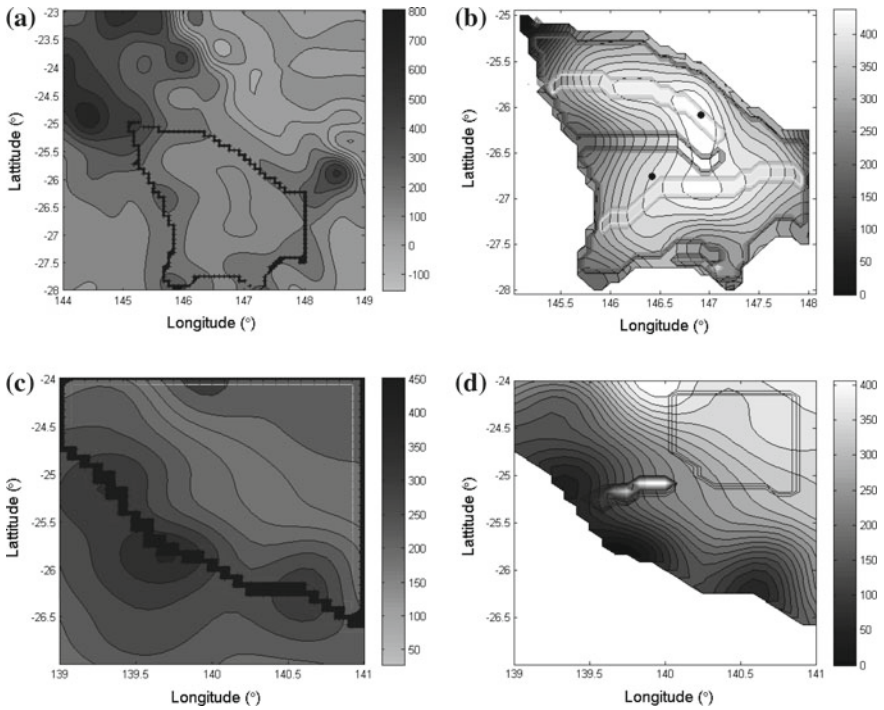


Fig. 3.18 Two identified *GP*s on Queensland in the first column and their TPM's (according to colour-bar) and absorbing states (*dots*) plotted on *top* of landscape on second column. **a** A valley identified as a *GP*. **b** PTM of *GP* on *top* of its topography. **c** A slope on one side of a mountain range identified as a *GP*. **d** PTM of *GP* on *top* of its topography

The concept of *GP*s that encapsulate the dynamics of water-flow on static topographical primitives, is a main feature of this model. Through statistically summarising simultaneous events that are spread in geography, it allows a distributed updating algorithm leading to parallel computation. The better understanding of the distributed nature of flood propagation and their relation to any geographical feature permit clustering of these primitives. It was measured that the speed of calculations increase by almost 20 times with the use of TPM on *GP*s as opposed to using gradient based calculations for each prediction time step. At this point manual demarcation of *GP*s is discussed in order to simplify the computational process. However this could be extended to mathematically identifying *GP*s for example by using numerical methods such as Delaunay triangulation [39]. The efficiency increases further when predicting for multiple time steps using Markov chains. The accuracy was reduced by the resolution of the states of the TPM which resulted in a maximum possible error of 6 km (at transition states of the TPM) with the resolution used in simulations. This reduces further with time as the TPM would approach absorbing states where water would be stagnated resulting in floods. The significance of this

error is further reduced with the use of variable resolutions considering the type of *GP* (e.g. Using higher resolution in valleys and lower resolution in flat land).

Complimenting the accurate predictions, the presented stochastic process for identifying *GP*s, with overlapping boundaries, using complex intrinsic statistics of flood propagation, allows identification of areas with flood risk mathematically even before using any rain data. This knowledge can be used in optimising disaster management strategies including land use. The terrain approximation provides a differentiable function permitting the development of a Markov chain process based on simulation data that bootstraps the real time computation based on measured rainfall data.

The discussed model however, has a few limitations as described below. One considerable limitation is that in the rainfall prediction step, projection is limited to an area of 500 km around the eye of the cyclone. This neglects the resulting rainfall due to convergence along the coasts during Extratropical Transition (ET) and accumulation of rainfall in rain-band echoes over specific regions further away from the eye of the cyclone as pointed by [24]. Additionally, although simplicity has its advantages, the model could be further improved by including some of the most significant features affecting cyclone activity, which will have less effect on efficiency. A more complete treatment of hydrologic aspects can make the predictions more generic by relaxing some of the assumptions such as frictionless and impermeable surfaces. If there are already existing relevant datasets for the area, they can be readily used to get a better flood vector field for the use in demarcation of *GP*s. It is still worth noting that although such comprehensive approaches make the prediction more realistic, great care has to be taken to get high quality data since the accuracy of prediction is highly sensitive to the used data. However, one important feature in using Bayesian updating is that it is purely data-driven and therefore for the rainfall prediction step, any error arising from the initial assumptions is suppressed after a few calculating time steps. It was also noticed that, for the cyclone path prediction step, while the model's 12 h prediction has indicated to be better, the 72 h prediction of the JTWC advisory was generally better than the 72 h prediction of the proposed model. Therefore, the model's prediction could be further improved using these outputs as input to the model using its modularised structure. Meanwhile, it should be noted that the model produces the predictions 3 h before the official forecasts are available, which would be much helpful for Just-In-Time sensor deployment and management and the JTWC advisories would serve only to improve on that prediction.

Existing water storages need to be included in the calculations. This could be included simply by considering the water levels of existing storages as a part of the landscape and building the Gaussian approximation of land from it. The flood levels can be defined by reducing the maximum capacities of water storages.

Although validated with two case studies from Queensland, it is worth noting that the mathematical framework discussed in this chapter does not use any assumptions specific to Queensland. Proposed is a probabilistic model of which the output is dependent solely on data. Therefore datasets of any cyclone in any region could be used with the presented model without further modification.

Table 3.6 Performance summary of major steps of the presented model

Prediction step	Efficiency comparison	Accuracy comparison	References
Cyclone path prediction (12 h horizon)	3 h before[a] official forecast	6.5 % Increase[b,c]	Table 3.5
Rainfall prediction	10 % Reduction in[d] processing speed	33.7 % Increase[e,f]	Figures 3.7 and 3.8
Flood prediction	18.9 times faster[g]	Max Error = 6 km[h,i]	
	3 % faster with use of irregular *GP*		

[a]Compared with official JTWC advisories
[b]Overall improvement over 12 h horizon compared to official JTWC advisories
[c]Distance error compared
[d]Compared to R-CLIPER prediction
[e]Distance error between centroids of mean rainfall contours taken for comparison
[f]Averaged from 5 prediction points and compared to the R-CLIPER prediction
[g]Compared to gradient based method
[h]Depends on the resolution of TPM used in calculations. Error can be reduced with increase in resolution
[i]Significance of error further reduces with the use of variable resolution in *GP*s

3.5 Discussion and Conclusions

This chapter discussed a comprehensive case study of predicting *MFPT* for the natural disasters of flood induced by cyclone activity. A Bayesian framework combining the advantages of data-driven and heuristic modelling, is used on a topographical network divided into GPs identified through flood propagation vector fields. The data driven prediction ensures that any initial errors diminish with time. The presented stochastic process for identifying GPs with overlapping boundaries, using complex intrinsic statistics of flood propagation, allows mathematical identification of areas with flood risk even before using any rain data.

Methods presented herein provide several advantages over numerical methods including real-time calculation efficiency, high lead time, easy data assimilation, and flexible model integration. It was seen that a probabilistic approach considering the effects of network inhomogeneity and bias can provide accurate and efficient predictions for temporal dynamics of cyclone induced flood density distributions. Identifying GPs using flood propagation vector fields provide a natural division of the problem, thus addressing both issues of network inhomogeneity and high computational complexity. It was also observed that a topography optimum model can be obtained using geography dependant resolution. Another advantage of the described TPM based method is that it is capable of handling a large class of bias scenarios even including a non-linear Manifold type bias distributions. This is not possible with many generic *MFPT* prediction models which require any biasing factors to be modelled in network properties such as degree distribution [40].

Reconstructing the landscape to a differentiable function allows generation of unlimited off-line data of desired resolution and permits the development of a Markov chain process for prediction. Oriented bivariate Gaussian gave the best likelihood function. The limitations of the model results from its initial assumptions discussed earlier.

Though the methods presented in this chapter are specific to the application, they provide the background to the development of more generic methods discussed in the next several chapters and the concepts used such as GPs and Eigenmode analysis on TPMs are revisited throughout the rest of this book.

3.6 Summary

This chapter presented a comprehensive case study on developing methods to predict *MFPT* for cyclone induced flood propagation considering the effects of network inhomogeneity and bias in propagation. A modularised approach is presented dividing the process into three stages; cyclone track forecasting, cyclone induced rainfall forecasting and prediction of flood propagation. The distributed Bayesian framework defined across the inhomogeneous media divided into predefined set of GPs allow efficient real-time prediction by encapsulating the effects of bias on flood propagation into TPM for each GP. The model was compared with several established modes and verified with two real datasets of cyclone data from 2010–2011 cyclone season where it was observed that the model was capable of predicting up to 3 h ahead of official forecasts with a 33 % improvement of accuracy.

Appendix 1: Conversion Between Position and Distance

Latitude degrees are parallel to each other and therefore, the distance between two degrees of Latitude does not change with position (neglecting the earth's slightly ellipsoid shape) and could be taken roughly equal to 111 km. Distance between two Longitudinal degrees would significantly change with location as degrees of Longitude are further apart near the equator and approaches zero at the poles. The distance between two longitudinal points is therefore taken by approximating the earth to a sphere.

The radius of the circle at latitude ϕ is calculated as

$$R = \sqrt{\frac{(a^2\cos(\phi))^2 + (b^2\sin(\phi))^2}{(a\cos(\phi))^2 + (b\sin(\phi))^2}} \tag{3.16}$$

where a is the Earth's equatorial radius and b is the Earth's polar radius.

The *distance* in meters can be converted to degrees longitude by

$$\lambda = \frac{180}{\pi} \times \frac{distance}{R}. \tag{3.17}$$

References

1. Chris Landsea JF, Beven J (2014) Atlantic hurricane database (HURDAT2) 1851–2014
2. Klotzbach PJ, Barnston A, Bell G, Camargo S, Chan JC, Lea A, Saunders M, Vitart F (2010) Seasonal forecasting of tropical cyclones
3. Roy C, Kovordányi R (2012) Tropical cyclone track forecasting techniques-a review. Atmos Res 104:40–69
4. Liu GR, Kuo TH, Chen WJ, Lin TH (2012) Prediction of tropical cyclone rainfall potential in Taiwan mountainous areas. In: Rainfall forecasting, Nova Science Publishers, Chap 5:199–232
5. Asante KO, Arlan GA, Pervez S, Rowland J, Artan GA (2008) A linear geospatial streamflow modeling system for data sparse environments. Int J River Basin Manage 6(3):1–9
6. BBC News (2011) Australia: Queensland floods spur more evacuations. http://www.bbc.co.uk/news/world-asia-pacific-12097280
7. ABC News (2011) Flood costs tipped to top $30b. http://www.abc.net.au/news/2011-01-18/flood-costs-tipped-to-top-30b/1909700
8. IPCC (2012) Summary for Policymakers. In: Managing the risks of extreme events and disasters to advance climate change adaptation. Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ, Plattner GK, Allen SK, M. T. Tech. Rep. Cambridge University Press, Cambridge, UK, New York, USA
9. Asante K, Artan G, Pervez S, Rowland J (2008) A linear geospatial streamflow modeling system for data sparse environments. Int J River Basin Manag 6:1–9
10. Nanayakkara T, Halgamuge M, Sridhar P, Madni A (2011) Intelligent sensing in dynamic environments using markov decision process. Sensors 11(1):1229–1242
11. Albers S (2010) Energy-efficient algorithms. Commun ACM 53(5):86–96
12. Passos RM, Coelho CJN, Loureiro AAF, Mini RAF (2005) Dynamic power management in wireless sensor networks: an application-driven approach. In: Second annual conference on wireless on-demand network systems and services, WONS 2005, pp 109–118
13. Yee G, Shucker B, Dunn J, Sheth A, Han R (2006) Just-in-time sensor networks. In: Third workshop on embedded networked sensors
14. Ding C, Zhang J, Wang S (2010) Disaster prevention decision-making method based on bayesian analysis. In: 2010 3rd IEEE international conference on computer science and information technology (ICCSIT), vol 9, pp 449–451
15. Bates PD, De Roo APJ (2000) A simple raster-based model for flood inundation simulation. J Hydrol 236(1–2):54–77
16. Bates PD, Marks KJ, Horritt MS (2003) Optimal use of high-resolution topographic data in flood inundation models. Hydrol Process 17(3):537–557
17. Wijesundera I, Halgamuge MN, Nirmalathas T, Nanayakkara T (2013) A geographic primitive-based bayesian framework to predict cyclone-induced flooding*. J Hydrometeorol 14(2):505–523
18. Grayson R, Blöschl G (2001) Spatial modelling of catchment dynamics. Observations and modelling, Spatial patterns in catchment hydrology, pp 51–81
19. Dehotin J, Braud I (2008) Which spatial discretization for distributed hydrological models? proposition of a methodology and illustration for medium to large-scale catchments. Hydrol Earth Syst Sci Discuss 12(3):769–796
20. Zerger A, Wealands S (2004b) Beyond modelling: linking models with gis for flood risk management. Nat Hazards 33(2):191–208

21. Asante K, Macuacua R, Artan G, Lietzow R, Verdin J (2007) Developing a flood monitoring system from remotely sensed data for the Limpopo basin. IEEE Trans Geosci Remote Sens 45(6):1709–1714
22. Da Silva RI, Del Duca Almeida V, Poersch AM, Nogueira JMS (2010) Wireless sensor network for disaster management. In: 2010 IEEE network operations and management symposium (NOMS), pp 870–873
23. Lonfat M, Marks F Jr, Chen S (2004) Precipitation distribution in tropical cyclones using the tropical rainfall measuring mission (trmm) microwave imager: A global perspective. Mon Weather Rev 132(7):1645–1660
24. Lonfat M, Rogers R, Marchok T, Marks F (2007) A parametric model for predicting hurricane rainfall. Mon Weather Rev 135(9):3086–3097
25. Ebert EE, Turk M, Kusselson SJ, Yang J, Seybold M, Keehn PR, Kuligowski RJ (2010) Ensemble tropical rainfall potential (etrap) forecasts. Weather Forecast 26(2):213–224
26. Li L, Wang J, Leung H, Jiang C (2010) Assessment of catastrophic risk using bayesian network constructed from domain knowledge and spatial data. Risk Anal 30(7):1157–1175
27. Lu MM, Chu PS, Lin YC (2010) Seasonal prediction of tropical cyclone activity near taiwan using the bayesian multivariate regression method. Weather Forecast 25(6):1780–1795
28. Bermak A, Belhouari S (2006) Bayesian learning using gaussian process for gas identification. IEEE Trans. Instrum. Meas. 55(3):787–792
29. Madsen H, Jakobsen F (2004) Cyclone induced storm surge and flood forecasting in the northern bay of bengal. Coast Eng 51(4):277–296
30. Bessafi M, Lasserre-Bigorry A, Neumann C, Pignolet-Tardan F, Payet D, Lee-Ching-Ken M (2002) Statistical prediction of tropical cyclone motion: an analog-cliper approach. Weather Forecast 17(4):821–831
31. Hurricanes: Science and Society (cited 2012) Dynamical models. http://www.hurricanescience.org/science/forecast/models/modeltypes/dynamicalmodels/
32. Bender MA, Ginis I, Tuleya R, Thomas B, Marchok T (2007) The operational gfdl coupled hurricaneocean prediction system and a summary of its performance. Mon Weather Rev 135(12):3965–3989
33. Gall JS, Ginis I, Lin SJ, Marchok TP, Chen JH (2011) Experimental tropical cyclone prediction using the gfdl 25km resolution global atmospheric model. Weather and Forecasting 26 (in press)
34. Marks F, Kappler G, DeMaria M (2002) Development of a tropical cyclone rainfall climatology and persistence (r-cliper) model
35. CGIAR-CSI (cited 2011) Srtm 250m digital elevation data. http://srtm.csi.cgiar.org/selection/
36. Bureau of Meteorology, Australia (cited 2016) Map of flood affected cities and towns in Queensland December 2010 and January 2011. http://www.bom.gov.au/climate/data/index.shtml
37. Giovanni (cited 2011) Trmm online visualization and analysis system (tovas). http://disc2.nascom.nasa.gov/giovanni/tovas/
38. RAMMB (cited 2011) Real time tropical cyclone products 2011 season. http://rammb.cira.colostate.edu/
39. Lee D, Schachter B (1980) Two algorithms for constructing a delaunay triangulation. Int J Parallel Program 9(3):219–242
40. Fronczak A, Fronczak P (2009) Biased random walks in complex networks: the role of local navigation rules. Phys Rev E 80(1):016,107

Chapter 4
First Arrival Time for Natural Disasters Modelled as Biased Networks

Abstract Taking forward the concept of calculating the mean first passage time for natural disasters, this chapter focusses on discussing generic methods for calculations in directionally biased networks. Following a short discussion on the current state of art in *MFPT* prediction for biased random walks, a case invariant modification of transport variables is described as a viable solution to the problem of addressing directional bias. The empirical equation that modifies the walk dimension leads the way to significantly better *MFPT* estimations when compared to the use of conventional transport variables. A range of simulated random walks on hypothetical networks are used to describe this model. Following this, a comprehensive case study is presented to validate the model with the real dataset of archived cyclone tracks over the North Atlantic Ocean. The comparative studies of prediction results and the generic nature of model development indicate wider applicability of the presented methods.

4.1 Introduction

For a large class of random walks describing the processes behind natural disaster dynamics, the existence of direction dependent bias is an all too common phenomenon. Effects of such directional bias on the *MFPT* calculations is a question worth investigating while stepping towards an application independent approach, where the networks are viewed as generic node distributions (either continuous or discreet) on Euclidean state space.

The complexity inherent in natural processes lead to state transitions for natural systems to often behave as random walks. It is also very common that many such walks are biased by various environmental phenomena. This is true for widely distributed dynamical systems like the propagation of fire biased by wind and fuel distributions [1, 2], flood propagation biased by topography [3–5], epidemics biased by population density distributions [6] and cyclones biased by weather conditions [7], as well as for more localised dynamical systems. Scientists have long been interested in understanding the significance of network inherent bias on *MFPT* calculations. It is also interest-

I. Wijesundera et al., *Natural Disasters, When Will They Reach Me?*,
Springer Natural Hazards, DOI 10.1007/978-981-10-1113-9_4

ing to investigate on any common features of such walks in diverse applications which could lead to more generic approaches of addressing directional bias.

As discussed in detail under Chap. 2 current work on *MFPT* prediction for random walks has advanced to finding efficient and accurate methods for random walks on a range of special networks including those on complex scale invariant media [8], fractals [9], scale-free networks [10], and lévy flights [8]. But when the random walks are results of state transitions for real world dynamic systems, such ideal conditions are not always met. These walks are commonly biased towards attractors formed by non-linear interaction dynamics between the system and the environment.

Bias, or drift occurs naturally in many real world dynamic systems [11] including those resulting from source-sink type topologies, and advection on flow potential [11, 29] among many others. This chapter specifically concentrates on random walks affected by directional bias and in the context of this chapter, it is assumed that the networks are homogeneous in transport characteristics; or in other words, when the properties of the random walker are independent on which part of the network the walks currently is.

Literature gives various definitions for bias in random walks. The most common way is describing bias in terms of node degree distribution as described in Chap. 2. For networks that cannot be described by a degree distribution (i.e. when the links between the nodes are dynamic and probabilistically dependent on the bias [12, 13], biased random walks are described in different terms for different types of networks. Fronczak et al. [14] defines it as "preferential transition probability in networks having nodes with arbitrary degree distribution". Some have extended this method to include the node weight distribution in bias correction [15]. Skarpalezos et al. [16] describe it as "the bias is represented as the probability P of the particle to travel along the shortest path to the target node".

