

**ECOLOGICAL DYNAMICS OF VECTOR-BORNE DISEASES IN CHANGING
ENVIRONMENTS**

by

Luis Fernando Chaves

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in The University of Michigan
2008

Doctoral Committee:

Professor Mark L. Wilson, Co-Chair
Associate Professor Mercedes Pascual, Co-Chair
Professor John H. Vandermeer
Assistant Professor Edward L. Ionides

“Se puderes olhar, vê. Se podes ver, repara”

Jose Saramago

© Luis Fernando Chaves

All rights reserved

2008

To my grandfather (Don Fernando),
with love for his support and advice

ACKNOWLEDGEMENTS

I am indebted to many people, primarily to my committee, they helped me in directing and defining my research agenda for this dissertation. First I'd like to thank my two chairs, Dr. Mercedes Pascual and Dr. Mark Wilson. They are very different in almost everything you can imagine, with the exception of being excellent scientists. With one I'm very grateful for the financial support, technical advice and input on scientific questions. With the other I'm grateful for the inspiration to build a comprehensive and dialectic epistemological framework for my career, and for the constant support through several steps in this segment of my journey in Academia. Dr. Ed Ionides was very helpful through his classes and time to instruct me with tools that are fundamental to the analysis of the problems presented here, as well as with the insights and perspectives of somebody with a completely different training. Dr. John Vandermeer is definitively one of my most admired colleagues, he showed me that being a successful ecologist can only be enhanced by being actively involved in the context of our objects/subjects of study.

I am also very grateful to all the comrades from the three discussion groups I regularly attended during my stay at the University of Michigan: Plant Animal interactions (Rathcke and Hunter Labs), Tropical Biology (Perfecto, Dick and Vandermeer labs) and New World Agriculture and Ecology Group (too many

people to mention). These groups widened my perspectives on the science of Ecology and led me to an intense collaboration with Sandra Yap, Rachel Vanette, Brian Sedio, Javier Ruiz and Shalene Jha, to whom I'm deeply grateful for all what they taught me about their research and for inviting me to apply my knowledge in areas outside where it is usually applied. Besides, all of them are among the best friends somebody could wish. Casey Taylor, Sandra Yap and Brian Sedio were also very generous with their time carefully reading and editing some of my manuscripts. I am also grateful to my labmates from the Pascual and Wilson labs, especially Justin Cohen for sharing ideas on spatial analysis and his input on epidemiological methods, and his generous help leading to the publication of Chapter VI.

I am also grateful to all those friends with whom I kept contact and constantly gave me input on my research: Chepina Hernandez and Diego Rodriguez at Universidad Central de Venezuela; Tomas Revilla at Groningen University, Sander Koenraadt at Wageningen University, Menno Bouma at London School of Tropical Medicine and Hygiene, Andy Dobson at Princeton University and Dick Levins at Harvard School of Public Health. Christian Lengeler and Tom Smith helped me with their expertise on malaria bednets (chapters IV & V) during a visit to the Swiss Tropical Institute. Akira Kaneko and Anders Björkman gave significant input to chapters IV and V, and generously invited me

to give a series of Lectures in their Malaria course at Karolinska Institutet in Stockholm during fall 2007, and to eventually join their group.

Finally, I want to thank all the good friends I met in Ann Arbor: Liz Vivas, Jose Manuel Gomez, Jasun Gong, Courtney Murdock, David Alonso, Katie Pethan, Jerry Smith, Catherine Badgley, Robyn Burham, Dan Green, Alan Stapledon, Stefano Allesina, Lia Florey, Liz Lehman, Pete Boldenow, Jose Siri, Jahi Chappell, Kacey Ernst, Ting-Wu Chuang, Ben-Yang Liao, Dave Nicoletti Allen, Michael H. Reiskind, Anne Lippert, Doug Jackson, Julie Maslovsky, Eladio Marquez, Roosevelt Garcia, C. I. Flores, Diego Fernando Alvarado, Susana Pereira, Kat Connors and Natalie Hubbard. They made joyful and pleasant the time in Ann Arbor.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF APPENDICES	xii
ABSTRACT	xiii
CHAPTER I INTRODUCTION	
ECOLOGICAL NATURE OF HUMAN DISEASES	1
MALARIA AND AMERICAN CUTANEOUS LEISHMANIASIS	5
RESEARCH AGENDA	7
REFERENCES	11
CHAPTER II CLIMATE CHANGE AND THE ABILITY TO FORECAST DISEASES: I. NON-STATIONARY PATTERNS FOR AMERICAN CUTANEOUS LEISHMANIASIS IN COSTA RICA	
INTRODUCTION	15
METHODS	17
RESULTS	20
DISCUSSION	22
REFERENCES	34
CHAPTER III CLIMATE CHANGE AND THE ABILITY TO FORECAST DISEASES: II. NON-LINEAR TOOLS APPLIED TO AMERICAN CUTANEOUS LEISHMANIASIS FORECASTS IN COSTA RICA	
INTRODUCTION	38
METHODS	40
RESULTS	43