This chapter targets a different class of random walkers where the network connections are temporal [13] and decided only at the time each step is taken. The reason behind this focus is that most common bias present in processes behind natural disaster activity is of this kind. For example, when looking at fire propagation, the next fuel point that catches fire is dependent on a range of factors including the temperature, wind direction, wind speed, fuel type, distance, etc. Therefore, for this example although the effects of bias are clear, a node degree distribution does not make sense.

More specifically, networks showing temporal connectivity resulting from biased transitions need to be looked at [17]. Past literature has shown that the first passage time characteristics for such walks are largely dominated by bias [18]. For a state space of a dimension larger than one (i.e. $d > 1$), a biased random walk mostly occupies nodes within a narrow cone along the bias direction. While the length of this cone is proportional to time t, the width is proportional to $t^{1/2}$ giving a density of visited nodes within the cone as $t/t^{(1+(d-1)/2)}\ t^{(1-d)/2}$ [18].

The transport characteristics of biased random walks can be summarised into a range of transport variables which are commonly used for *MFPT* calculations.

4.2 Predicting Failure Time Using Transport Variables

Extensive work has been done on predicting *MFPT* using various transport properties introduced in Chap. 2. Two of the most common transport properties used in *MFPT* prediction methods [8, 19–21] are the two characteristic dimensions; the fractal dimension (*df*), and the walk dimension (*dw*). These two metrics jointly have been used to describe the transport characteristics of random walks.

4.2.1 The Fractal Dimension (df)

For a given random walk on a d dimensional state space the *df* can be introduced as; If there are N number of nodes within a d dimensional sphere of radius r, then N for a fractal network shows as proportional to r^{df} [8, 20]. The *df* is thus known to characterise the density of nodes in a network or in other words the reachability of nodes for the random walker.

The general expression for *df* is

$$N \propto r^{df}. \tag{4.1}$$

For a continuous state space, *df* follows the dimension of that state space.

For complex networks with a scale-free degree distribution, an alternative to the standard fractal dimension has been proposed to be the *Box dimension* (d_B). This is when renormalising is possible by covering with N_B non overlapping boxes of size l_B where $N_B/N \propto l_B^{-dB}$ [8].

4.2.2 The Walk Dimension (dw)

The path resulting from a random walk can be viewed as a random object which has been found out to be self-similar as per [22]. The fractal dimension of such a trail of a random walk is called the '***walk dimension***' of that walk and denoted as ***dw.*** This characterises the mean square displacement from the source as a function of the number of moves.

The length-scale invariant properties of the random walks result in the scaling of the form

$$t_{exit} \propto r^{dw} \tag{4.2}$$

defining the walk dimension for an unbiased network. Here, the t_{exit} is the exit-time from a sphere of radius r from the source S.

The dw is a measure of the ease with which the walker can move away from the source and therefore accounts for any obstacles and potential fields influencing the walk. The higher the dw is, the harder it is for the walker to leave its immediate neighbourhood. The $dw = 2$ for regular diffusion but in fractals formed by random walks, generally it tends to be that $dw \neq 2$ resulting in '*anomalous diffusion*'.

4.2.3 Network Exploration by a Random Walk

Important insights of a random walk on a given network can be observed by comparing the magnitudes of df and dw [23]. If $dw < df$, it signifies that the walk follows a 'non-compact exploration' meaning that the walker does not need to traverse a large part of the immediate neighbourhood to exit from it. The walker moves further away from the source with time. The dependence of source location is lost when the distance to the target is higher. In contrast, networks with $dw \geq df$ shows 'compact-exploration' where the starting point always matters since each node is eventually visited as obstacles make it harder for the walker to move away from the source. Random walks with $d < 2$ would result in the walker visiting each site within the sphere infinitely often thus making way for compact exploration always [18] and time to return to origin diverges for $d \leq 2$.

Relaxation time (refer Chap. 2) has also been used to describe network exploration characteristics. Random walks with short relaxation times are also known as random walks with non-compact exploration [15]. In the context of random walk based searching, it has been shown that the physical principle underlying an improved search efficiency is an optimised balance between the search space being much larger, equal and smaller than the target [24]. In other words, it has been shown that non-compact exploration is more useful in bringing the random walker closer to the target but once closer, compact exploration performs better in hitting the target from closer proximity.

In the context of this book, a network homogeneous in df and dw is called a network primitive (NP) which will be described in detail in Chap. 5. For this chapter, it is assumed that the network consists of a single NP.

With respect to the relative values of df and dw, Condamin et al. have presented a powerful $MFPT$ prediction method analysed using the pseudo Green function [8]. They have shown relations of $MFPT$ to the network size (N) and distance between

the source and the target (r). The general applicability of their result to non-fractal networks has been discussed in [15, 25]. This generic theory allows reasonably accurate evaluation of the *MFPT* in complex media for a range of stochastic processes characterised by length-scale invariant properties [8].

The next part of this chapter discusses an adaptation of the methods presented in [8] and extending it for directionally biased networks. Condamin et al. [8] present a powerful result which facilitates simplicity and compatibility with older models [23]. Condamin et al. [8] show that for a length-scale-invariant network, the *MFPT* can be obtained as

$$MFPT \sim \begin{cases} N(A - Br^{dw-df}), & \text{if } dw < df \\ N(A + B\ln x), & \text{if } dw = df \\ N(A + Br^{dw-df}), & \text{if } dw > df \end{cases} \tag{4.3}$$

where N is the number of nodes, $\sim$ indicates large N asymptotic equivalence, and A & B are domain dependent constants. The rest of this chapter describes that this method can be extended for networks showing directional bias, by bias-modifying the transport variables df & dw.

4.2.4 Assumptions

Several assumptions are used in the prediction methods presented in the rest of this chapter.

- Networks are homogeneous in transport properties.
- The effects of bias at any location can be quantified as an *intensity* towards a *direction* on Euclidean state space.
- Network connections are temporal and the global connectivity is not known at the time of prediction. Although this is not a limitation, the effectiveness of the presented methods is more significant for such applications.
- For the case study, it is assumed that best-track data are correct and represents the actual path of the eye of the cyclone.

4.3 Time to Reach in Biased Media

The main focus of this chapter is on discussing the methods of predicting *MFPT* for dynamic systems which result in random walks showing directional bias.

A **biased network** can be defined as **a network where there are concentration gradients in exit points in a sphere of any radius** r. The degree of bias depends on the relative magnitude of the gradient in one direction relative to others. In a slightly different definition, for biased random walks the $MFPT$ from a given source (S) to a given target (T) is directed (i.e. $MFPT(S \rightarrow T) \neq MFPT(T \rightarrow S)$).

Therefore for such application, any $MFPT$ prediction method needs to account for the aforementioned biased motion common in real world dynamic systems. This chapter shows that this can be done by bias-modifying the transport variables df and dw. As df is merely a property of the network it is not altered. In contrast, dw should indeed account for any potential fields shaping the walk. Firstly, the probability of reaching (P_r) any given T is dependent on the relative angle between the line connecting S to T and the bias direction θ (Fig. 4.1a), and the bias intensity U. Therefore, with increasing U, although the rate of displacement from S is higher, the angle segment in which the targets have significant P_r's lowers. Thus, the overall reachability of the network decreases with increasing U. For a generic prediction model, it was required to find a generic bias modified walk dimension dw_b, for all nodes within an NP, such that the variances of dw_b and df are negligible.

Bias-modified walk dimension (dw_b)
The characteristic dimension which defines the relationship between t_{exit} and r for a directionally biased random walk.

In a biased network where there are concentration gradients in exit points in a sphere of any radius r, the effective bias modified dw is dependent on the relative direction of the walk from the source. The next section discusses the process of empirically obtaining dw_b. For this, a dataset constructed of simulated random walks on a set of hypothetical 2D networks was used as described in the next subsection.

4.3.1 Simulated Biased Random Walks

In order to discuss the empirically developed expression for a generic dw_b, a set of hypothetical unconnected 2D networks is used. To derive a large set of NP's, a uniform network with evenly spaced nodes is modified changing the locality of the nodes by moving them within a fixed standard deviation. Starting with a uniformly distributed 50×50 NP (a section shown in Fig. 4.1b), several NP's are derived by adding Normal noise with a fixed standard deviation (σ) as

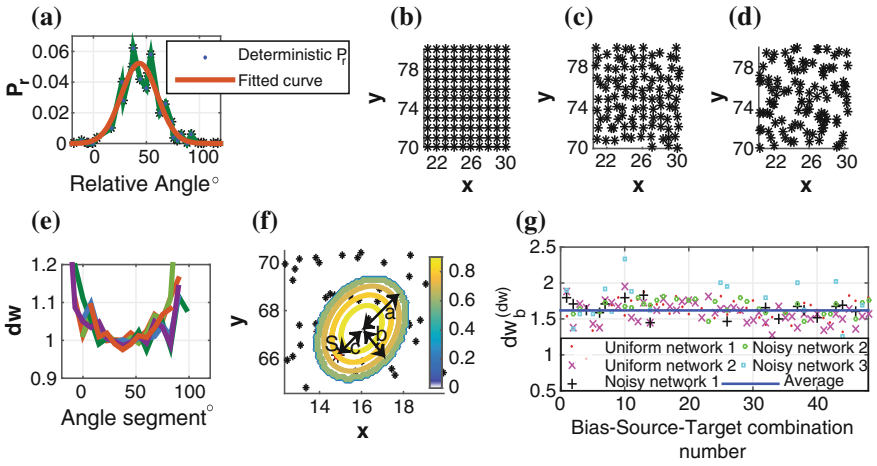


Fig. 4.1 Empirical derivation of a case invariant relationship between dw and dw_b. **a** An example profile for P_r for a network ($\sigma = 0.3, \theta = 45°$) with respect to relative angle. Several NP's are used including **b** a uniformly distributed NP and **c, d** NP's derived from that in b with added σ's of 0.15, and 0.5. **e** The profile for dw for a similar network as (a) with 6 initialisations of the network. **f** An example Gaussian probability density distribution (P_{DD}) for selecting next step in a random walk ($\theta = 45°$). Nodes shown as stars and dimensions of P_{DD}'s base ellipse with major Eigen vector having the same direction as bias. **g** The relationship obtained empirically through comparing the values calculated for dw with the optimal dw_b for 48 Bias-Source-Target combinations for each of five network scenarios ($\sigma = 0$, 0 (with higher sparsity), 0.15, 0.5 and 0.8)

$$\forall_i(x_{ni}, y_{ni}) = (\mathcal{N}(x_i, \sigma), \mathcal{N}(y_i, \sigma)) \tag{4.4}$$

where (x_i, y_i) is the position of the the ith node without noise (Fig. 4.1c, d). The σ's are kept low ($0 < \sigma \leq 0.9$) for all nodes to belong to a single NP. Each initialisation of such network is essentially different from the next even though both share the same noise level. Several initialisations are used with each σ to assure the consistency of results.

Random walks are simulated with a bi-variate Gaussian probability density distribution (P_{DD}) at each step of the walk as

$$P_{DD}(x, y) = e^{-(d(x-C_x)^2+2e(x-C_x)(y-C_y)+f(y-C_y)^2)} \tag{4.5}$$

where (C_x, C_y) is the centre of the base ellipse after shifting the current position a distance of c in the θ direction (Fig. 4.1f), $d = cos^2\theta/2\sigma_x^2 + sin^2\theta/2\sigma_y^2$, $e = -sin2\theta/4\sigma_x^2 + sin2\theta/4\sigma_y^2$, and $f = sin^2\theta/2\sigma_x^2 + cos^2\theta/2\sigma_y^2$. The values for σ_x and σ_y, are chosen from $a = (R+R/HB)/2LB$ (minor axis) and $b = (R+R/HB)/2$ (major axis) depending on θ. $c = b - R/HB$ where R = rate of spread (speed of the walk without bias), Length-to-Breadth ratio $LB = e^{0.25U}$ and Head-to-Back ratio $HB = LB^{2.5}$. The LB and HB are taken such that for an environment without bias, LB would be 1 and HB would be an augmentation of LB. This P_{DD} was inspired by

a fire propagation model [26] that used Huygens principle, which is generally known to be a universal model of propagation [27] for natural environments (for a short summary see Appendix 1 at the end of the chapter). This elliptical P_{DD} also relates to the variability in ellipses of cyclone forecast error profiles (see [28]) as well as many other natural propagation scenarios.

4.3.2 Calculating Transport Properties for Each NP

A large collection of NP's obtained as described above was used to empirically develop a generic expression for dw_b. For each NP, df was calculated using Eq. (4.1) and dw using a Monte-Carlo approach described below. Figure 4.1e shows the dw for six initialisations of an NP with $\sigma = 0.3$ and $\theta = 45°$. The dw is lowest at θ where the walker experiences highest ease of motion.

4.3.3 Calculating the Walk Dimension (dw)

The steps followed to calculate the dw are summarised in Algorithm 1. The dw is pre-calculated and recorded in a lookup table with respect to the bias intensity U ($0.6 \leq U \leq 1.6$ for the simulations) and relative direction ($\Phi = -\pi : \pi/24 : \pi$). The calculation for each U is iterated a large number of times ($mIter = 500$ in the simulations) for the average dw to converge into a stable value. Random walks are generated having a predefined maximum hops per walk (max_hops). The next node is selected using P_{DD} (Fig. 4.1f). The distance (r) and Φ for the new node are calculated and recorded alongside the hop number with respect to the angle segment of the new node. Finally dw is calculated using curve fitting ($dwfit$) for the relation $t_{exit} \propto r^{dw}$ by non-linear least squares method.

4.3.3.1 Dependence of *dw* on Relative Direction of Propagation

To capture the effect of bias on the ease of which the walker can move in directions relative to bias the results are divided into angle segments in obtaining the dw of the hypothetical networks on 2D space using Monte-Carlo simulations. Figure 4.2c, f, i and l show the dw response for 4 systems with varying noise levels (of standard deviations 0.1, 0.3, 0.5 and 0.7, respectively) and a uniform bias towards a 30° angle. Six initialisations of each noise level are considered in each graph. It can be seen that dw tends to be lowest in the direction of bias which is explained by the fact that the ease with which the walker moves further away from the source is highest in the direction of bias. Similar results are obtained with simulations having different values for θ, R and σ.

Algorithm 1 Obtaining dw table at a given source node

$[c_{sx}, c_{sy}]$ = coordinates of source node
define $mIter, max\_hops$
initialise vectors U, Φ
for $\forall u \in \{U\}$ **do**
  for $i = 1 : mIter$ **do**
    for $s = 1 : max\_hops$ **do**
      current node $\rightarrow [c_{1x}, c_{1y}] = \begin{cases} [c_{sx}, c_{sy}], & \text{if } s = 1. \\ [c_{2x}, c_{2y}], & \text{otherwise.} \end{cases}$
      define $\theta, U \leftarrow f(u), R$
      $P_{DD} \leftarrow f(c_{1x}, c_{1y}, \theta, U, R)$ {Eq. 4.5}
      Next node $[c_{2x}, c_{2y}] \leftarrow f(P_{DD}, F)$
      $dist = \sqrt{((c_{2x} - c_{1x})^2 + (c_{2y} - c_{1y})^2)}$
      $hop\_dir = \pm tan^{-1}((c_{2y} - c_{1y})/(c_{2x} - c_{1x}))$
      $hop\_segment \leftarrow f(\Phi, hop\_dir)$
      if $max(t_{exit}(u, hop\_segment, :)) < dist$ **then**
        $t_{exit}(u, hop\_segment, end + 1) = s$
        $r(u, hop\_segment, end + 1) = dist$
      end if
    end for
  end for
  for $\forall seg \in \{\Phi\}$ **do**
    $dw(S, u, seg) = dw_n \leftarrow dwfit(t_{exit}, r; t_{exit} = ar^{dw_n})$
  end for
end for

4.3.3.2 Bias Modified dw_b

A generic relationship for obtaining dw_b is the main objective at this point. With the aim of obtaining a case invariant relationship between dw with dw_b, a large number of scenarios were considered (9 NPs ($0 < \sigma \leq 0.9$) $\times$ 6 initialisations of each $\times$ 8 (S, T) sets $\times$ 6 U's). In order to get deterministic dw_b values, the $MFPT^*$ values are obtained numerically using Monte-Carlo simulations from a fixed S to reach a set of T's at different distances. Each network was considered as a small portion of a much larger domain and therefore a large value L (L =10,000 in simulations) was used as the FPT for instances where the target isn't reached within the max_hops and $MFPT^* = (MFPT^r - L)P_r + L$ where $MFPT^r$ is the $MFPT$ given the target is reached. This simulation was repeated for all 48 bias-source-target combination for each NP initialisation. The best curve fit for each scenario was obtained using non-linear least squares method for the custom equation $MFPT = N(A - Br^D)$ following work of [8] (Eq. 4.3) where N is the number of nodes in the network, r is the distance between S and T and (A, B, D) being variables. From the achieved best fit, dw_b was obtained using $D = dw_b - df$.

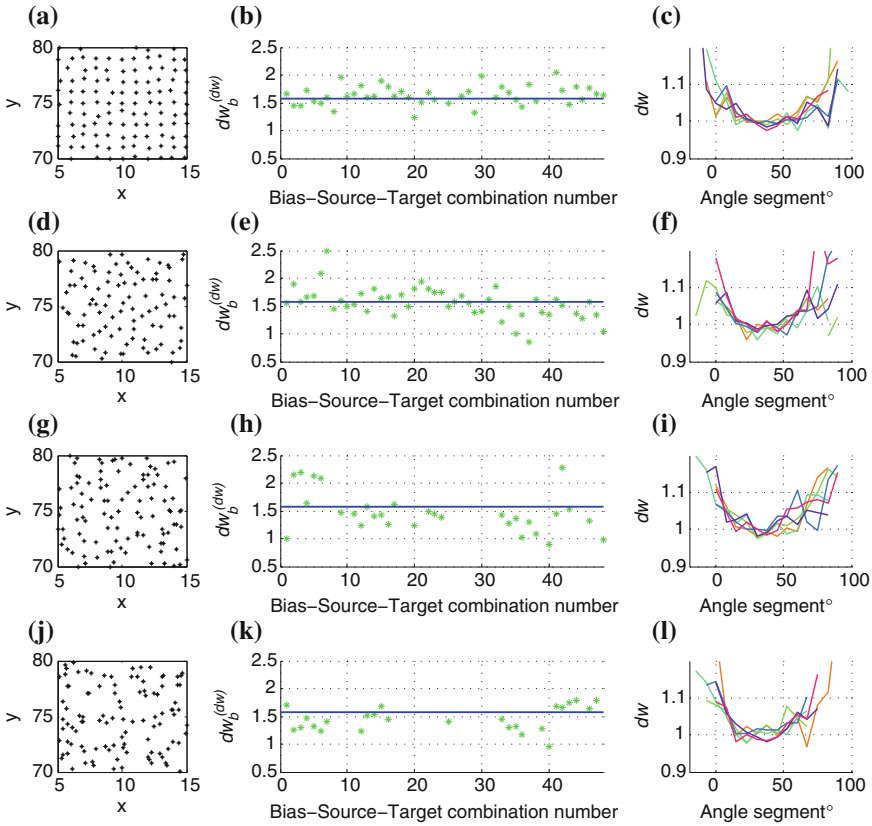


Fig. 4.2 Obtaining a relationship between dw and dw_b. a, d, g, j Samples of networks having noise levels of 0.1, 0.3, 0.5 and 0.7 respectively. **b, e, h, k** The relationship of dw_b to dw obtained empirically, for network on the *left*, through comparing the values calculated for dw with the optimal dw_b. The values are taken from 48 Bias-Source-Target combinations for each network. **c, f, i, l** Dependence of dw on the relative angle of the line connecting the source and target to the direction of bias for 6 initialisations of networks with noise levels same as leftmost figures in each row

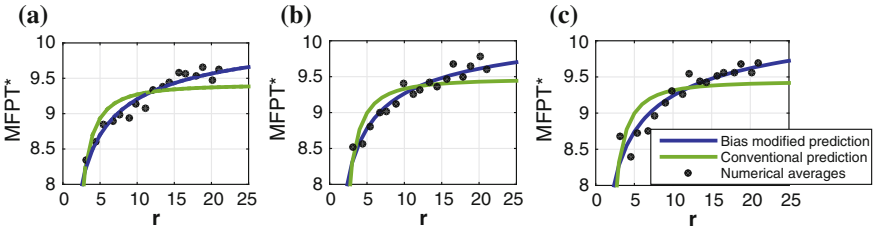


Fig. 4.3 The average *FPT*s compared to theoretical predictions for networks with **a** $U = 1.2$ **b** $U = 1.0$, and **c** $U = 0.8$. The r-squared value sets obtained for predictions using conventional and bias-modified variables for the three cases were [0.743, 0.941], [0.733, 0.932], and [0.607, 0.881] respectively

The values for $(dw,\ dw_b)$ for all scenarios were compared and for each a relationship was observed to follow an inverse power function of the form

$$dw_b \propto C^{1/dw} \tag{4.6}$$

where C is a constant. The constant C was calculated empirically as being approximately $\pi/2$ for all tested scenarios (Fig. 4.1g, Fig. 4.2) thus producing

$$dw_b = \left(\frac{\pi}{2}\right)^{1/dw} \tag{4.7}$$

as a generic relationship.

A few comparisons of *MFPT* predictions are shown in Fig. 4.3 while Fig. 4.4 shows a set of comparisons for a range of noise levels as well as bias scenarios. When compared to the deterministic counterparts, the figures prove how the modified transport variables facilitate far superior predictions. In the next section the presented prediction model for predicting *MFPT* is validated using a case study with past data of cyclone tracks over the NA Ocean.

4.4 Case Study: Calculating the Expected Arrival Time for Cyclone Motion Modelled as Biased Random Walks

This section presents a case-study using an archived dataset of past cyclone tracks over the North Atlantic (NA) Ocean.

4.4.1 Background

Tropical cyclones (TC) are one of the most destructive types of natural disasters. An estimated 1.9 million TC related deaths have been recorded according to [29], indicating the extent of damage caused by TC activity. Accurate forecasts of TC track, intensity, structure and TC induced rainfall are of extreme importance in terms of prevention and mitigation of associated losses. Forecasting cyclone activity is an area plagued with complexity. Cyclone activity is governed by a range of factors including but not limited to wind pressure, sea surface temperature, Coriolis effect, ocean currents and tropical temperature profiles [7]. Combining these parameters with basin specific factors, climate change effects and other factors to facilitate accurate and efficient numerical predictions is indeed a formidable task. The complexity

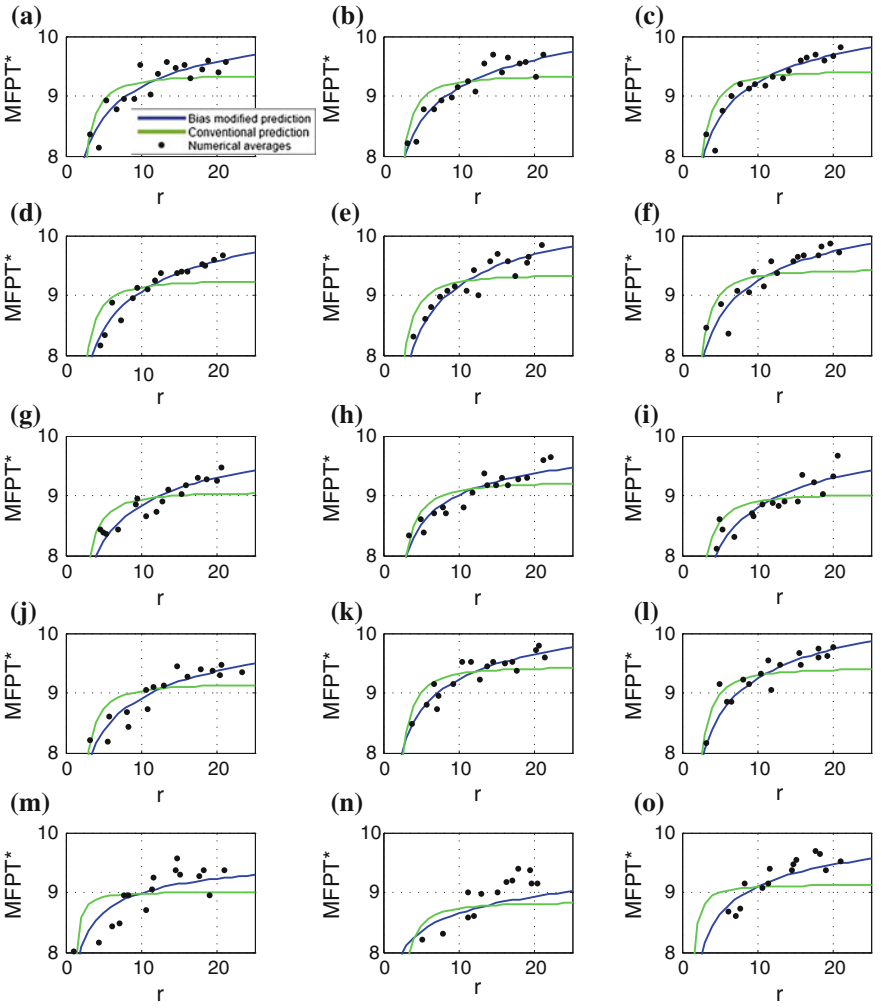


Fig. 4.4 Comparisons of *MFPT* prefictions for random walks starting from a fixed source S to reach a set of targets having the same relative angle to S but at increasing distances from S for five different networks with respective noise levels of 0.1, 0.3, 0.5, 0.7, 0.9 are represented in each row of sub figures. The *three columns* of sub figures correspond to bias intensities of 0.8, 1.0 and 1.2. The *bold dots* show the numerically obtained mean first passage times using Monte-Carlo simulations compared to the theoretical *curve* fits using conventional and bias modified transport variables. The r-squared value sets obtained for curve fitting using conventional and bias-modified variables for **a–o** are ([0.6302, 0.81632], [0.63888, 0.90035], [0.63348, 0.90279], [0.64693, 0.95868], [0.73777, 0.92564], [0.48401, 0.8083], [0.83784, 0.92949], [0.61473, 0.86012], [0.81793, 0.8981], [0.52259, 0.84617], [0.57366, 0.84689], [0.65008, 0.82429], [0.52259, 0.84617], [0.57366, 0.84689], [0.65008, 0.82429], [0.35366, 0.7097], [0.38286, 0.58259], and [0.40216, 0.78947]) respectively

increases further when TC modifies its environment and thus induces feedback to the system [30].

Forecasting TC tracks have shown continuous improvements in skill levels resulting from rigorous efforts during the past several decades [31, 32]. These advancements are greatly due to improvements in numerical weather prediction (NWP) models, including improvements in resolution, methods of initialisation, better physical parameters, and more skill in models resulting in consensus forecasts [28]. Cyclone track prediction models have shown much higher skill improvements with respect to other aspects of prediction such as cyclone induced rain, cyclone intensity and structure. However there still exists instances where models' sensitivity to initial conditions and model error could lead to considerable deviations in track forecasts [33] between ensemble members of the same model as was the case in predicting cyclone Sandy in 2012 according to [30].

An estimate of the mean first passage time ($MFPT$) for a cyclone to reach a known location is a useful metric for disaster planning operations. A prediction of the expected time for a cyclone to reach any target location, can compliment the cyclone forecasts of track, intensity, structure and rainfall. Rumpf et al. [34] have shown that TCs can be statistically modelled as random walks. This case study presents a novel approach of predicting the $MFPT$ for the eye of the cyclone to reach any known location (given in latitude and longitude coordinates) using the coupled properties of the of the random walker (spatial motion of the eye of the cyclone) and the network (set of states the walker can occupy) as described earlier in this chapter. These transport properties are used to predict the expected arrival time at any target location (T) given the knowledge of the current position (S) of the TC. Although there has been debate over methods of determination of the centre of a cyclone [35], such errors are neglected for this case study. The transport properties can be obtained from past data, and simulations using physical modelling. For this case study only past data is used for simplicity but it is worth noting that the accuracy of the model increases with the knowledge of the system.

Ideal conditions required for many $MFPT$ prediction methods found in literature are rarely met when random walks are results of state transitions for natural systems such as motion of cyclones. These walks are commonly biased towards attractors formed by non-linear interaction dynamics between the system and the environment. Given the relationship of wind and temperature through thermal wind balance, properties such as shallow convection which determine the tropical temperature profile can lead to biases in TC motion [36]. This section presents a bias modified $MFPT$ prediction model useful under such biased environments.

The spatial motion of the eye of the cyclone can be viewed as a random walk and it can be shown that the methods described earlier to predict the $MFPT$ based on coupled properties of the walker and the network is a suitable model for prediction. This method can provide useful insight into the $MFPT$ profile for a range of target points even for areas outside of the forecast cones [37] thus allowing the possibility of complex paths.

4.4.2 The Dataset

The dataset used in this section consists of best-track data from HURDAT2 [38] dataset for getting the training and validation data for the presented model. HURDAT2 is a publicly available dataset providing a record of TCs in the North Atlantic Ocean observed from 1851 to 2013, available online at http://www.nhc.noaa.gov/data/#hurda. The data of all cyclones above the NA Ocean from 1950 to 2011 are used as the training dataset for simulations in this section. The dataset consists of the observed position of the eye of the cyclone in 6 hour time intervals. For this dataset any errors possible in determining the centre of a cyclone was neglected.

4.4.3 MFPT Predictions for Case Study

With each prediction time step, the cyclone tracks going through the closer neighbourhood of the current position of the cyclone are filtered from the total dataset. The reason being the inhomogeneity in transport properties throughout the entire NA Ocean area. This inhomogeneity has also been addressed in previous work in [39] and further addressed in detail in Chap. 5. An example subset of the dataset used for predicting *MFPT* when the current location is at longitude $-35°$ and latitude $0°$ is shown in Fig. 4.5a.

Initially, the transport dimensions (df and dw_b) are calculated from the training dataset. As the state-space of all possible locations is continuous, df follows the dimension of the continuous state space (i.e. $df = 2$). The values for dw and dw_b are calculated from Eqs. (4.2), and (4.7) for angle segments of $\pi/24$ limited by the dataset (refer Sect. 4.3.3 and Algorithm 1). Figure 4.5b compares the numerically obtained average *FPT*s to reach a fixed group of targets at different distances in the direction of bias, to predictions using bias modified transport variables. The prediction gave

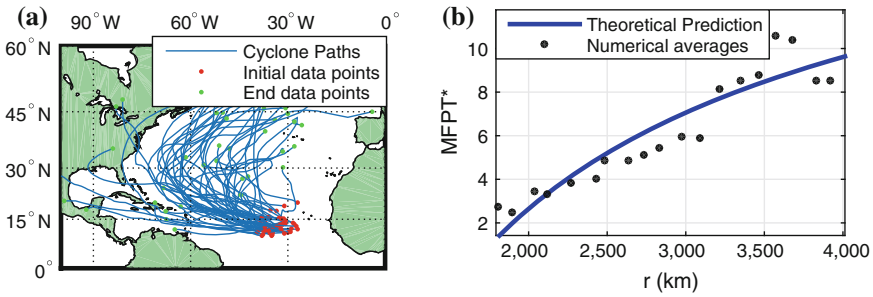


Fig. 4.5 *MFPT* prediction for TC motion over the NA Ocean. **a** Filtered set of cyclone tracks as training dataset for prediction at $S \approx [0, -35]$. **b** Comparison of numerically obtained *MFPT*s with predictions using bias modified transport variables

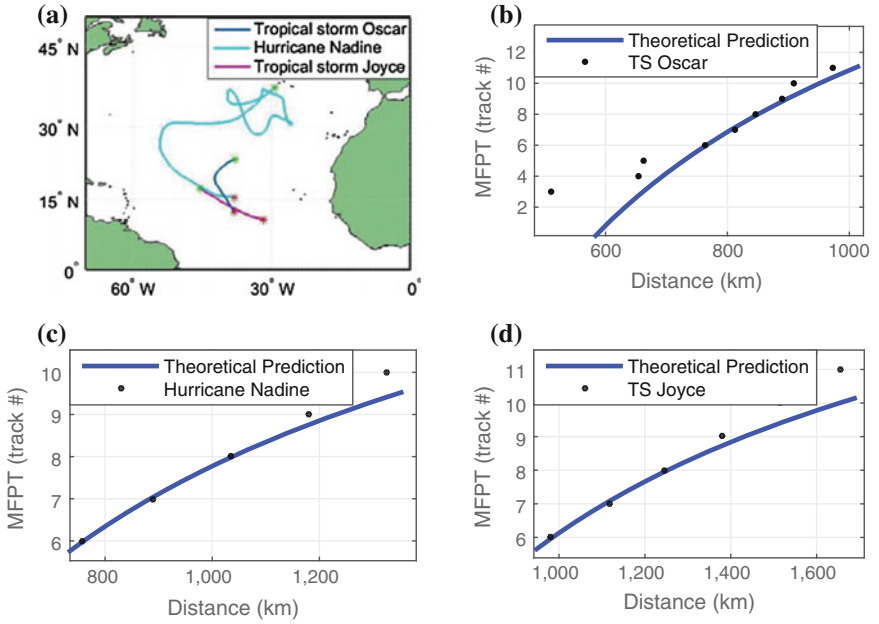


Fig. 4.6 *MFPT* predictions using the presented model compared to best-track data for **a** 3 TC's in 2012; **b** Tropical storm Oscar (for angle segment 8 $[\frac{-5}{12}\pi, \frac{-4}{12}\pi]$), **c** Hurricane Nadine (for angle segment 3 $[\frac{-10}{12}\pi, \frac{-9}{12}\pi]$), and **d** Tropical storm Joyce (for angle segment 4 $[\frac{-9}{12}\pi, \frac{-8}{12}\pi]$). Forecasts were done at time period 8 (i.e. at the 8th forecast point) for all three cases shown here

an r-square value of 0.866 with the real data indicating high goodness of fit by the presented theoretical prediction.

The calculated transport variables (using the training dataset of TC tracks between 1950–2011) are used to predict *MFPT* for comparing with predictions for three separate cyclones during the year 2012 which originated within the region considered in Fig. 4.5a. Figure 4.6 compares theoretical predictions using presented model to actual data for TC's 'Oscar', 'Nadine' and 'Joyce' in their most common segments. These prediction curves gave r-square values of 0.93, 0.96 and 0.92 respectively showing high goodness of fit with real data. The predictions were done with df and dw_b pre-calculated using the training dataset and the variables A and B calculated using a moving window of 3 data points of the current TC track data available at the time of prediction.

Prediction error profiles(normalised) for four cyclones from 2013 cyclone season are shown in Fig. 4.7 with the training dataset consisting of best track data of all cyclones between 1950–2012. Predictions were done at each available best-track location update for each cyclone. Figure 4.7 depicts these error profiles for cyclones 'Dorian', 'Humberto', 'Ingrid', and 'Gabrielle'. Finally a comparison of *MFPT* values predicted using the presented model to official forecasts from the National Hurricane Centre data archive [40] and best track data is presented. For these com-

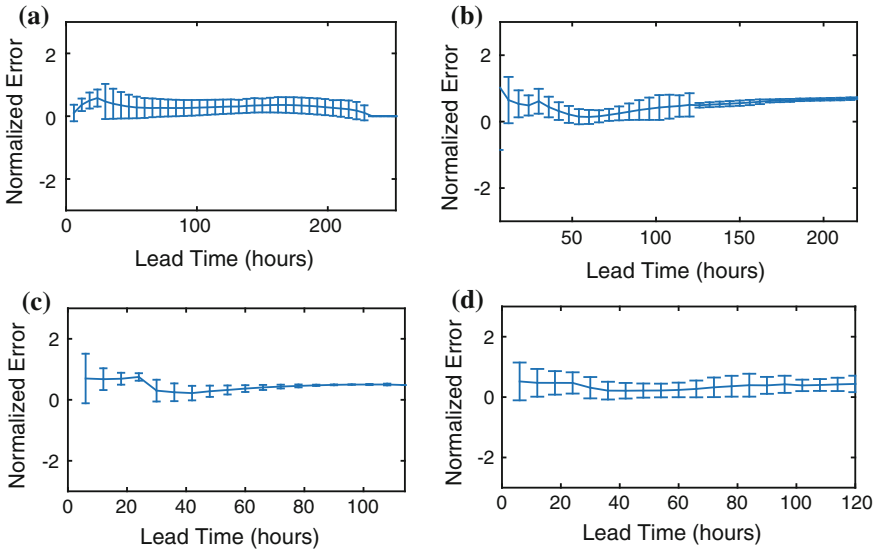


Fig. 4.7 Normalized prediction error profiles (mean and standard deviation) of four cyclones in 2013 comparing error of prediction with respect to actual FPT. **a** Cyclone Dorian **b** Cyclone Humberto **c** Cyclone Ingrid **d** Cyclone Gabrielle

parisons, it was required to find instances where a forecast overlaps (or at least get very close to) a best track data point after a re-curving or looping track. Figure 4.8 shows three such instances. Figure 4.8a depicts the track of cyclone Ingrid in a solid line with the dots showing the actual best-track positions at foretasted times. The dashed line and squares show forecast locations from the issued third forecast for Ingrid. One comparison point is shown here where the forecast point is very close to a best track location. Figure 4.8b shows two such comparisons for cyclone Humberto with the 24th official forecast. The resulting times to reach for these three comparisons for official forecast, best track data and presented prediction are **(i)** [72 hours (3 days), 96 hours (4 days), 4.1927 days] **(ii)** [48 hours (2 days), 72 hours (3 days), 3.3456 days] **(iii)** [72 hours (3 days), 96 hours (4 days), 3.9646 days] respectively.

4.4.4 Case Study: Discussion

Following the presented comparison scenarios, several observations prevail. Firstly, the high goodness of fit resulting in Figs. 4.5 and 4.6 show the suitability of the presented model for *MFPT* prediction for cyclone motion. The error profiles in Fig. 4.7 show that although the normalised error profiles appear promising, there are some inconsistencies when comparing the four graphs. In addition, especially in Fig. 4.7b, a sharp change of error standard deviation is observed. Both these effects

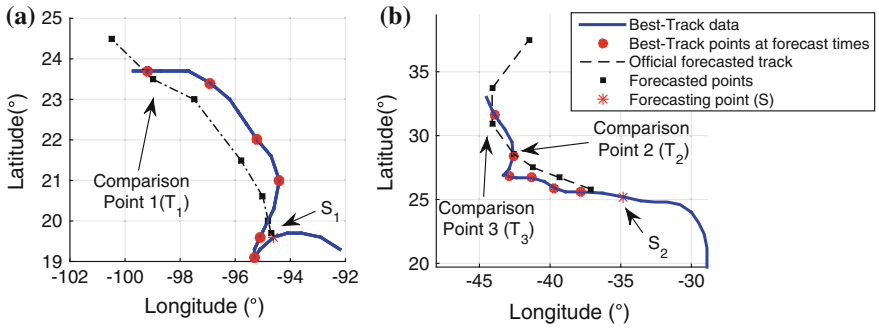


Fig. 4.8 Three comparisons of the expected time to arrive using official forecasts to best track data for **a** 3rd available forecast for cyclone 'Ingrid' and **b** 24th available forecast for cyclone 'Humberto'. The forecasts are available for lead times of [12 24 36 48 72 96 120] hours from the forecast point. The *MFPT* predictions using the presented method were higher in accuracy for the three comparisons by 20.2 %, 44.9 % and 24.1 % respectively (with respect to official forecasts)

are resulting from the limited training dataset used. The predictions are better in segments where a larger training dataset is used. One other limitation of the discussed model is that when there is zero data available for some segment, no prediction is possible. Both these issues can be addressed by increasing the quantity of high quality training data used. This can be done by incorporating as much knowledge available of the random walker and the network (i.e. using complex physics behind cyclone formation and motion and the interaction with the surrounding environment) into the training phase. With this, in addition to best track data of past cyclones, simulated data using the physical modelling of the complex processes governing the large-scale and inner-core characteristics of TCs and their interaction with the environment can be used to derive the transport variables. As this is out of the scope of this book, for now it is stressed that the skill of the model increases with knowledge of the system. Figure 4.8 presents a few instances showing how the presented model could compliment TC forecasts by estimating the expected arrival time at any location.