	DISCUSSION	44
	REFERENCES	54
CHAPTER IV	MODELLING DISEASES FROM FIRST PRINCIPLES: SELF REGULATORY FEEDBACKS, BIOLOGICAL INTERACTIONS AND CLIMATIC FORCING, THE ECOLOGICAL DYNAMICS OF VIVAX AND FALCIPARUM MALARIA	
	INTRODUCTION	59
	METHODS.....	62
	RESULTS.....	70
	DISCUSSION	71
	REFERENCES	82
CHAPTER V	RESILIENCE TO THE EFFECTS OF ENVIRONMENTAL VARIABILITY: BEDNETS AND MALARIA IN VANUATU	
	INTRODUCTION	90
	METHODS.....	92
	RESULTS.....	98
	DISCUSSION	101
	REFERENCES	119
CHAPTER VI	THE ECOLOGICAL DYNAMICS OF VECTOR-BORNE DISEASES UNDER THE EFFECTS OF A MULTIDIMENSIONAL CHANGING ENVIRONMENT: SPATIAL PATTERNS OF CUTANEOUS LEISHMANIASIS; THE ROLE OF SOCIAL MARGINALITY AND DEFORESTATION IN COSTA RICA	
	INTRODUCTION	127
	METHODS.....	129
	RESULTS.....	135
	DISCUSSION	138
	REFERENCES	157

CHAPTER VII CONCLUSIONS

SUMMARY OF MAJOR FINDINGS AND RESEARCH IMPLICATIONS..... 162
SUGGESTIONS FOR FUTURE RESEARCH..... 163
REFERENCES 165

APPENDICES..... 167

LIST OF TABLES

Table 2.1. Model Selection and Parameter Values	25
Table 3.1. Models and predictive R^2	50
Table 4.1. Model Search.	76
Table 4.2. Parameter estimates for the best models	77
Table 5.1. Parameter values and % reduction for <i>Plasmodium falciparum</i> and <i>Plasmodium vivax</i> rate before and after the breakpoint obtained by using the variance of the Kolmogorov-Zurbenko adaptive filter.	109
Table 6.1. Breakpoint values for the natural logarithm of minimum rainfall, the marginalization index, and the percentage of people living close to the forest for the studied models.....	142
Table 6.S1. Ecosystems and number of locations where human biting sand fly species have been caught in Costa Rica (S6, S7)	143
Table 6.S2. Principal Components Analysis (PCA) for the landscape units where human biting sand fly species have been caught in Costa Rica	144
Table 6.S3. Factor Loadings for Ecosystems in components 1 & 2	145
Table 6.S4. Parameters, smooth function degrees of freedom and significance for the GAM described in equation 1.....	146
Table 6.S5. Comparison of Linear Mixed Effects models	147
Table 6.S6. Parameters for the model presented in Figure 6.2.....	148
Table 6.S7 Analysis of Covariance for the model in (5)	149
Table 6.S8 Parameters for the linear model in (5).....	150

LIST OF FIGURES

Figure 2.1. Time Series	27
Figure 2.2. Dominant Frequencies in the Data	28
Figure 2.3. Cross-Wavelet Coherency and Phase.....	29
Figure 2.4. Cross-correlations and Predictive R^2	30
Figure 2.S1. Autocorrelation function.....	31
Figure 2.S2. Cross-Wavelet Coherency and Phase	32
Figure 2.S3. Mean Square Error.....	33
Figure 3.1. Time Series	51
Figure 3.2. Multidimensional plots for the square root transformed ACL cases	52
Figure 3.3. Bootstrap experiment.....	53
Figure 4.1. Feedback loops for top-down and bottom-up regulation of multi-species malaria infections.....	78
Figure 4.2. Time Series	79
Figure 4.3. Bednets and Parasite Clearance	80
Figure 4.4. Fitted vs Observed values for the best models.....	81
Figure 5.1. Time Series	110
Figure 5.2. Regime Shift for falciparum and vivax malaria.....	111
Figure 5.3. Bednets	112
Figure 5.4. Cross-Wavelet Coherency and Phase.....	113