When the transport variables (df and dw) are calculated off-line, efficient real time predilections are available within a few seconds after each recorded TC location update, showing the models low sensitivity to computational resources in real time prediction. All simulations and analysis presented herein were done using MATLAB R2014b on an Intel(R) Core(TM) i7-4770 CPU.

4.5 Conclusions

This chapter focussed on discussing generic methods for predictions for processes governing natural disasters that state changes take place in directionally biased networks. Following a short discussion on the current state of art in *MFPT* prediction

for biased random walks, a case invariant modification of transport variables is described as a viable solution to the problem of addressing directional bias. With pre-processed transport variables, real-time prediction is shown to be less dependent on computational resources and predictions are available as soon as location updates are received. The methods described here are very generic and therefore indicates wider applicability.

The dependence of the prediction accuracy on the size of the training dataset as seen in some case studies can be addressed by including various environmental conditions or storm characteristics in terms of complex physical modelling of TCs in calculating the transport variables. This is outside the scope of this book and is left for future work. Another limitation of the transport variable based method is that it is only capable of handling directional bias. In a non linear manifold type bias for example, where there is bias which can not be modelled in Euclidean state space, the discussed method is not applicable. The TPM based method presented in Chap. 3 is more suitable for such random walks.

Finally, it is worth noting that although the methods presented in this chapter were verified for 2D networks, they are equally applicable to random walks on higher dimension state spaces.

Several important conclusions can be made from the material discussed in this chapter. Firstly it was shown that a large class of processes that translate to state space random walks show direction dependent bias. Then the ability was shown to incorporate directional bias into *MFPT* calculations through bias modifying transport variables used in the calculations. The comparisons showed that bias modified transport variables show higher prediction accuracy (for directionally biased random walks) when compared to methods assuming isotropic motion of random walks. Simulations using the real dataset of archived cyclone track data verify the presented methods. Further, it was shown that an estimation of *MFPT* to reach a set of target locations compliments cyclone forecasts allowing disaster preparation for areas even outside of the track forecast cones.

Following these conclusions, the next chapter concentrates on addressing network inhomogeneity in *MFPT* calculations for natural disaster systems.

4.6 Summary

This chapter discussed the problem of addressing network specific bias in predicting *MFPT* for processes governing natural disaster activity. A novel method of predicting the *MFPT* for a directionally biased random walker to reach any target state is presented. The transport variables df and dw are modified to account for the effects of bias. A generic expression is obtained that represent a relationship that allows direct computation of bias modified

walk dimension. This relationship is tested with simulated random walks on many hypothetical networks with varying noise levels and bias conditions.

A comprehensive case study is presented using the archived dataset of cyclone tracks over the NA Ocean since 1950. The motion of TC's are treated as biased random walks and the joint transport properties of the random walker and the network of locations the walker can occupy are used to predict the *MFPT* for the walker to reach any target point. The transport properties are bias-modified to account for the potential fields that shape the random walks. Comparisons with official forecasts demonstrate the usefulness of the predictions in complimenting forecasts. The model's ability to predict the expected time of arrival at any given location gives valuable information for expected lead times even for areas outside of the forecast cones even allowing the possibility of complex paths. This knowledge compliments the official cyclone track forecasts which are generally available 3 hours after the best-track location updates.

Appendix 1: A Fire Spread Model Using Huygens Principle

Most simulated random walks used in this book had probability density distributions to select the next step based on Huygens principle (HP) as it is generally known to be a universal model of propagation [27] for natural environments. It has been shown that HP applies to any propagation phenomena which can be described through explicit linear differential and difference equations.

The two-dimensional network with dynamic connectivity described in Chaps. 4 and 5 was inspired by a two-dimensional deterministic fire growth model named FARSITE [26]. The section of FARSITE model that describes surface fire spread as the base was used for obtaining the probability density distribution (P_{DD}) for identifying the next hop of a random walk. The main approach is to apply HP at each vertex at the spread perimeter to shape and orient an elliptical wavelet at each time step. The size is determined by the spread rate and the length of a computation time step and the shape is determined by the direction and effective intensity of bias. The LB (Length-to-Breadth ratio) an HB (Head-to-Back ratio) for the FARSITE model are found using

$$\begin{aligned} LB &= 0.936e^{(0.2566U)} + 0.461e^{(-0.1548U)} - 0.397, \\ HB &= (LB + (LB^2 - 1)^{0.5})/(LB - (LB^2 - 1)^{0.5}) \end{aligned} \tag{4.8}$$

formulated empirically [26]. In the simulations, the wavelet was converted into a probability density distribution, thus reducing the spread to a random walk.

References

1. Anderson HE (1983) Predicting wind-driven wild land fire size and shape. US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden
2. Achtemeier GL, Goodrick SA, Liu Y (2012) Modeling multiple-core updraft plume rise for an aerial ignition prescribed burn by coupling daysmoke with a cellular automata fire model. Atmosphere 3(3):352–376
3. Asante K, Macuacua R, Artan G, Lietzow R, Verdin J (2007) Developing a flood monitoring system from remotely sensed data for the Limpopo basin. IEEE Trans Geosci Remote Sens 45(6):1709–1714
4. Bates PD, Marks KJ, Horritt MS (2003) Optimal use of high-resolution topographic data in flood inundation models. Hydrol Process 17(3):537–557
5. Zerger A, Wealands S (2004a) Beyond modelling: linking models with GIS for flood risk management. Nat Hazard 33(2):191–208
6. Aparicio JP, Pascual M (2007) Building epidemiological models from r0: an implicit treatment of transmission in networks. Proc Biol Sci 274(1609):505–512
7. Roy C, Kovordányi R (2012) Tropical cyclone track forecasting techniques-a review. Atmos Res 104:40–69
8. Condamin S, Benichou O, Tejedor V, Voituriez R, Klafter J (2007b) First-passage times in complex scale-invariant media. Nature 450(7166):77–80
9. Song C, Havlin S, Makse HA (2006) Origins of fractality in the growth of complex networks. Nat Phys 2(4):275–281
10. Song C, Havlin S, Makse H (2005) Self-similarity of complex networks. Nature 433(7024):392–395
11. Nicolaides C (2011) Anomalous transport in complex networks. Thesis
12. Perra N, Baronchelli A, Mocanu D, Gonçalves B, Pastor-Satorras R, Vespignani A (2012) Random walks and search in time-varying networks. Phys Rev Lett 109(23):238–701
13. Starnini M, Baronchelli A, Barrat A, Pastor-Satorras R (2012) Random walks on temporal networks. Phys Rev E 85(5):056–115
14. Fronczak A, Fronczak P (2009) Biased random walks in complex networks: the role of local navigation rules. Phys Rev E 80(1):016–107
15. Lee ZQ, Hsu WJ, Lin M (2014) Estimating mean first passage time of biased random walks with short relaxation time on complex networks. PLoS One 9(4):e93–348. doi:10.1371/journal.pone.0093348
16. Skarpalezos L, Kittas A, Argyrakis P, Cohen R, Havlin S (2014) Efficiency of message transmission using biased random walks in complex networks in the presence of traps. arXiv preprint arXiv:14062437
17. Wijesundera I, Nanayakkara T, Halgamuge MN, Nanayakkara T, Mean first passage time for cyclone motion modelled as biased random walks. Under Review
18. Redner S (2001) A guide to first-passage processes. Cambridge University Press, Cambridge
19. Condamin S, Benichou O, Moreau M (2005) First-passage times for random walks in bounded domains. Phys Rev Lett 95(26):260–601
20. Tejedor V (2012) Random walks and first-passage properties. Thesis
21. Tejedor V, Benichou O, Voituriez R (2011) Close or connected: distance and connectivity effects on transport in networks. Phys Rev E 83(6):066–102
22. Ben-Avraham D, Havlin S (2000) Diffusion and reactions in fractals and disordered systems. Cambridge University Press, Cambridge
23. Shlesinger MF (2007) Mathematical physics: first encounters. Nature 450(7166):40–41
24. Godec A, Metzler R (2015) Optimization and universality of brownian search in quenched heterogeneous media. arXiv preprint arXiv:150300558
25. Lau HW, Szeto KY (2010) Asymptotic analysis of first passage time in complex networks. EPL Eur Lett 90(4):40–005
26. Finney M, Station RMR (1998) FARSITE, Fire area simulator-model development and evaluation. Us department of agriculture, forest service, rocky mountain research station, Ogden

27. Peter E (1996) Huygens' principle and the modelling of propagation. Eur J Phys 17(4):226
28. Hamill TM, Whitaker JS, Fiorino M, Benjamin SG (2011) Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble kalman filter. Mon Weather Rev 139(2):668–688
29. Shultz JM, Russell J, Espinel Z (2005) Epidemiology of tropical cyclones: the dynamics of disaster, disease, and development. Epidemiol Rev 27(1):21–35
30. Torn RD, Whitaker JS, Pegion P, Hamill TM, Hakim GJ (2015) Diagnosis of the source of gfs medium-range track errors in hurricane sandy (2012). Mon Weather Rev 143(1):132–152
31. Rappaport EN, Franklin JL, Avila LA, Baig SR, Beven JL, Blake ES, Burr CA, Jiing JG, Juckins CA, Knabb RD, Landsea CW, Mainelli M, Mayfield M, McAdie CJ, Pasch RJ, Sisko C, Stewart SR, Tribble AN (2009) Advances and challenges at the national hurricane center. Weather Forecast 24(2):395–419
32. Soloviev AV, Lukas R, Donelan MA, Haus BK, Ginis I (2014) The air-sea interface and surface stress under tropical cyclones. Sci Rep 4
33. Lai Z, Hao S, Peng S, Liu B, Gu X, Qian YK (2014) On improving tropical cyclone track forecasts using a scale-selective data assimilation approach: a case study. Nat Hazard 73(3):1353–1368
34. Rumpf J, Rauch E, Schmidt V, Weindl H (2006) Stochastic modeling of tropical cyclone track data. In: 27th Conference on Hurricanes and Tropical Meteorology, AMS, 5A–6
35. Chaurasia S, Kishtawal CM, Pal PK (2010) An objective method of cyclone centre determination from geostationary satellite observations. Int J Remote Sens 31(9): 2429–2440
36. Torn RD, Davis CA (2012) The influence of shallow convection on tropical cyclone track forecasts. Mon Weather Rev 140(7):2188–2197
37. Meuel T, Prado G, Seychelles F, Bessafi M, Kellay H (2012) Hurricane track forecast cones from fluctuations. Sci Rep 2
38. Chris Landsea JF, Beven J (2014) Atlantic hurricane database. HURDAT2: 1851–2014
39. Rumpf J, Weindl H, Höppe P, Rauch E, Schmidt V (2007) Stochastic modelling of tropical cyclone tracks. Math Method Oper Res 66(3):475–490. doi:10.1007/s00186-007-0168-7
40. Center NH (2015) Tropical cyclone advisory archive. http://www.nhc.noaa.gov/data/

Chapter 5
Calculating *MFPT* for Processes Mapping into Random Walks in Inhomogeneous Networks

Abstract Dynamic processes leading to natural disasters often translate into random walks in state spaces which are inhomogeneous in transport characteristics. In other words, such random walks will behave differently in different parts of a network which would have different values for transport properties (df and dw if using methods from Chap. 4). Thus, for such networks the application of *MFPT* calculation methods introduced in Chap. 4 along with many other methods described in literature are not straight forward. This chapter proposes that using the novel concept of dividing the node distribution into patches/clusters known as network primitives (NPs) where all nodes within each primitive share common transport variables, and adopting a 'hop-wise' approach to calculate *MFPT* between any source and target pair as an extension to the methods described under Chap. 4, can be a viable solution for predicting *MFPT* for random walks in inhomogeneous networks. This methodology's potential is demonstrated through simulated random walks and with a case study using the dataset of past cyclone tracks over the North Atlantic Ocean. The predictions using the presented method are compared to real data averages and predictions assuming homogeneous transport properties.

5.1 Introduction

Majority of natural real world systems have been proven to show strong variability in transport properties [1]. This inhomogeneity affects the many stochastic processes which occur on that network such as random walks.

Network inhomogeneity generally increases the network analysis complexity. However, some applications such as random search processes also find advantages in network inhomogeneity through higher rates of arrival at the target [2]. They have shown that in optimal inhomogeneous search tests, *MFPT* on such networks are completely dominated by those trajectories heading directly towards the target.

Network inhomogeneity has been defined in various ways in literature. The most common definition is the inhomogeneity in terms of node degree distribution [1, 3, 4]. One of the most common such network type that is studied in literature are

I. Wijesundera et al., *Natural Disasters, When Will They Reach Me?*,
Springer Natural Hazards, DOI 10.1007/978-981-10-1113-9_5

the Scale Free (SF) networks. These are networks which show a power-law degree distribution [1].

Network inhomogeneity is an issue common for a range of real world applications. This inhomogeneity is defined in this chapter as the spatial inhomogeneity in transport properties (e.g. node density, ease of propagation) experienced by a random walker on that network. This chapter discusses the effect the inhomogeneity of transport characteristics have on *MFPT* calculations on such networks. It specially focusses on the fact that many such inhomogeneous networks consists of *homogeneous patches* in developing methods to handle this inhomogeneity in *MFPT* calculations. A discussion on how to identify such homogeneous partitions is followed by an extension of MFPT calculation methods from Chap. 4 to accommodate network inhomogeneity.

Some real world examples of inhomogeneous networks (on which random walks take place) are the atmosphere on which a cyclone would travel through [5], a terrain on which flood will propagate through [6], an area where fire will spread through [7, 8], or even the density of people through which a disease will spread [9]. In the real world, the networks representing the mobility pattern of individuals among different subpopulations are in many cases highly inhomogeneous [10]. In animal motion it has been shown that the calculation for expected time needed for a predator to locate small patches of prey in a 2-D landscape has two components; random and directed [11]. Although animal motion itself is isotropic they are possibly spatially inhomogeneous due to inhomogeneity in landscape. Random walks in all these scenarios will go through different portions of a network where the network itself will behave in such different ways making the estimation of arrival time at a given destination extremely difficult to calculate as a whole.

5.1.1 Random Walks on Inhomogeneous Networks

Random walks are among the basic stochastic processes which are affected by inhomogeneity of a network. As discussed earlier in this book, random walks have been studied extensively for decades on regular lattices, fractal networks, and many other specialised networks [1]. Much work has considered network nodes as *well-mixed homogeneous populations* [9, 10, 12]. Although there exists many mathematical methods that produce excellent results on estimating *MFPT* on different types of networks, a common requirement is that every node in the network shares some transport properties that are used in *MFPT* calculations. In other words, the properties of the network and the random walker need to be length-scale invariant. Most of these methods share the necessity that some transport properties that describe how a random walker would move through the network are common at all nodes of the network. For example, the method presented in [13] works only for networks where df and dw remain spatially homogeneous.

5.1.2 Current Trends in MFPT Estimation in Inhomogeneous Networks

Predicting *MFPT* in inhomogeneous networks have been given much attention specially for networks which possess characteristics such as small-world property, and scale-free property [4]. The findings in [14] show that the *MFPT* depends on source-target distance and the degree-distribution for scale-free networks. They have also shown significant differences of these properties when the networks show non-compact explorations as opposed to compact explorations (Chap. 4). The transit and commute times tend to diverge when the network consists of more than one cluster because the graphs are not connected (this is presented in [4] with respect to *MFPT* calculations on an Erdös-Renyi random graph).

A common approach of addressing network inhomogeneity is through providing a global result which is the average of *MFPT* over a set of starting points distributed uniformly over all the other nodes of the graph [1, 2]. A large focus has been set on scale-free networks [2]. The reason being that there are a considerable class of networks which are shown to satisfy the scale-free condition. However, the focus of this chapter is on estimating the *MFPT* for random walks on networks that are inhomogeneous in transport properties and which are not necessarily scale free.

Earlier in this book, a method of reducing the calculation complexity by dividing the problem of predicting flood propagation through identifying 'Geographic Primitives' based on the terrain slope profile was presented (Chap. 3). However in this chapter, a somewhat different approach is taken [15] following a more generic method to identify primitives using the transport variables df and dw introduced in Chap. 4. Using the knowledge that many inhomogeneous networks commonly include frequent homogeneous patches (e.g. fuel distributions for fire propagation, distribution of cities for disease propagation, etc.), it is understood that if it is possible to divide the network into portions where all nodes within that sub-network have (approximately) equal values for df and dw, the complexity of the problem would reduce considerably as *MFPT* prediction between any two points within such network portion would be straight forward using methods from Chap. 4. With this knowledge, the concept of network primitives (NP) is introduced.

5.1.3 Network Partitioning

Following the above discussion, the focus is now brought on to calculating *MFPT* through addressing network inhomogeneity by dividing the network into simpler portions: that is to follow a divide-and-conquer method to reduce the complexity of the problem. The next question is which properties of a network should be considered when trying to divide the network. There are numerous ways in which a network could

be divided by. The work in [2] has considered the network as two concentric regions with piece-wise constant diffusivity. Ferguson et al. has shown in [16], many ways in which an inhomogeneous network can be looked at in terms of epidemic transmission. The most basic method is to assume homogeneous mixing. However, this suppresses many features of the transmission and does not give adequate intelligence of the system. The first method of network division proposed was by age/social structure to account for different probabilities of contact between these subgroups. The next was to identify a static network structure. But this has been shown to be difficult if not impossible for many systems [9]. The fourth method is to identify homogeneous patches in the network. Such patch, or a subpopulation is generally a natural unit of study. Although deterministic models which give exact solutions provide rapid simulations, it is extremely important to account for stochasticity in transmission events [16].

The patch identification method is given more emphasis in this chapter for the purpose of dividing an inhomogeneous network into homogeneous patches. This chapter proposes that if the network is divided into patches homogeneous in the transport properties used to calculate *MFPT* within a patch, that would be a viable solution for calculating *MFPT* for random walks in networks that are inhomogeneous in transport properties. A hop-wise approach can be followed in obtaining the *MFPT* for any predefined source-target set on an inhomogeneous network divided into homogeneous 'primitives'. The paths could be weighted to get the final path-integral representation of random walk properties.

Assumptions

Some assumptions that are used in the methods discussed in this chapter are as follows:

- Random walks show only directional bias with one prominent bias direction (see Sect. 5.2.2).
- Inhomogeneous networks consist of homogeneous patches.
- Bias (direction and intensity) is static during any given prediction interval.
- There are no loops in biasing vector fields.

In addition to these, the assumptions used in Chap. 4 are also used in this chapter. Refer to Sect. 4.2.4 for details.

5.2 Network Primitives (NPs)

In the context of transport property based *MFPT* calculations for an inhomogeneous node distribution, an NP can be technically defined as follows:

An NP ($\mathfrak{N}$) is defined as a subset cluster of a network with ν governing transport variables where $\forall\ \zeta \subseteq \mathfrak{N},\ \exists\ \Delta_i\ s.t.\ \sigma_i < \Delta_i, i = 1, 2, \ldots, \nu$ where σ_i is the standard deviation of the transport variable d_i for all nodes in ζ and Δ_i is sufficiently small.

In other words, an NP is a subset/patch within a given network where all nodes within an NP are homogeneous in transport properties (i.e transport variable variation below a small threshold)

Since in this chapter follows the *MFPT* computation methods presented in Chap. 4, the governing transport properties are *df* and dw (or dw_b in the existence of directional bias). Therefore to identify NPs, the only necessity is that the *df* and dw distributions are known for the network. These can be either calculated directly from the network structure or from prior knowledge either in the form of prior data of random walks on the network or knowledge which can help in simulating random walks on that network structure. Calculating these properties was discussed in detail in Chap. 4. It is also worth noting at this point that if a different method is used for computing *MFPT* within an NP, transport properties used in that particular calculation method should be used in identifying the NPs.

For the calculation of *MFPT* over several NP's to be possible, the NP identification process uses the distributions of *df* and dw over all nodes of that network. One main requirement is that every node in the network should belong to an NP. It is also understood that the smaller the NP's are, the higher the complexity of calculation would become. Therefore, one aim in an NP identification process should always be to identify NPs to be as large as possible.

5.2.1 Hop-Wise MFPT Estimation

As the final prediction output, getting back to the bigger picture, the *MFPT* needs to be calculated for a random walker initially at source S to reach a known target T, irrespective of which NPs the S and T belongs to. A hop-wise calculation approach is adapted to make this possible. Figure 5.1 shows a network divided into 15 NPs on which a random walk takes place. Since the walk crosses many NP boundaries, first passage time calculation can be broken down to n number of hops as shown in Fig. 5.1. This is possible because *MFPT* for any one hop can be calculated using methods described in Chap. 4. The final *MFPT* can be obtained for this particular hop sequence from:

$$MFPT = \hat{T}_1 + \hat{T}_2 + \cdots + \hat{T}_n$$

where T_i is the *MFPT* to reach the NP exit point at the ith hop. In a real prediction scenario, as there would be many possible exit points for a random walker to leave

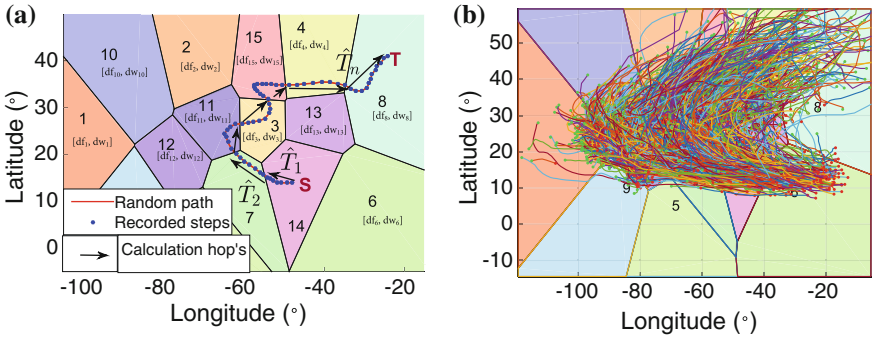


Fig. 5.1 Hop-wise *MFPT* calculation. **a** An example calculation of *MFPT* to reach a known target (T) over several NPs. **b** The dataset for this example consists of the cyclone tracks over the North Atlantic Ocean between 1950 and 2012 (Sect. 5.2.3.1)

its current NP, the hop-wise prediction is done using parallel processing of many exit points weighed by a probability density distribution (P_{DD}) which is covered in detail later in the chapter. The final prediction of MFPT would be

$$< T >= \sum_{i=1}^{n} w_i MFPT_i \tag{5.1}$$

where n is the number of alternate paths considered and w_i is the probability of path i which is calculated from the training dataset as will be discussed later in the chapter.

The calculation complexity of the hop-wise estimation is greatly reduced by identifying NPs to be convex in shape by reducing multiple cross-over points between two adjacent NPs, and eliminating enclosed NPs and possible infinite loops in calculation. The next section discusses the steps followed in developing an algorithm to identify NPs such that the following aims (as summarised below) are achieved.

NP identification : Aims

- All nodes belonging to any given NP should be homogeneous in df and dw values (having a standard deviation below a given threshold).
- Every node should belong to one and only one NP (total coverage and no overlapping). $\underset{i,j}{\forall} NP_i \cap NP_j = \emptyset$. At this point the NPs are limited to be non-overlapping to reduce the complexity of identifying NP exit points.
- NPs should be convex (for unique exit points per hop).
- It is desirable to have larger NPs (to reduce calculation complexity).

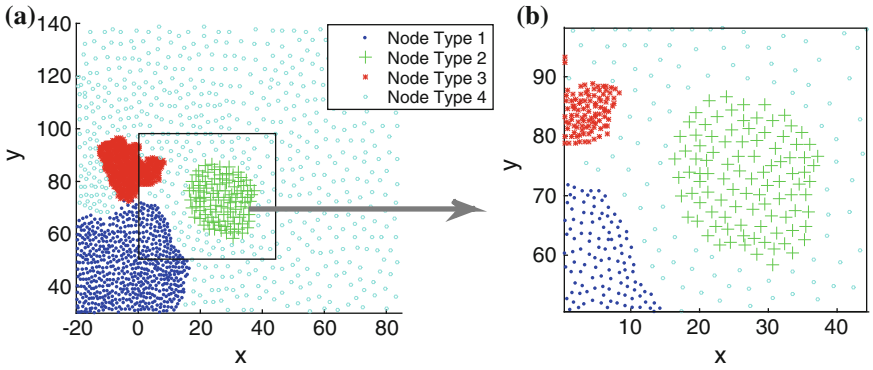


Fig. 5.2 'Network Y': Node distribution. **a** The extended network used for calculating the transport properties. **b** Portion of Network Y considered for MFPT calculations

The following sections detail the steps followed in identifying NPs taking a generic two dimensional node distribution. Nevertheless, these concepts are applicable to networks of higher dimensions as will be discussed later in the chapter.

A hypothetical 2D node distribution, hereafter called *network Y* is introduced in order to discuss the network identification process (Fig. 5.2). For simulated random walks, since df and dw calculations are sensitive to the edges of the considered portion of the network (with non reflecting boundaries), a margin around this network section is considered in the calculations (Fig. 5.2a). The transport properties for this generic discrete 2D network are calculated at each node similar to the calculation in the hypothetical networks in Chap. 4 Sect. 4.3 with a consistent and uniform bias at an angle of 45° clockwise to the north assumed. The only difference in the calculation is that as multiple node types are assumed (shown in different colours/symbols in Fig. 5.2), the R value (Rate of spread) of these node type differ from each other. This in turn changes the P_{DD} for selecting the next step for each node type thus making it an anomalous walk [17]. This is a common characteristic in many naturally occurring random walks and can be visualized with the example of fire propagation where depending on the fuel type (e.g. short grass, timber grass, short brush, dormant brush, hardwood litter, etc. [18]), the rate of spread differs. Figure 5.3 shows an example random walk and P_{DD} at each step. From this it is clear how the P_{DD} changes with node type. The rest of this discussion assumes that the values of df and dw (or dw_b in directionally biased media) are known for every node in the network.

5.2.2 Identifying Homogeneous NPs

The most important step of the hop-wise *MFPT* calculation process is identifying the different hypothesis that exists as 'NP types' in a given network with known df

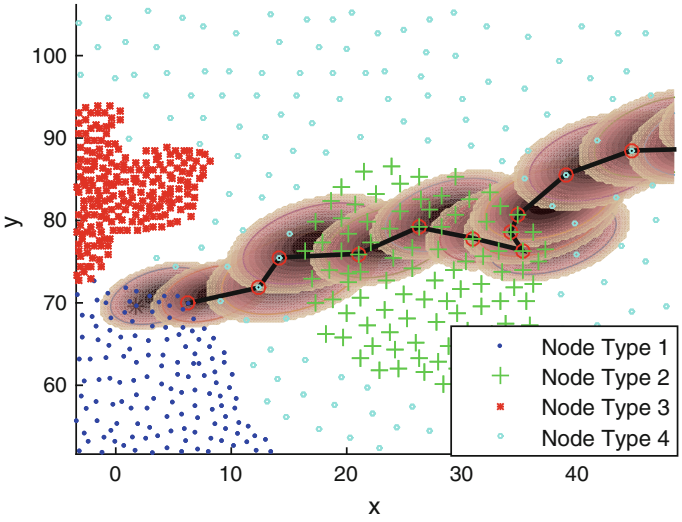


Fig. 5.3 The effect of node type on probability density distribution for next step selection. The footprint of P_{DD} differs for different node types

and dw distributions, and which type each node is of. A hypothesis or NP type is a type of predefined behaviour at each node and in this case is defined by a unique pair of transport variables $H_i = [df_i, dw_i]$. Depending on the complexity of the network, different hypotheses can be identified either by manually observing the distributions of df and dw or by using model selection techniques such as Gaussian Mixture Models (GMM) on the same [19]. The final output will be a limited set of NP types (η) where each type is identified by a unique vector $H_i, i \in [1, 2, \ldots, \eta]$. The calculation efficiency of the presented hop-wise *MFPT* prediction method is higher for networks with a smaller set of NP types (which would result in a smaller NP set).

Similar to methods described in Chap. 4, the dw is a function of the relative angle (with respect to bias) of the walk. Therefore in order to identify possible hypotheses for NP type identification process, it is required to select one dw value for each node. Using the highest probable routes, the dw used for NP identification is chosen to be in the direction of the bias. The calculated distributions of df and dw for network Y are shown in Fig. 5.4. One limitation of this method is that when dw_b distribution with θ has many separate peaks, the effectiveness is lowered as prominence is only given to one peak.

A common method often used for data clustering when clusters are of different sizes and correlation is by using the Gaussian Mixture Model (GMM). It is a parametric probability density function represented as a weighted sum of Gaussian component densities [19, 20]. GMM uses an iterative algorithm to select components to maximise posterior probability. GMM is used for hypothesis selection for

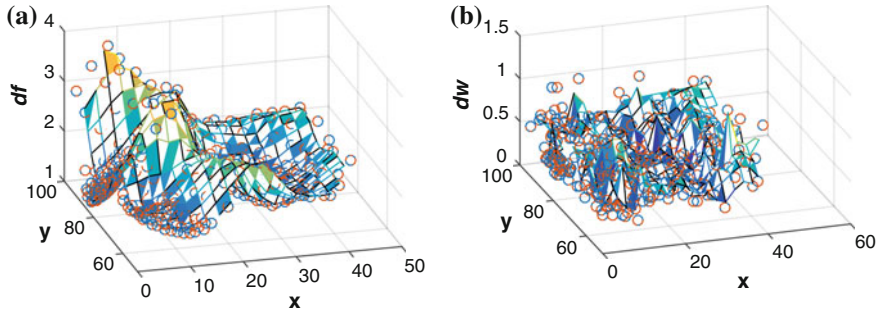


Fig. 5.4 The transport property distributions **a** df and **b** dw for network Y, with respect to the coordinate location of the network nodes

network Y where the number of hypothesis (η) is selected using both Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to avoid any error from underestimation or overestimation [21]. Figure 5.5 shows these values obtained with different number of components used in using GMM for hypothesis selection. A number of 4 NP types ($\eta = 4$) were selected considering these results and desired simplicity of real-time computation. Figure 5.6 shows the combined distribution of resulting Gaussians on top of the discrete (df, dw) distribution.

For network Y ($\eta = 4$), the hypotheses are defined as $H_1 : H_4$. The NP type at each node was calculated using the following steps.

1. Identify the set of η number NP types/Hypotheses using GMM (i.e. identify the values of df and dw for each NP type/hypothesis).

$$\underset{i=1:\eta}{\forall} H_j = [df_j, dw_j] \tag{5.2}$$

2. Calculate likelihood of each node belonging to each hypothesis given the df and dw at each node. (Denoted as $P(H_i|dw)$ and $P(H_i|df)$; $i = 1 : \eta$)

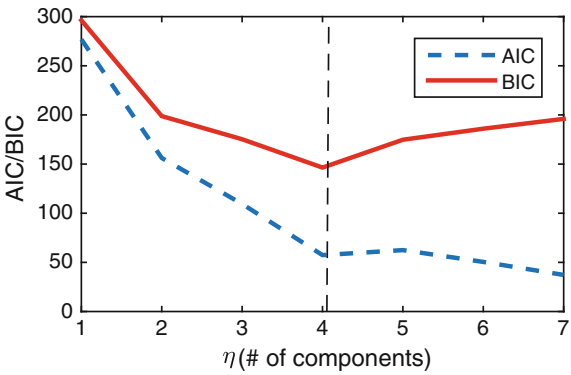


Fig. 5.5 The number of components/NP types or GMM for network type identification selected using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for network Y

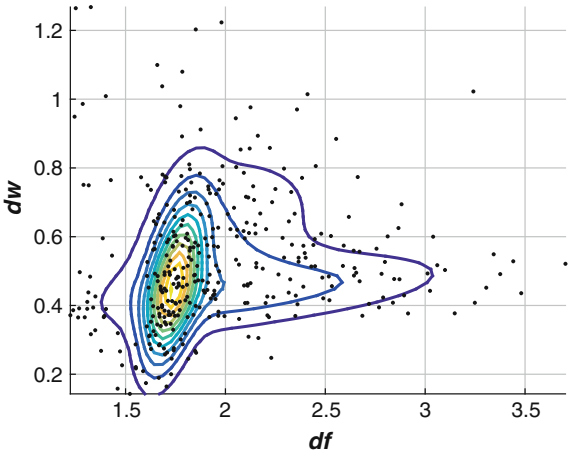


Fig. 5.6 The *df* versus *dw* distribution and GMM with $\eta = 4$ components for network *Y*

$$\begin{cases} \underset{i=1:\eta}{\forall} P(H_i|df) = \dfrac{\frac{1}{|df - H_i(1)|}}{\sum\limits_{i=1:\eta} \frac{1}{|df - H_i(1)|}} \\ \underset{i=1:\eta}{\forall} P(H_i|dw) = \dfrac{\frac{1}{|dw - H_i(2)|}}{\sum\limits_{i=1:\eta} \frac{1}{|dw - H_i(2)|}} \end{cases} \tag{5.3}$$

3. Get joint probability for $H_1 : H_\eta$ at each node. This is obtained by element-wise multiplication of the above likelihood matrices.

$$\underset{i=1:\eta}{\forall} \Theta_i = P(H_i|df) * P(H_i|dw) \tag{5.4}$$

4. Use Maximum A'Posteriori (MAP) estimate at each node to identify which hypothesis each node is most likely to be of.

$$H = \arg\max_{df,dw} \Theta \tag{5.5}$$

The resulting NP type distribution for network *Y* (i.e *H*) is shown in Fig. 5.7 and the next stage of the hop-wise estimation process is clustering the nodes of the same NP type into separate NPs. Following the set of aims of NPs described earlier in the chapter, a heuristic approach is taken to identify NPs where the nodes of the same NP type are initially clustered into ellipses and then optimised using self-organising hierarchical particle swarm optimiser (HPSO) [22] and finally broken into convex Voronoi regions. Initial clustering of nodes (using a GUI) of same NP type into areas of elliptical shapes is done because of the convexity of the shape as well as the ease of identification/demarcation. Then the number, size, shape, position, and orientation of ellipses are optimised using HPSO as it is capable of handling the possible multiple

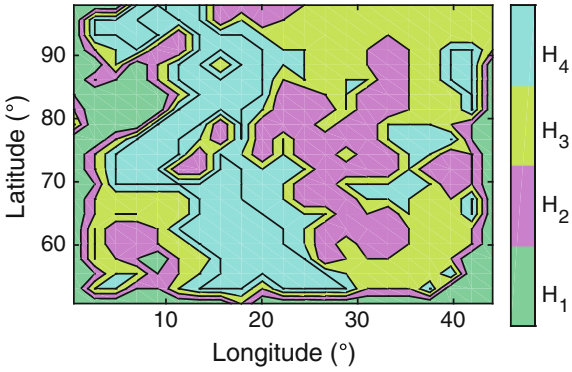


Fig. 5.7 Distribution of NP type (hypotheses) node-wise. Every node belongs to one of NP types $H_1 : H_4$ for network Y

local optima and converging to a global optimum through individual and social behaviour of particles as well as reinitialising velocities of particles stagnated at local optima [22]. These steps are discussed in detail in the following subsections and using a case-study of cyclone tracks over the NA Ocean (Sect. 5.2.3 and Fig. 5.13).

5.2.2.1 Initial Manual Demarcation of NPs

In the first step of the demarcation process, the simplest approach of clustering nodes is by manually drawing ellipses over a visual graph (Fig. 5.7) of the network in order to identify patches of nodes of similar NP types together. This gives a basis for generating the initial population to work on to optimise the NP demarcation. Ellipses are identified by (centre(x,y), major axes length, eccentricity, and orientation). Figure 5.8 shows the manually selected ellipses for network Y.

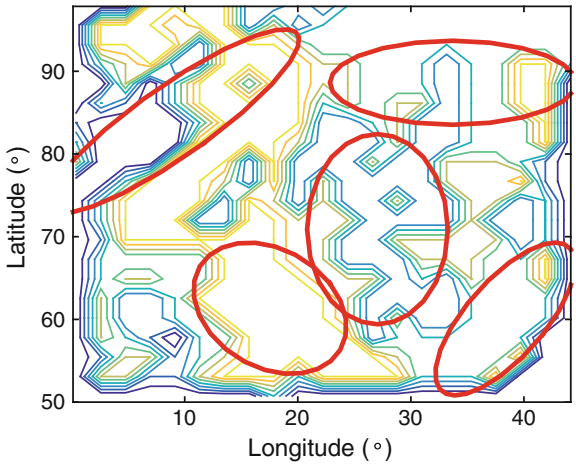


Fig. 5.8 Network Y: Manual NP demarcation using an interactive MATLAB plot. Most prominent NP's are identified with ellipses over a contour plot of NP type distribution

5.2.2.2 Optimising Elliptical NPs Using Self-organising Hierarchical Particle Swarm Optimisation (HPSO)

The initially selected ellipses have to be optimised for NP identification. For this, the particle swarm optimisation (PSO) method is used. It is a population-based, self-adaptive search optimisation technique introduced in 1995 by Kennedy and Eberhart [23]. A population of 'particles' (defined by the dimensions) are allowed to move around the 'search-space' until it converges to an optimal solution. With both cognitive and social components, PSO has been known to converge to a reasonably good solution very quickly [22, 24]. As there are many networks where the network connectivity pattern and the random walk process dynamics are unfolding on the same time scale [25], the quick optimisation process of PSO is useful when networks change in real time. In order to eliminate particles from stagnating at suboptimal solutions, the extended version called self organising hierarchical particle swarm optimisation (HPSO) is used [22].

Particle swarm optimisation (PSO)

PSO [22] is a population based search optimisation technique originating from animal social behaviour such as fish schooling or bird flocking, developed by Kennedy and Eberhart [23] in 1995. Being a population based optimisation algorithm, it initiates with n number of particles in m dimensional search space where m is the number of variables for optimisation. Unlike genetic algorithms which are also population based search algorithms, PSO itself does not include genetic mixing of properties between particles. Each particle achieves optimal position considering its individual behaviour as well as social behaviour. A reward function defined by the application decides the fitness of each particle after each generation/iteration. For each particle a location and a velocity is defined at each iteration. The position of each particle after each generation is given for each dimension d as

$$x_{id} = x_{id} + v_{id} \tag{5.6}$$

where $x_{id} = (x_{1i}, x_{2i}...x_{mi})$ is the current position of the ith particle and $v_{id} = (v_{i1}, v_{i2}, ...v_{mi})$ is the current velocity of the particle's trajectory towards its optimal position. v_{id} is calculated as

$$v_{id} = v_{id} + c_1 \times rand(1) \times (p_{id} - x_{id}) + c_2 \times rand(1) \times (p_{gd} - x_{id}) \tag{5.7}$$

where v_{id} of previous generation provides the momentum for the particle to keep moving towards its optimal position. The second component of the statement is called the *cognitive component* and is the bias towards the best performance for the individual particle whereas the third component or the

social component is the drive for the particle towards global optimal position. $P_{id} = (p_{1i}, p_{2i}, ...p_{mi})$ and $P_g = (p_{1g}, p_{2g}, ...p_{mg})$ gives the best position so far for the ith individual particle (giving the highest reward) and the best position among all particles respectively.

The initial population can be initiated randomly or based on some prior knowledge. The fitness of each particle is evaluated and stored in memory along with the positions of all particles (P_{id}). After each new generation, the fitness values are compared to the best so far and updated accordingly. Generally, in order to keep the particles from drifting outside the search space, a maximum velocity (V_{maxd}) is defined for each dimension. Whenever this is exceeded, the velocity in that dimension is set to V_{maxd}. For the simplicity of implementation and fast convergence rates, PSO has become very popular and it is used in the Chap. 5 in the process of identifying NPs.

Ratnaweera et al. [22] have presented an extension which improves the convergence rate of PSO considerably. This method, known as HPSO-TVAC (Hierarchical Particle Swarm Optimiser with Time-Varying Acceleration Coefficients) is discussed below.