Figure 5.S1. Time series.....	114
Figure 5.S2. Breakpoints for the rate of slide examination	115
Figure 5.S3. Breakpoints for the <i>Plasmodium falciparum</i> and <i>P. vivax</i> rates	116
Figure 5.S4. Breakpoints for Temperature and Rainfall.....	117
Figure 5.S5. Cross-Wavelet Coherency and Phase <i>Plasmodium vivax</i> - <i>P. falciparum</i> rates	118
Figure 6.1. Patterns of clustering and Schmalhausen’s law	151
Figure 6.2. Breakpoints and discontinuous patterns of association.	152
Figure 6.3. Cutaneous leishmaniasis in Costa Rica: Deforestation and El Niño Southern Oscillation (ENSO).....	153
Figure 6.S1. Supplementary maps	154
Figure 6.S2. Generalized Additive Model smooth functions	155
Figure 6.S3. Landscape dimension reduction.....	156

LIST OF APPENDICES

Appendix S1	167
Appendix S1	173
Appendix S1	179
Appendix S1	181

ABSTRACT

ECOLOGICAL DYNAMICS OF VECTOR-BORNE DISEASES IN CHANGING ENVIRONMENTS

by

Luis Fernando Chaves

Co-Chairs: Mark L. Wilson and Mercedes Pascual

One of the major threats for the current functioning of the world we know is the uncertainty about the effects of global climate change. This dissertation aims to understand the effects of a changing environment on the ecological dynamics of vector-borne diseases, one of the major burdens for human populations worldwide. Vector-borne diseases are expected to be highly sensitive to the effects of climatic change, because of the natural history of both the vectors and parasites, which are highly sensitive to small changes in precipitation and temperature.

This dissertation investigates several aspects of the effects of changing environments in vector-borne diseases: (i) The plausibility of early warning systems to predict the future dynamics of a disease based on its association to climatic forces, using a time series for cutaneous leishmaniasis cases from Costa Rica, 1991-2001 (ii) The mechanisms regulating co-infections of malaria parasites

in the island of Santo, Vanuatu, 1983-1997, and (iii) Abrupt dynamical changes in diseases along smoothly changing environments, temporally for malaria in the archipelago of Vanuatu, 1983-1999, and spatio-temporally for cutaneous leishmaniasis in Costa Rica, 1996-2000. Methods involved a suite of qualitative and quantitative techniques in order to robustly assess the reliability of results. Frequency, time and time-frequency domain statistical techniques for time series analysis were used to study the association between disease dynamics and climate, time models predictive ability for early warning systems was tested with “out-of-fit” data. Signed digraph loop analyses and quantitative discrete time models were used to discern working hypothesis about parasite species co-infection regulation. Statistical techniques for breakpoints were used to study abrupt dynamical changes. In addition, spatial clustering techniques were used as guidance to establish transmission risk factors.

Results show that early warning systems are feasible goals, that malaria parasites and their interactions seem to be regulated in a bottom-up fashion, and that abrupt changes on the sensitivity to the effects of climate change are dependent on the context of transmission. Finally, all the results confirm the importance of considering the whole environmental context in which vector-borne diseases are transmitted and the need for abstraction to understand and manage the underlying complexity.

CHAPTER I

INTRODUCTION

“Malaria, like every other epidemic, obeys the universal law of rhythm, that is to say, it manifests itself in periodical cycles, in the course of centuries, of years and months ”

Celli (1933)

Ecological Nature of Human Diseases

Understanding the universal laws of rhythm that drive population numbers in time and space is one of the main themes of Ecology as a scientific discipline. As one of the main themes that define the research agenda of a group of people with different backgrounds and perspectives about nature, explanations for the observed patterns of population abundance across diverse landscapes have had very different “plausible” explanations through time, sometimes in opposition. In ecology, the classical example for this opposition was the discussion held in 1957 at the Cold Spring Harbor Symposium on quantitative Biology, where Nicholson [1], who worked with lab populations of moths, criticized the lack of density-dependent mechanisms for changes in populations whose growth was forced by environmental factors [2,3]. These irreconcilable ideas, through the synthesis of evidence, eventually led to the proposal that all populations need to

have regulatory mechanisms for negative feedback, which are sensitive to the forcing by changing environments [4,5].

Vector transmitted diseases (VTDs), those diseases for which the transmission of a pathogen between two individuals of the same species (in some cases different species) is mediated by an individual of a second species (generally an insect), are one of the main problems currently faced by humankind. The main three vector transmitted diseases affecting humans, Malaria, Schistosomiasis and the Leishmaniases, are in part responsible for a diminished quality of life (shorter life span, poverty, etc) in at least one tenth of the world population[6]. VTDs are also phenomena that involve both organisms and their environment. This fundamental property makes them suitable to ask similar questions to those that have been the object of study by ecologists for a long time. For example, what determines their distribution and prevalence?

In general, the main approach to answer this question is one that tends to separate factors, assessing their relative importance. This approach is probably a by-product of assigning a meaning to the output of classical quantitative techniques like linear statistical models [7].