The original form of PSO has the characteristic that the relative magnitudes of cognitive to social component result in different behaviours of the motion of particles. Higher cognitive components result in the particles wandering over the search space while the opposite will lead to particles settling prematurely at local optimums. Kennedy et al. [23] have suggested setting $c_1 = c_2 = 2$. Although this improves the optimisation and reduces convergence times to some extent, the suggestions in [22] provide a more logical approach to better tackle the problem. Taking in to consideration the general desire of encouraging particles to wander about the search space in the initial generations in order to avoid being stagnated at a local optimal point, and contrastingly, the necessity of converging quickly after the global optimal point is found, the goal of this development has been to time-vary the coefficients $c1$ and $c2$ in a way to meet both objectives so that the global optimal point is found efficiently. Under TVAC, Ratnaweera et al. [22] suggest that the coefficients are varied as

$$c_1 = (c_{1f} - c_{1i})\frac{iter}{MAXITR} + c_{1i} \tag{5.8}$$

$$c_2 = (c_{2f} - c_{2i})\frac{iter}{MAXITR} + c_{2i} \tag{5.9}$$

where c_{1i}, c_{1f}, c_{2i}, and c_{2f} are constants, *iter* is the current iteration number and $MAXITR$ is the maximum number of allowable iterations. The authors have found the values for $c_{1min} = 0.5$; $c_{1max} = 2.5$; $c_{2min} = 0.5$; $c_{2max} = 2.5$; to give optimal solutions for many search problems and therefore the same coefficients are used in this chapter. The next improvement suggested in [22] is the "self-organising hierarchical particle swarm optimiser (HPSO)"

where the problem of particles stagnating at local minima for multimodal systems (systems with many local minima), the effect of velocity of the particle reaching zero is avoided by resetting the velocity to a reinitialisation velocity proportional to V_{max} as

$$v_{id} = \begin{cases} r_2 V_{max}, & r_1 < 0.5 \\ (-)r_3 V_{max}, & otherwise \end{cases} \tag{5.10}$$

where r_1, r_2, and r_3 are random values between 0 and 1.

In identifying NP's, HPSO is used to optimise these ellipses to maximise a reward function concerning coverage, overlap and uniformity. The optimisation is done in two phases with the first phase restricting the movement of the ellipse centres and the number of ellipses. And the second phase is developed to optimise on the uniform coverage.

NP Optimisation: Phase I

The variables for this phase are the length of major axis, eccentricity and orientation of each of the ellipses. The total number of dimensions (n_d) for the optimisation problem is the number of variables for each ellipse ($n_d = 3$ for Phase I) times the number of ellipses (M) while the centres and the ellipse numbers are kept fixed. The optimisation problem can be stated as;

$$\begin{aligned} \min_{d_1, d_2, \ldots d_{n_d}} \quad & f_0(d_1, d_2, \ldots d_{n_d}) \\ \text{subject to} \quad & d_i > 0, \ i = 1, 4, \ldots, n_d - 2 \qquad \text{For length of ellipse} \\ & 0 < d_i < 1, \ i = 2, 5, \ldots, n_d - 1 \qquad \text{For eccentricity} \end{aligned} \tag{5.11}$$

where f_0 is the cost function. The adaptation of HPSO algorithm to solve the optimisation problem stated above is detailed in Algorithm 5. A population of N particles are initiated in n_d dimensions in terms of a position vector X_i where $i \in (1, 2, \ldots, N)$ and a velocity vector V. This initial population of particles with position vectors X_i are obtained from adding a Gaussian noise with a small standard deviation to the original particle obtained from manually drawn ellipses. In particle motion, the maximum velocity in each dimension is defined as $V_{max_d} = max(X_d)$ as usually done in PSO to reduce roaming outside the search-space [22]. The initial velocities are defined as fractions of V_{max}. The best localisation of each particle P_{id} (which shows the highest fitness), and localisation of fittest particle so far (P_g) is updated with each iteration. The new positions of each particle are updated with new velocities.

Reward Function for Optimisation Phase I (f_{0I})

The objective of Phase I is to identify the main clusters which would form the largest NPs. This is an important step as the final *MFPT* estimation would greatly depend

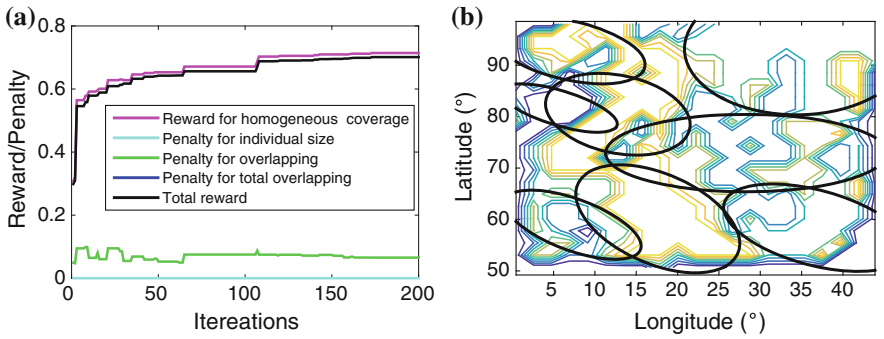


Fig. 5.9 Network Y: optimisation Phase I. **a** HPSO reward function. **b** Identified elliptical NPs at the end of Phase I

on the accuracy of calculation within these largest NPs. With this objective in mind, the reward function for this phase (f_{0I}) is formulated as a weighted sum for the following.

- Reward for uniform coverage (i.e the proportion of nodes covered by a single ellipse); R_{UC}
- Penalty for the number of overlapped nodes; P_O
- Higher penalty for ellipses that are totally overlapped by others; P_{TO}
- Penalty for individual size of ellipses being below a threshold; P_{IS}

The fitness for each of the N particles is calculated for Phase I as

$$f_{0I}(x) = w_1 R_{UC} - w_2 P_{IS} - w_3 P_O - w_4 P_{TO}$$

where $f_{0I}(x_i)$ is the fitness of the ith particle (x_i), $w_1 : w_4$ are the weights for each reward or penalty. The weights $w_1 : w_4$ for the rewards and penalties used were selected by observing the convergence rate and fitness for a range of possible fitness functions. The reward function is used for network Y with weights as $[w_1, w_2, w_3, w_4] = [1, 0.4, 0.2, 0.4]$. Figure 5.9a shows the individual rewards and the total reward function with respect to the iteration numbers and Fig. 5.9b shows the NP cluster ellipses at the end of Phase I.

NP Optimisation: Phase II- HPSO for Optimising Coverage and Identifying Smaller Primitives

The main (largest) NP's are identified roughly by the end of Phase I. The next stage is concerned with clustering the isolated nodes outside of the first set of ellipses either by merging into larger NPs or by forming smaller NP's by themselves. A function called 'grains' that is used in MATLAB for image processing for clustering similar areas to identify object boarders is used to identify the centroids of uncovered areas. The next generation of ellipses are initialised partially with these centroids and the

Algorithm 5 (HPSO in Phase I)

begin
Initialise population of $\mathbf{N}$ *particles*
$X_1 = [X_{maj1}, X_{ecc1}, X_{ori1}, ...X_{majM}, X_{eccM}, X_{oriM}]$ {Position vector for 1st particle}
$\underset{2<j<N}{\forall} X_j = X_1 + [\eta_{1j}, \eta_{2j}, \ldots, \eta_{3Mj}], \eta \sim \mathcal{N}(0, \sigma^2)$ {Add noise to get other particles}
$V_{max} = max(X_j)$ {Initialize velocity matrix}
$\underset{j}{\forall} V_j = 0.1\ V_{max}$
Define
fitness threshold = f_{th}
$f_{0I} = 0, F = [0\ 0\ \ldots\ 0]_{1\times N}$
maximum iterations = I_{max}
$[c1_{min}, c1_{max}, c2_{min}, c2_{max}] = [0.5, 2.5, 0.5, 2.5]$ {Acceleration coefficients}
$[w_1, w_2, w_3, w_4]$ {Reward function weights}
$I = 0$;
while f< f_{th} & $I < I_{max}$ **do**
  $c1 = (c1_{min} - c1_{max})I/I_{max} + c1_{max}$;
  $c2 = (c2_{max} - c2_{min})I/I_{max} + c2_{min}$;
  $i = 1 : N$
  $f_{0I}(x) = w_1 R_{UC} - w_2 P_{IS} - w_3 P_O - w_4 P_{TO}$ {Calculate fitness}
  Update P_{id} ***and*** P_g
  if $F_i < f_{0I}(x)$ **then**
    $F_i = f_{0I}(x)$ {Update best fitness matrix}
    $P_{id} = X_i$ {Best position of each particle}
    if indexOF(max(F))==i **then**
      $P_g = P_{0I}$ {Fittest particle so far}
    end if
  end if
  for d = 1: 3M **do**
    Update velocity
    $V_{id} = c_1 \times \text{rand}(1) \times (P_{id} - X_{id}) - c_2 \times \text{rand}(1) \times (P_{gd} - X_{id})$ {New velocity}
    if $V_{id} == 0$ **then**
      if rand(1) < 0.5 **then**
        V_{id} = rand(1) x $V_{max}(d)$
      else
        V_{id} = (-)rand(1) x $V_{max}(d)$
      end if
    end if
    V_{id}= sign(V_{id}) × min(abs(V_{id}),V_{max})
    $X_{id} = X_{id} + V_{id}$
    if limits violated **then**
      reset X_{id}
    end if
  end for
  $I = I + 1$;
end while

rest with the output of the earlier phase. A modified reward function is used in this stage to give more prominence to total coverage and added penalties for uncovered area. The reward function for this phase is

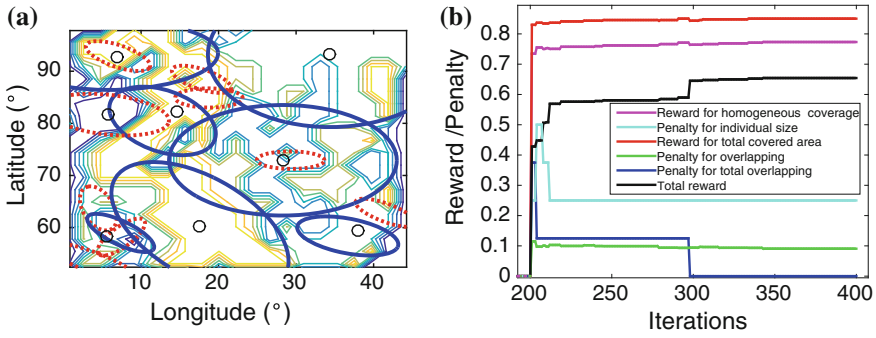


Fig. 5.10 Network Y: optimisation Phase II. **a** NP clusters at the end of optimisation Phase II. The newly added patches are marked in *red dashed lines*. **b** Reward function ($f_{0\text{II}}$) with respect to iterations in optimisation Phase II. The sharp changes in the curves correspond to removal of smaller ellipses in subsequent iterations

$$f_{0\text{II}}(x) = w_1 R_{UC} + w_2 R_{IS} - w_3 P_O - w_4 P_{IS} - w_5 P_{TO}$$

where R_{IS} is the reward for individual NP sizes. The weights of the individual rewards were $[w_1, w_2, w_3, w_4, w_5] = [0.33, 0.67, 0.5, 0.5, 0.5]$. Sparsely interlaced iterations were used to identify and delete ellipses smaller than a fixed lower limit. Iteration outputs for optimisation Phase II for network Y are shown in Fig. 5.10a, b gives the individual and total reward functions ($f_{0\text{II}}$) with respect to iterations.

NP Final Demarcation Using Voronoi Regions

In real-time computation, the hop-wise MFPT calculation complexity increases largely with the number of NP's in a network. Therefore, the ellipses resulting from the previous optimisation steps can be divided as ellipses of considerable size and ones which, if considered as individual NPs are very small and will increase the *MFPT* estimation complexity. Therefore, the set of ellipses are divided into two groups depending on their coverage areas. The threshold is taken as having a major axis at least three times the rate of spread. This is done in the sense that if the walker is able to reach the furthermost end of the primitive in less than three steps, such NPs are considered inefficient by increasing calculation complexity. Nodes within ellipses such as these are merged into the NPs closest to them. The centres of gravity for larger ellipses are adjusted by a weighted average of coverage areas with the positions of added clusters.

As the final step of the NP identification process, the total continuous area is divided into convex network primitives using the centres of gravity for the remaining ellipses as centres of Voronoi regions. Finally, the infinite edges of Voronoi regions are clipped off at the network boarder in order to make hop-wise *MFPT* estimation feasible. The final NP demarcation for network Y is depicted in Fig. 5.11 with a total of eight NPs identified.

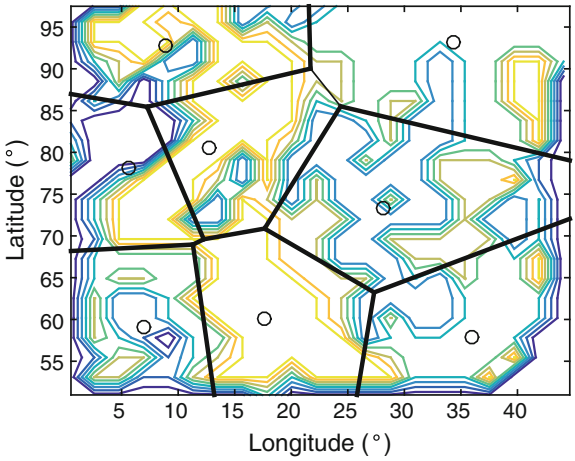


Fig. 5.11 Network Y: final demarcation of NPs identified through Voronoi regions with the infinite edges clipped off with the network boarder

This method of NP identification is presented here for a 2D unconnected node distribution but the concepts are adaptable for higher dimensions using higher dimension ellipsoids and Voronoi regions and even for complex connected networks.

NP Transport Properties

As described at the beginning of this chapter, an NP is a part of a network uniform in transport properties. Therefore, for each identified NP, transport properties have to be identified. Calculating the *MFPT* for a random walker to reach any target T within the same NP as the source point (i.e one hop) is straight forward using methods from Chap. 4 given that the values for df and dw and constants A and B in Eq. (4.3) are known. If another method is used for computing the *MFPT*, this calculation is still straight forward if the transport properties for that method were used in identifying NPs. Ideally every node within an NP should have the same values for df, dw and bias (and thereby A and B). However, since the NP identification process needs to allow some robustness when dealing with real world networks, there could be slight variations in these values between nodes in the same NP. Yet, it is important for the hop-wise *MFPT* calculation that common values for these dimensions are found for each NP such that Eq. (4.3) can be applied. One way of obtaining these values are using the df and dw values set by the hypothesis of the most common NP type of within the NP. But to add robustness through customisation, the transport variables df, dw, α, $A(\alpha)$, & $B(\alpha)$ for each of the primitives were recalculated dividing the dataset between the NPs and again into Ψ number of angle segments i.e. $\alpha \in \Phi \mid \Phi\{(-\pi, -\pi + 2\pi/\Psi), (-\pi + 2\pi/\Psi, -\pi + 4\pi/\Psi), \ldots, (-\pi + 2(\Psi - 1)\pi/\Psi, -\pi + 2\Psi\pi/\Psi)\}$. It was decided to use $\Psi = 24$ angle segments for the simulations after compromising between (limited) training dataset size and accuracy.

In the next section the NP identification process is used and extended to the hop-wise *MFPT* prediction presented with the case study of cyclone motion in the NA Ocean.

5.2.3 Case Study: MFPT for Cyclone Motion

The methods discussed that estimate *MFPT* for a random walker at a known source point S to reach a given target point T in inhomogeneous networks are now discussed using a case study of predicting *MFPT* for cyclone motion using a dataset of the paths followed by cyclones over the NA Ocean.

5.2.3.1 Case Study: Dataset

The dataset used for this case study is the same as what was used in Chap. 4 which gives the complete set of cyclone track archive for the cyclones observed over the NA Ocean since 1950 [26]. It can be observed that with respect to cyclone track history over other Oceans, the dataset over NA Ocean shows a clear correlation with the Beta effect and thus becomes a good candidate for an example biased random walks. This gives a good case study where random walks are biased by environmental factors. Although the complexity of the atmosphere and stochastic nature of the cyclone motion itself creates an uncorrelated set of random walks, the biasing factor imposes broad limitations to the movement.

This data source provides 6 hourly best track coordinate locations of the eye of the cyclone for every documented cyclone in the categories of Tropical Depression, Tropical Storm, Hurricane, Extra tropical cyclone, Subtropical cyclone, Low, Tropical wave and Disturbance. The selected complete dataset is shown in Fig. 5.12 that discusses how the dataset consists of a network inhomogeneous in transport properties.

5.2.3.2 Case Study: Calculating Transport Properties

The *df* for the case study obtained as being equal to 2 which results from the continuous two-dimensional Euclidean state space. Using the dataset of cyclones between 1950–2012 as the training dataset, dw_b was calculated using Eqs. (4.2) and (4.7) following Algorithm 4. When calculating dw at different points in the state space, it is observed that the values vary considerably throughout the considered area (i.e. state space). In order to get the distribution of dw_b, the area is divided into a grid-world and used the datasets of cyclones originating from each grid cell to calculate dw with respect to direction.

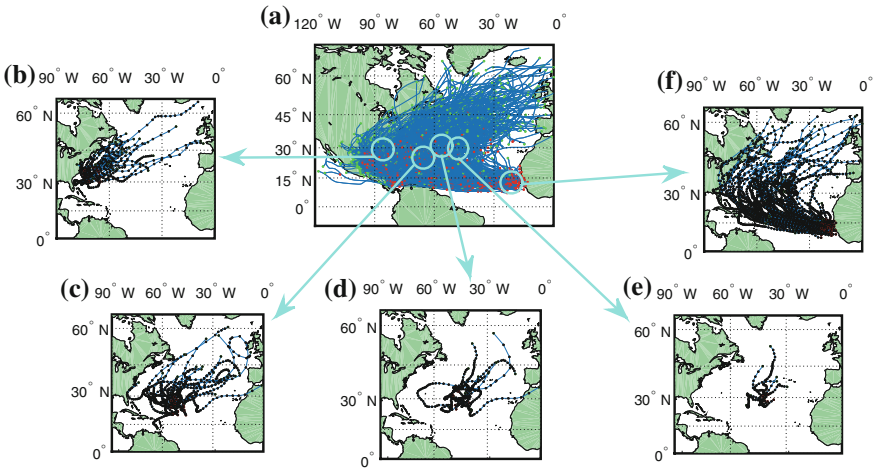


Fig. 5.12 Case study: past cyclone tracks showing inhomogeneous transport characteristics. **a** The complete dataset of archived cyclone track data over the NA Ocean since 1950. **b–f** Subsets of tracks filtered by the initial track location (initially observed within the *circles* shown at the base of each *arrow* in **a**). The *dots* show 6 hourly positions of the eye of the cyclones. It is seen how the bias direction (directional inhomogeneity) and bias intensity(inhomogeneity of density of nodes) is changing with location

5.2.3.3 Cases Study: NP Identification

The NP type/hypothesis at each location are obtained using the dw_b and df profiles similar to network Y. Due to the network being continuous, it was required to use the above mentioned grid-world for NP type identification. Also due to the state space being continuous, the df remained constant for every point in the state space and therefore it was only required to identify hypotheses for dw_b, as it is the only variable transport property of importance for NP identification. Similar to network Y, the dw_b distribution used for NP type selection is composed of dw_b values in the direction of maximum bias.

The hypotheses/NP type identification is done manually by observation since the only variable is dw_b. Four hypotheses were identified from this distribution which fixed the NP type set at $\eta = 4$. The point-wise NP type distribution is shown in Fig. 5.13b. NP identification process for this case study follows the procedure explained with the example of network Y and the outputs and intermediate steps are illustrated in Fig. 5.13.

A total number of 15 convex NPs are obtained using the 4 identified hypotheses. The values for df and dw_b for each of the 15 NPs were identified by using the dataset divided by the NP in which each cyclone was originally recorded in. (i.e. from the initial recorded position). A unique value set of $\{df, dw_b(\alpha), A(\alpha), B(\alpha), \alpha\}$ where

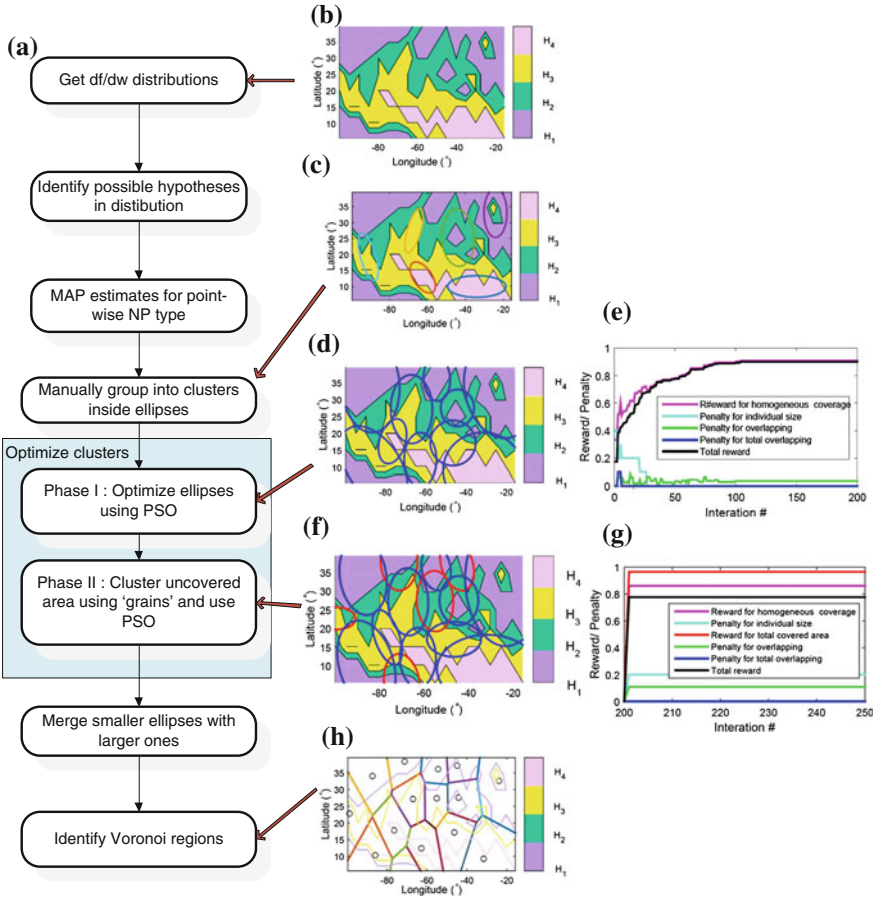


Fig. 5.13 Identifying NPs for the case study of cyclone tracks over the NA Ocean. **a** The flowchart of NP identification process. **b** Point-wise NP type/hypothesis distribution. **c** Manual demarcation of elliptical NP patches. **d**, **e** Resulting NP distribution at the end of optimisation Phase I, and its reward function. **f**, **g** The same for Phase II. **h** Identified NP set

α is the direction of the hop is stored for each NP for the real-time *MFPT* prediction for each hop using Eq. (4.3). The angles are divided into $\Psi = 24$ segments limited by training dataset. These values are stored in a cell structure to be accessed in real-time computation.

5.2.3.4 Hop-Wise *MFPT* Estimation

A hop-wise *MFPT* prediction method follows the NP identification process. Unlike the point to point estimation as in the case of a single NP where there is only one pair

of S and T considered by the walker in a single attempt, when the network is divided into patches with different characteristics, the paths a walker would take in the next 'hop' differ largely depending on the exit point of the previous NP. Therefore, it is required that multiple exit points are considered for each starting points within the current NP. This as a result will increase the calculations exponentially with each 'hop'. The scalability of the presented methods will be discussed later using calculation complexities.

The hop-wise estimation starts with selecting the exit points for each NP the walker enters (given that the final target point T, is not inside the NP the walker enters). There is one assumption in this hop-wise *MFPT* calculation that is that there are no loops in the hops. If that was the case, the calculation might crash in an infinite loop until the maximum hops are reached. Although it is possible to extend the calculation to include this case, due to the limitation of computational resources, this assumption is enforced as a limitation to the number of hops.

Focussing back on the question of selecting exit points for a random walker to exit from its current NP, for directionally biased random walks, the main influence is from the network bias. When the bias direction is known, this is the most common exit direction from the NP. If the bias intensity is also known, the size of the angle segment of possible exit points can also be calculated. Without a loss of generality, an angle segment of Δ is chosen around the bias direction with 95 % confidence interval that random walkers in that NP would exit through. This is obtained from the training dataset used in getting the transport variables. In getting the actual exit points, it is straight forward for a discrete network by selecting the set of nodes closest to the boundary of the NP within the angle segment of possible exit. When the network is continuous, it is required to find a discrete number of exit points from the current NP. This number is a compromise between the required precision and the affordable computational complexity. In order maximise the utilisation of knowledge encapsulated in transport variables calculated from previous data, an exit point is used for each angle segment transport variables are pre-calculated for (Φ) and which are within the probabilistic region. For isotropic random walks, this region extends the complete NP boundary.

5.2.3.5 Probabilistic Path Selection

Regardless of the selection of NP exit points, it is inherent in its biased motion that it is more likely that some paths are taken than others. The probability of exiting through a node directly in the direction of bias is more likely than in other directions. Therefore, probability of path selection obtained from past data, is integrated into the hop-wise *MFPT* prediction. The probability of any given random path being the actual random walk is the joint probability along the hops of that path. Thus, the final *MFPT* estimation is an average *MFPT* weighted by the probability of that particular path. Figure 5.14 shows a few examples of the selected hop 'branches' for a selected

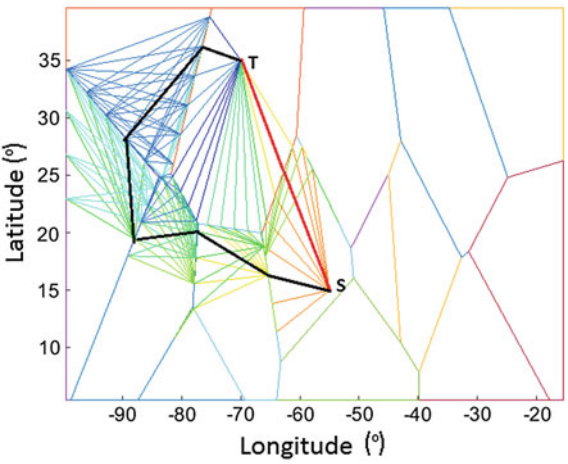


Fig. 5.14 Hop-wise *MFPT* estimation for the case study of cyclone tracks over NA Ocean for one pair of S and T. The 'branches' of hop-wise routes considered in the *MFPT* estimation process are shown. Different hops are shown in *different colours* and two sample routes (collection of hops) are highlighted in *red* and *black*

source and target set. It is worth noting that the plots are not random walks but rather the connectivity of a few hops of random walks.

5.2.3.6 Real-Time Prediction with Parallel Processing

The computational complexity resulting from the need to consider many alternate paths of a single random walk depending on transit points between NPs is a drawback in using the hop-wise *MFPT* prediction in real-time. This increases the number of computations required exponentially with each NP interface a random walker needs to cross in order to reach the target point. This makes computational resources a bottleneck for the estimation. But an important factor in this calculation is that although multiple paths of a single random walker has to be considered, these paths are independent of each other because in reality although multiple paths are considered a walker can only take one at a given time. To make use of this independence, the MATLAB code developed to implement this methods in simulations uses parallel processing functions to improve calculation efficiency. To preserve the efficiency provided by parallel processing, variables used in each parallel process needs to be independent of each other. Therefore parallel processing is used only per hop to calculate *MFPT* via multiple exit/transit nodes which are independent.

The hop-wise *MFPT* prediction method has a real-time computational complexity of $\mathcal{O}(n_d^h)$ where n_d is the total number of dimensions of the state space and h is the maximum number of hops considered. Therefore with scaling in terms of higher dimension random walks, the real time computational complexity will not be effected. However, in obtaining *df* and *dw* distributions, the calculation complexity will exponentially increase with dimensions of the state space. This could be significantly reduced by parallel processing as calculations are still mutually independent.

5.2.3.7 *MFPT* to Reach Any Target T

The *MFPT* estimates obtained for each hop of the calculation are stored in a (MATLAB) cell structure until the final estimation converges or if the hop number has reached a threshold. When the networks are biased and when the bias does not form loops, after a limited number of steps, if the target is not reached, it can be assumed that the *MFPT* tends to infinity. This number of steps is reduced with the increase in bias intensity. With the number of iterations set to the converging threshold of the network, Algorithm 6 presents the pseudo code followed in the final *MFPT* estimation process.

5.2.3.8 Case Study: Hop-Wise *MFPT* Estimation

For this case study, the number of maximum hops is limited to $h_{max} = 20$ by observation since the results generally converged by 10 hops. Figure 5.15 shows the *MFPT* prediction for ten (S,T) pairs with respect to the number of hops considered (here the branches which have not yet reached the target are considered as misses). The (S, T) pairs are the same as the ten cases presented in Table 5.1. It can be concluded that after 20 hops, any walkers which have not yet reached the target have indeed missed it as the probability of returning against the bias path is extremely low (given that there are no loops in the bias vector field in this example). If there were loops in the bias vector field, convergence would be delayed.

5.2.4 Case Study: Results Comparison

The predictions made using the hop-wise model are compared with real data averages and predictions using homogeneous transport properties for several (S,T) pairs. Ten such comparisons are presented in Table 5.1 and from this it is shown that breaking the problem into NPs reduces the average error from 18.8 to 5.4 % with respect to the homogeneous network assumption, but at the cost of 12 times increase in processing time required. However, this increase in processing time is subjective and the importance changes with the application. For example, in the cyclone track dataset, new data points were available every 6 h and the increase in processing time in the order of one or two minuets does not make any difference. It is also noted that the processing time is dependent on the resolutions used. Figure 5.15 presents the *MFPT* convergence rate with the number of 'hops'. All simulations and calculations presented in this chapter were done using MATLAB coding on an Intel® core™ i7-4770 CPU with 16 GB available RAM. In conclusion, while requiring more computational power, breaking the problem of *MFPT* estimation on an inhomogeneous network by dividing the network using the concept of NPs increases the estimation accuracy considerably. Comparisons were done with only

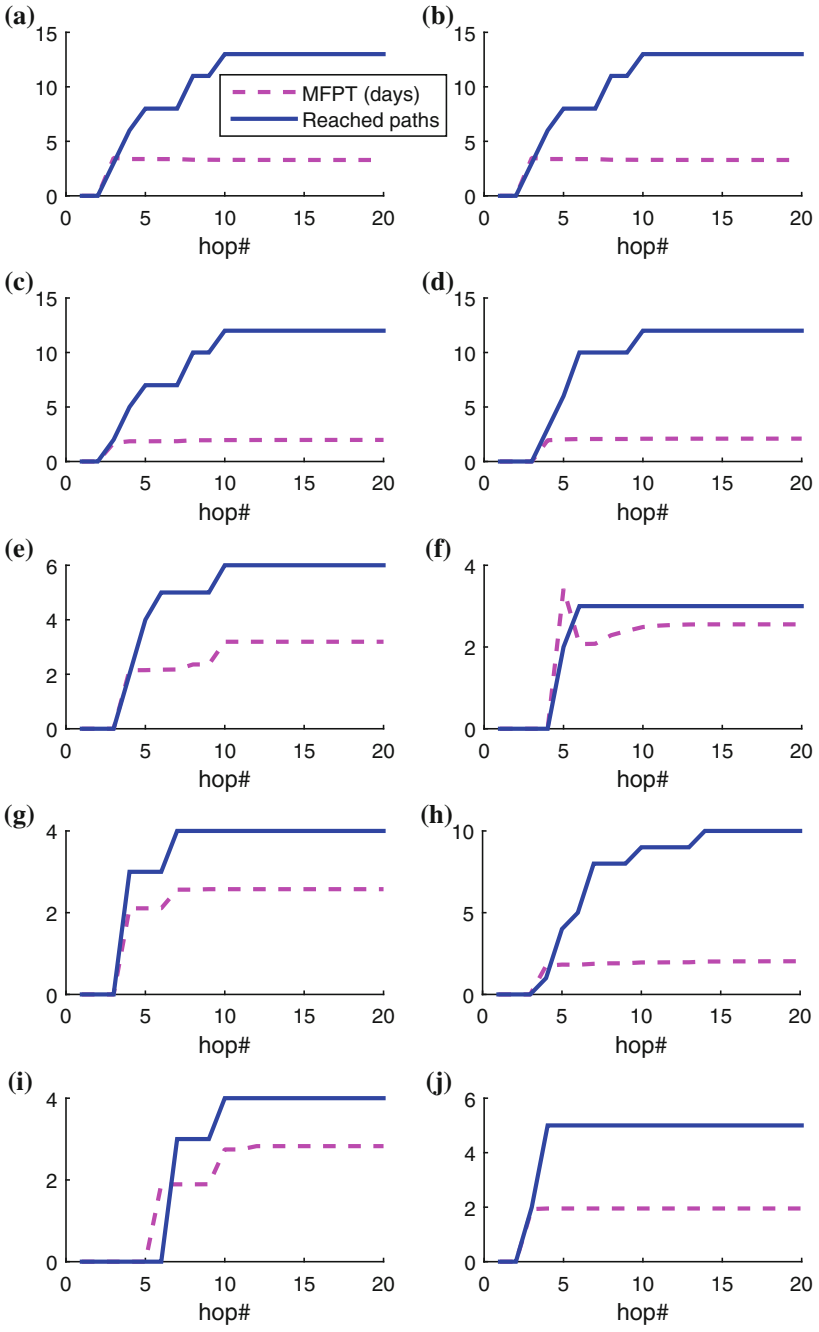


Fig. 5.15 *MFPT* estimation for the 10 presented comparisons in Table 5.1. The convergence of the iterative *MFPT* estimation is observed

Algorithm 6 Hop-wise *MFPT* calculation

Begin
Define S, T, $frontline = S$ {Input Source and Target}
Load NPs, df, dw, A, B, P_{DD} {Import NP & Transport properties}
Identify NP containing $S\&T \rightarrow NP_S, NP_T$
set h_{max} {maximum number of hops}
if $S == T$ **then**
  $hops\{1\} = \{source = S, Pr = 1, mfpt = 0, prev_{NP} = 0, reached = 1\}$
else
  $hops\{1\} = \{source = S, Pr = 1, mfpt = 0, prev_{NP} = 0, reached = 0\}$
  for $hop = 2 : h_{max}$ **do**
    for n=1:length(frontline) **do**
      For every boundary node from previous hop
      if $hops\{hop\}.reached > 0$ **then**
        Append $hops[hop]|hops[hop-1](n)$ {Append nth elements}
      else
        $[newfrontline, mfpt_{all}, NP_S, reached_{all}]$=$getNexthop$
        $Pr_1 = P_{DD}\{NP_S\}.hops\{hop-1\}.Pr(n)$
        Append $hops\{hop\}|\{source = frontline, Pr = Pr_1, mfpt = mfpt_{all}, reached = reached_{all}, prev_{NP} = NP_S\}$
      end if
    end for
    $frontline = newfrontline$;
  end for
end if

getNexthop
if $NP_S == NP_T$ **then**
  calculate final *MFPT* using Eq. (4.3)
  return $T, mfpt, NP_S, reached = 1$
else
  Identify exit points from $NP_S \rightarrow$ frontline
  Start ‖ processing
  for every point on the previous frontline **do**
    if target reached **then**
      calculate final *MFPT* using P_r
      return $T, mfpt, NP_S, reached = 1$
    else
      calculate new frontline
      calculate *MFPT* to reach every ith point on frontline
      return $\{frontline, mfpt, NP_S, reached = 0\}$
    end if
  end for
  End ‖ processing
end if

larger distances between S and T where the hop-wise calculation made sense (for S and T to be in different NPs). Currently 120 h is the maximum lead time official forecasts can provide.

Table 5.1 Comparisons of *MFPT* predictions with actual average FPTs for the dataset of past cyclone tracks over the NA Ocean discussed in Sect. 5.2.3.1

Test case	Source [Lat,Lon]	Target [Lat,Lon]	Real *MFPT* (days)	Hop-wise prediction		Prediction assuming network homogeneity	
				MFPT	Error (%)	*MFPT*	Error (%)
1	[12–20]	[32–70]	11.2917	11.3066	0.13	10.6305	5.86
2	[15–55]	[35–70]	7.25	7.8085	7.70	10.2541	41.44
3	[15–60]	[32–70]	6.0417	6.0417	0.00	7.712	27.65
4	[15–58]	[23–64]	7.4722	7.9218	6.02	5.3626	28.23
5	[15–50]	[25–60]	8.3611	8.3791	0.22	6.8037	18.63
6	[18–45]	[27–67]	8.0278	8.3881	4.49	9.4525	17.75
7	[12–45]	[35–65]	11.8333	12.7713	7.93	11.0873	6.30
8	[13–21]	[29–69]	10.8333	10.2158	5.70	10.4282	3.74
9	[14–48]	[34–69]	9.7222	10.3003	5.95	10.6472	9.51
10	[20–45]	[25–65]	9.6111	8.119	15.52	6.829	28.95

A set of 10 source and target sets are presented to compare the average FPTs with the hop-wise prediction presented in this chapter, and with the predictions obtained using methods predicted in Chap. 4 assuming homogeneous transport properties. It is shown that breaking the problem into NPs using the presented methods reduces the average error from 18.8 to 5.4 % with respect to the homogeneous network assumption

5.3 Conclusions

This chapter discussed the effect of inhomogeneity of transport variables are shown on state spaces onto which processes leading to natural disasters translate into. The concept of dividing an inhomogeneous network into NPs is introduced. This enables straight forward *MFPT* calculation between any two nodes within an NP and makes way for hop-wise *MFPT* estimation to reach any target point in the inhomogeneous media when the NPs are convex in shape. An algorithm to identify convex NPs for any given 2D node distribution is discussed which can be extended to higher dimensions with the use of higher dimension ellipsoids and Voronoi regions. A 'Hop-wise' *MFPT* prediction method is shown to get a final approximation of *MFPT* to reach any target point in the complex media. The effect of probability of reaching a target given the effect of biasing factors is demonstrated in this hop-wise *MFPT* estimation.

Several important conclusions can be observed from material discussed in this chapter. Firstly it was shown that a large class of processes leading to natural disaster dynamics translate into random walks that encounter inhomogeneity in networks' transport properties. Such inhomogeneity can not always be modelled through node degree distributions or weight distributions especially when connectivity within networks are temporal. Dividing the network into homogeneous network primitives is shown to be a feasible approach to address this problem. It was shown that PSO techniques can be used to optimise the NP identification process that is described through a generic algorithm that identifies NPs using the transport property distributions. The discussed methods were demonstrated using simulated and real datasets where predictions showed better results with the dataset of archived cyclone tracks when compared to predictions assuming homogeneous networks.

5.4 Summary

Random walks resulting from state changes of processes leading to natural disaster activity commonly show network inhomogeneity. Therefore, this chapter investigated the possibility of estimating *MFPT* for a random walker to reach a target T starting from a source node S in inhomogeneous media using the concept of NPs to divide the problem. NPs are identified by their transport properties df and dw, and clustered to be convex in shape to facilitate hop-wise *MFPT* prediction. The methods are described initially with random walks on a hypothetical two dimensional generic discrete node distribution denoted as network Y. Then a case study using the dataset of past cyclone tracks over the NA Ocean for the past 65 years was used to verify the methods by comparing results with numerical *MFPT* values and estimation results assuming homogeneous transport properties using methods from Chap. 4. Comparison results showed a 3.5 times less error with the presented methods.

One issue with the method is that when there is zero prior knowledge of cyclone tracks in an NP in some direction, this method cannot be used to predict for that direction (as no transport variables can be calculated). Finally, it is worth noting that this method of using multiple possible exit points from an NP can be directly used to extend this method in calculating *MFPT* for random spreads such as fire propagation.

References

1. Agliari E, Burioni R (2009) Random walks on deterministic scale-free networks: exact results. Phys Rev E 80(3):031125
2. Godec A, Metzler R (2015) Optimization and universality of Brownian search in quenched heterogeneous media. arXiv preprint arXiv:150300558
3. Noh JD, Rieger H (2004) Random walks on complex networks. Phys Rev Lett 92(11):118701
4. Sood V, Redner S, Ben-Avraham D (2005) First-passage properties of the Erdös-Renyi random graph. J Phys Math Gen 38(1):109
5. Roy C, Kovordányi R (2012) Tropical cyclone track forecasting techniques-a review. Atmos Res 104:40–69
6. Zerger A, Wealands S (2004a) Beyond modelling: linking models with GIS for flood risk management. Nat Hazards 33(2):191–208
7. Anderson HE (1983) Predicting wind-driven wild land fire size and shape. US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station
8. Achtemeier GL, Goodrick SA, Liu Y (2012) Modeling multiple-core updraft plume rise for an aerial ignition prescribed burn by coupling daysmoke with a cellular automata fire model. Atmosphere 3(3):352–376
9. Aparicio JP, Pascual M (2007) Building epidemiological models from R0: an implicit treatment of transmission in networks. Proc Biol Sci 274(1609):505–512
10. Colizza V, Vespignani A (2007) Invasion threshold in heterogeneous metapopulation networks. Phys Rev Lett 99(14):148701
11. Kurella V, Tzou JC, Coombs D, Ward MJ (2015) Asymptotic analysis of first passage time problems inspired by ecology. Bull Math Biol 77(1):83–125
12. Giuggioli L, Pérez-Becker S, Sanders DP (2013) Encounter times in overlapping domains: application to epidemic spread in a population of territorial animals. Phys Rev Lett 110(5):058103
13. Condamin S, Benichou O, Tejedor V, Voituriez R, Klafter J (2007b) First-passage times in complex scale-invariant media. Nature 450(7166):77–80
14. Tejedor V, Benichou O, Voituriez R (2011) Close or connected: distance and connectivity effects on transport in networks. Phys Rev E 83(6):066102
15. Wijesundera I, Nirmalathas T, Halgamuge MN, Nanayakkara T, *mfpt* calculation for random walks in inhomogeneous networks. Under Review
16. Ferguson NM, Keeling MJ, Edmunds WJ, Gant R, Grenfell BT, Amderson RM, Leach S (2003) Planning for smallpox outbreaks. Nature 425(6959):681–685. doi:10.1038/nature02007
17. Ben-Avraham D, Havlin S (2000) Diffusion and reactions in fractals and disordered systems. Cambridge University Press, Cambridge
18. Finney M, Station RMR (1998) FARSITE. Fire Area Simulator-model development and evaluation, US Department of Agriculture, Forest Service, Rocky Mountain Research Station
19. Reynolds D (2009) Gaussian mixture models, Springer US, book section 196, pp 659–663. doi:10.1007/978-0-387-73003-5_196

20. González-Fierro M, Hernández-García D, Nanayakkara T, Balaguer C (2015) Behavior sequencing based on demonstrations: a case of a humanoid opening a door while walking. Adv Robot 29(5):315–329
21. Celeux G, Soromenho G (1996) An entropy criterion for assessing the number of clusters in a mixture model. J Classif 13(2):195–212
22. Ratnaweera A, Halgamuge SK, Watson HC (2004) Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. IEEE Trans Evol Comput 8(3):240–255
23. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks, vol 4, pp 1942–1948. doi:10.1109/ICNN.1995.488968
24. Eberhart RC, Shi Y (2001) Particle swarm optimization: developments, applications and resources. In: Proceedings of the congress on evolutionary computation, vol 1. IEEE, pp 81–86
25. Perra N, Baronchelli A, Mocanu D, Gonçalves B, Pastor-Satorras R, Vespignani A (2012) Random walks and search in time-varying networks. Phys Rev Lett 109(23):238701
26. Chris Landsea JF, Beven J (2014) Atlantic hurricane database (HURDAT2) 1851-2014

Chapter 6
Conclusions and Future Research Directions

This book discusses natural disaster prediction methods by analysing the state changes of the dynamics behind natural disasters. It shows that many such state changes behave as random walks in their respective state spaces. Therefore, this book mainly focusses on methods that predict the *MFPT* for random walks resulting from natural processes leading to natural disasters viewed in an engineering perspective. The main objective of this book is to address network inhomogeneity and directional bias that affect random walks' *MFPT* calculations. The motivation behind this discussion is the large class of process behind natural disaster dynamics of which the state transitions translate to random walks where the non-linear interactions with the environment result in biased random walks showing inhomogeneous transport characteristics.

The first part of this book presents a discussion of an end-to-end prediction model for the specific application of cyclone induced flood prediction. This particular application was chosen due to its construction based on many interdependent random processes affected by network inhomogeneity and biasing factors. Chapter 3 presents a modularised model allowing flexible data assimilation and easy integration of predictions from other models. A novel concept of dividing the terrain into *Geographic Primitives (GP)* is used to accommodate network inhomogeneity where a transition probability matrix (TPM) per GP is used to summarise the effects of bias. Efficient real-time predictions are obtained through the use of the TPM in a Markov chain and properties of the network identified through Eigen-value analysis. A Bayesian framework combines the advantages of data-oriented and heuristic modelling and the proposed model is validated using two complete datasets of past cyclones from 2010 to 2011 season.

The second part of the book concentrates on addressing the directional bias affecting a large class of random walks that result from real-world dynamic systems. In Chap. 4 *MFPT* calculation using bias modified transport variables is suggested as a feasible generic approach to this issue. An algorithm is developed to capture the effects of bias on the 'walk dimension' and a case invariant relationship is obtained

I. Wijesundera et al., *Natural Disasters, When Will They Reach Me?*,
Springer Natural Hazards, DOI 10.1007/978-981-10-1113-9_6

empirically to calculate the novel metric called the *bias modified walk dimension* that is shown to improve the *MFPT* prediction accuracy in the presence of directional bias. A question might arise on the dependence of the proposed model on the set of hypothetical networks used to derive this relationship. However, the diversity of the case studies that are used throughout the book to validate the methods indicate wider applicability. A case study with the complete dataset of cyclone tracks over the North Atlantic Ocean is presented validating the methods proposed in Chap. 4.

The final part of the book concerned the development of methods to address inhomogeneity in network transport properties experienced by a random walker. Chapter 5 presents a novel concept of dividing the network into homogeneous patches known as *Network Primitives (NP)* and adapting a hop-wise prediction method. An algorithm is heuristically developed to identify convex NP's for a given network optimised using particle swarm optimisation methods. Simulations using the same cyclone track dataset (that was also used in Chap. 4) show the improvement of prediction accuracy from this method when compared to predictions assuming homogeneous transport properties.

Overall, this book has presented that state changes for a large class of natural disaster dynamics can be translated into biased random walks in inhomogeneous media and a prediction of the *MFPT* to reach some target states can be a useful tool in management of such situations. One common property of all presented methods are that they are data-driven. Therefore, these methods are easily applicable to systems where stochasticity makes numerical analysis extremely difficult. While data-driven models facilitate the use of all current knowledge of the system, the downside is that the accuracy of the models depend on the quality and quantity of training data used to tune the models. A limitation of data-driven models come from the increase of computational resources needed for manipulating large volumes of data. The methods presented in this book have always used parallel processing techniques to reduce such effects wherever required.

6.1 Future Research

A range of applications that can exploit the generic methods discussed in this book make way for an array of possible future research topics. Some proposed future research objectives are summarised below with respect to the contributions of this book.

Prediction of Cyclone Induced Flood Propagation

The objective of Chap. 3 was to analyse an end-to-end real world natural disaster dynamic system consisting of interdependent random processes affected by network inhomogeneity and biasing factors. However, this model has given promising results for a basis of a practical cyclone induced flood prediction model which can be improved by:

- Using the modularised approach to integrate outputs of sophisticated prediction models at each stage as and when they become available. This would facilitate the maximum use of the knowledge of the system while having the ability to generate initial predictions well ahead of time with minimal available data.
- In the presented model, several assumptions were made to simplify the calculations. A more complete treatment of the hydrological features affecting propagation can improve the prediction accuracy by relaxing some of the assumptions such as surfaces being frictionless and impermeable.
- The rainfall prediction Bayesian framework can be extended for non-cyclone induced rainfall prediction scenarios.
- The Eigen-value analysis used for the flood prediction stage can be taken in abstract to an array of prediction applications given available data at reasonable resolutions.

MFPT Calculation for Biased Random Walks

The objective of Chap. 4 was to develop generic methods to predict *MFPT* for natural processes translating to random walks under the influence of directional bias.

- Due to the generic nature followed in developing the relations and algorithms, the possible future research applications are vast. As long as the systems' state transitions can be represented as directionally biased random walks, the proposed methods can be applied for these natural disaster systems.
- Although all hypothetical and real case studies used in Chap. 4 concerned 2-dimensional random walks, the methods are developed to be generic and extendible to networks of higher dimensions. A useful future objective is to extend these methods and verify with example datasets.

MFPT Calculation for Random Walks in Inhomogeneous Networks

The methods proposed in Chap. 5 can be extended to a range of natural disaster prediction applications that can be modelled as random walks in inhomogeneous networks with identifiable homogeneous patches. These methods customised to such applications give rise to an array of possible future research projects.

The main extension that can be proposed for methods proposed in Chaps. 4, and 5 is for the *MFPT* prediction for branching random walks and random spreads. Several complications arise when using the proposed methods to random spreads.

Figure 6.1 shows the simulations of fire spread on a hypothetical fuel distribution. A uniform node distribution is considered where nodes in some areas did not carry any fuel, thus making them obstacles for the propagation. Figure 6.1 presents the fire spread at 3, 8, 13 and 18 time-steps. The area with and without fuel are indicated on the figures. The red ellipses show the probability density distributions of the next steps of the spread. The initial bias towards a westerly direction is changed suddenly at time-step 13 towards a south-easterly direction. This is indicative of possible changes of wind direction possible in fire propagation. A change of spread indicative of the bias change is observed in Fig. 6.1d. The transport properties df and dw for this is obtained in methods similar to those used in Chaps. 4 and 5. An extended version of the algorithm to calculate dw for fire propagation is included in Appendix 1 at the end of this chapter.

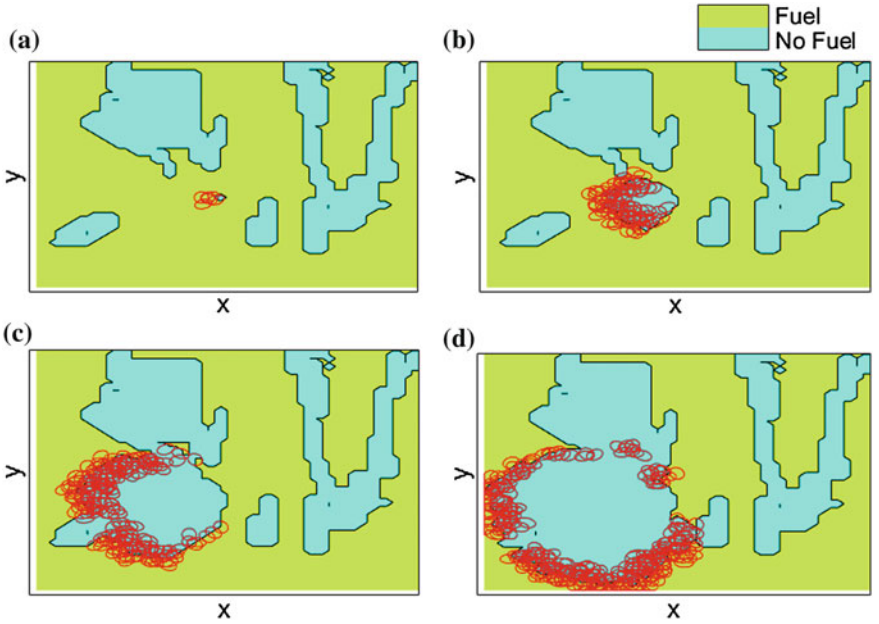


Fig. 6.1 Simulating fire propagation as a random spread to calculate transport dimensions df and dw_b. The fire spread and the available fuel distributions after **a** 3, **b** 8, **c** 13, and **d** 18 time steps

From these figures, some observations present the challenges in extending these methods for predicting fire propagation as a random spread. The large number of steps per time-step results in the increase of computational complexity. This is worsened by the interdependence of simultaneous walkers. For fire propagation, the complexity further increases as the node distribution is dynamic (i.e. the fuel point might be occupied by a parallel walker or run out), and the vulnerability of biasing factors to sudden changes. These are common attributes of real world random processes. Therefore future research can address the higher computational complexity through methods such as parallel processing, distributed computing and optimisation.

Appendix 1: Calculating the Walk Dimension (dw) for a Dynamic Network

The steps followed to calculate the dw at a given source point (S) of a known dynamic 2D network is summarised in Algorithm 7. This is an extension to Algorithm 4 from Chap. 4, including the consideration of dynamicity in network node distribution. The dw is calculated prior to the $MFPT$ calculation and recorded in a matrix with respect to bias intensity ($bias\_range = 0.6 - 1.6$ for the simulation) and relative direction.

Algorithm 7 Obtaining *dw* table for given source node

begin
  $[c_{sx}, c_{sy}]$ = coordinates of source node
  define *max_iterations*, *max_hops*
  initialise vectors *bias_range*, Φ
for $\forall u \in \{bias\_range\}$ **do**
  for $i = 1 : max\_iterations$ **do**
    $obstacle\_reached = 0$
    for $s = 1 : max\_hops$ **do**
      coordinates of current node $\rightarrow [c_{1x}, c_{1y}] = \begin{cases} [c_{sx}, c_{sy}], & \text{if } s = 1. \\ [c_{2x}, c_{2y}], & \text{otherwise.} \end{cases}$
      define $\theta, U \leftarrow f(u), R$
      dynamic node status matrix $\rightarrow F$
      if $obstacle\_reached == 1$ **then**
        break;
      end if
      $P_{DD} \leftarrow f(c_{1x}, c_{1y}, \theta, U, R)$ {Obtain P_{DD} for identifying next hop}
      coordinates of next node $[c_{2x}, c_{2y}] \leftarrow f(P_{DD}, F)$
      if $[c_{2x}, c_{2y}]$==null **then**
        $obstacle\_reached = 1$
        continue
      end if
      $dist = \sqrt{((c_{2x} - c_{1x})^2 + (c_{2y} - c_{1y})^2)}$
      Find angle segment the node belongs to
      $hop\_dir = \pm tan^{-1}((c_{2y} - c_{1y})/(c_{2x} - c_{1x}))$
      $hop\_segment \leftarrow f(angle\_segment\_range, hop\_dir)$
      Update vectors
      if $max(t_{exit}(u, hop\_segment, :)) < dist$ **then**
        $t_{exit}(u, hop\_segment, end + 1) = s$
        $r(u, hop\_segment, end + 1) = dist$
      end if
    end for
  end for
  Update table
  for $\forall seg \in \{angle\_segment\_range\}$ **do**
    $dw(S, u, seg) = dw_n \leftarrow dwcurvefit(t_{exit}, r; t_{exit} = a \times r^{dw_n})$
  end for
end for

The results are divided into angle segments ($\Phi = -\pi : \pi/24 : \pi$ in the simulations) to capture the effect of bias on the ease with which the walker can move in different directions relative to bias. Therefore, the calculation is repeated for the possible range of bias intensities and iterated a large number of times ($max\_iterations = 500$ in the simulations) such that the average *dw* would converge into a stable value. Random walks are generated having a predefined maximum number of hops per walk (*max_hops*). The status of each node is kept in a reference matrix (F), to capture the dynamicity of node availability and to identify obstacles. The next node at each hop is identified using a probability density distribution (P_{DD}) which is a function of bias intensity (U), bias direction (θ) and the rate of spread (R) that is the

speed of the walk without bias and is a property of the network. For the simulated random walks in the main text, the Length-to-Breadth ratio (*LB*), Head-to-Back ratio (*HB*), length of major axis (b_e), length of minor axis (a_e) and the shift of centre in the direction of bias (c_e) (Fig. 4.1b) were obtained using

$$\begin{aligned} LB &= e^{0.25U}, \\ HB &= LB^{2.5}, \\ a_e &= 0.5 * (R + R/HB)/LB, \\ b_e &= (R + R/HB)/2, \& \\ c_e &= b - R/HB \end{aligned} \tag{6.1}$$

respectively. The equations for *LB* and *HB* are taken such that for an environment without bias, *LB* would be 1 and *HB* would be an augmentation of *LB*. The coefficients for *LB* and *HB* were selected at random. The orientation of the ellipse is equal to θ. The Gaussian P_{DD} for selecting the next node is

$$P_{DD}(x, y) = exp(-(d_e(x - C_x)^2 + 2e_e(x - C_x)(y - C_y) + f_e(y - C_y)^2)) \tag{6.2}$$

where (C_x, C_y) are the coordinates of the centre of the ellipse obtained by shifting the current position a distance of c in the θ direction. $d_e = cos^2\theta/2\sigma_x^2 + sin^2\theta/2\sigma_y^2$, $e_e = -sin2\theta/4\sigma_x^2 + sin2\theta/4\sigma_y^2, f_e = sin^2\theta/2\sigma_x^2 + cos^2\theta/2\sigma_y^2$ where σ_x and σ_y, the standard deviations in x and y directions, are chosen from a_e and b_e depending on θ. The next node is selected using P_{DD} and F. The distance and relative direction for the new node is calculated and recorded alongside the hop number (i.e. time-to-reach) with respect to the angle segment the new node belongs to. Finally dw is calculated using curve fitting (*dwcurvefit*) for time-to-reach (t_{exit}) versus distance (r) using linear least squares method.

Glossary

Adjacency matrix A matrix of which the element $A_{ij} = 1$, if there is a link from i to j ($A_{ii} = 0$ otherwise).

Anisotropy The property of being directionally dependent, as opposed to isotropy.

Anomalous Abnormal/deviating from what is standard, normal, or expected.

Anomalous diffusion Mean square displacement (r^2) is no longer proportional to t.

Asymptotic Approaching a value or curve arbitrarily closely.

Biased-random walk A random walk which is more likely to follow one path than another.

Bosonic systems Network nodes can be occupied by any number of walkers simultaneously.

Branching random walk A random walk where at each step, several steps are branched out.

Connected graph There exists a path from every node to every other.

Degree of node The number of links connected to a node.

Disordered media Structures with highly irregular geometry (e.g. Fractals).

Heuristic Approach in problem solving not guaranteed to be optimal or perfect but sufficient for immediate goals.

Levi Flights Step lengths $\rightarrow$ heavy tailed.

MCT Mean Coverage Time: Time taken to cover all nodes of a network.

MFPT Mean First Passage Time: Expected time for a random walks to first reach a target or set of targets.

MRT Mean Return Time: Expected time for a random walk to return to the source node.

Numerical weather prediction (NWP) Prediction modes that use mathematical models of the atmosphere and oceans to predict the weather based on current weather conditions.

Random walk A walk which follows no discernible pattern or trend.

Relaxation time The time taken for a system to converge to steady state (or a stationary distribution).

Scal-invariant network Networks where properties are conserves at all scales.

I. Wijesundera et al., *Natural Disasters, When Will They Reach Me?*,
Springer Natural Hazards, DOI 10.1007/978-981-10-1113-9

Scale-Free network A network where the degree distribution follows the power law (i.e. degree of connections declines as the power law [e.g internet]).

Self similar network Have exactly the same pattern at an infinite number of length scale (if scale-invariant networks have the same property in average at all scales, self-similar networks like deterministic fractals have exactly the same pattern at an infinite number of length scale).

Short relaxation time Networks with non compact exploration.

Small-world type network Diameter scales like the logarithm of volume (A random walk on such network can reach a given node from another one, following the path with the smallest number of links between the nodes, in a very small number of steps).

Stochastic Can be determined randomly; usually has a random nature that may be analysed statistically but may not be predicted precisely.

Undirected network A network where $A_{ij} = A_{ji}$.

Index

I. Wijesundera et al., *Natural Disasters, When Will They Reach Me?*,
Springer Natural Hazards, DOI 10.1007/978-981-10-1113-9