However, for any complex system, like a vector transmitted disease as malaria, the relevant question is not what factor(s) is (are) the most important in explaining any given pattern [8]. Such a perspective, the search for the main or most important factor (variable), basically biases any understanding of the complexity of the underlying processes. This is especially true for complex systems where relationships between variables can be described by non-linear

functions, and slight changes can lead to very different patterns, independently of whether systems are seen from a dynamic or static perspective [8,9]. On one scale, the differences in the life history of the species involved in the life cycle of the pathogen, and the pathogen itself, can have different sensitivities to both exogenous and endogenous factors. On a different scale, life history differences among individuals of the same species become relevant. In addition to these aspects of heterogeneity, organisms by themselves can actively modify the environment where they live, making it necessary to understand populations and their environment as a unity where both components dynamically interact with each other [8,10,11]. This view poses a challenge to our understanding of the complex ways in which both exogenous and endogenous factors interact to generate disease. This challenge is not only restricted to diseases, but applies to the interplay of the multiple components of an ecosystem [12]. Moreover, only a few elements of a system can be studied at the same time, either for logistic reasons (such as the impossibility of gathering and analyzing data for sufficiently powerful statistical tests), or by simple ignorance [13,14,15]. Thus, the challenges for understanding the dynamics of an ecological system such as a vector borne disease are enhanced by the limitations imposed by the methods of study.

There are two known strategies to approach the dynamics of an ecological system. Both strategies are similar because they rely on the isolation of specific components. The first strategy, which is mainstream in most contemporary scientific practice, is the reduction of the nature of a complex problem to one of

its specific parts, without recognizing the intrinsic abstraction that such a simplification implies. This strategy in general resorts to panglossian explanations [16], e.g., the search for meanings in definitions necessary to apply statistical or mathematical methods, sometimes creating the illusion of an ultimate explanation for patterns in nature, and potentially contributing to the delegitimatization of science as an evidence-based approach to understand the world. This strategy to define problems is very common in the field of disease ecology and ecology in general [e.g., 17]. This strategy is also a known pathway to failures. For example, the view of malaria as driven primarily by mosquitoes' abundance led to the spraying of insecticides. The latter led to resistant mosquitoes that are no longer susceptible to an otherwise successful control measure. The reduction of malaria by control measures targeting the pathogen led to the abuse of antimalarial-drugs, which eventually drove the drug resistance of parasites, with infections no longer controllable by drugs. Countless examples of these outcomes can be seen everywhere. The second strategy, less common, begins with the realization of the limitations of isolating a specific phenomenon from the wholeness of nature, and the need for abstraction in order to make the object of study apprehensible. Although the best tools are always limited to explain even a small part of nature, the use of several tools to study the same problem as seen from various perspectives can lead to the discovery of processes and mechanisms that can explain regular patterns in a robust way, i.e., independently of the method or its assumptions [13, 18].

With this framework in mind, a description of the biology relevant to the hypotheses in this dissertation regarding the two vector transmitted diseases, American Cutaneous Leishmaniasis and Malaria, is presented next.

Malaria and American Cutaneous Leishmaniasis

Although both diseases are caused by parasites transmitted by insects that fly, and both infect humans, they are quite different in several aspects. Malaria is caused by 4 species of parasites belonging to the genus *Plasmodium* (Haemosporida: Plasmodiidae); *P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale*. It is transmitted by mosquitoes (Diptera: Culicidae) in the genus *Anopheles* and, sometimes, can be lethal. By contrast, American Cutaneous Leishmaniasis is caused by at least 14 species belonging to genus *Leishmania* (Kinetoplastida: Trypanosomatidae) and grouped in the subgenera *Leishmania* and *Viannia*. The parasites are transmitted by sand flies (Diptera: Psychodidae) in the genus *Lutzomyia*. Infection generally leads to disfiguring cutaneous/muco-cutaneous lesions. ACL is a zoonotic vector transmitted disease, with several vertebrate host species besides humans [19] and evidence suggests that humans are sink hosts, i.e., they do not transmit the disease back to sand flies [20, 21]. Malaria is a vector transmitted anthroponosis [19], where humans are the only hosts that can infect vectors.

As a product of history a gap in the knowledge between the two diseases exist. American Cutaneous Leishmaniasis, as its name indicates, is a disease

restricted to the New World. Even though there are pottery samples from pre-columbian times where people are represented with cutaneous lesions similar to those caused by leishmania parasites and some Native American languages have names for it, no written records from those times are known [22]. Some records from early colonial times describe a common disease where people have lost their noses [23]. This might indicate that from very ancient times this infection was common. As of today, the Leishmaniases (including all the clinical forms present in the New World and Old World) represent the fourth most important neglected disease, with a burden of at least 2.1 million of infected people per year [6]. ACL occurs mostly in the tropical and sub-tropical regions of the new world [24].

By contrast, malaria which is currently considered an almost exclusively tropical disease, used to have a wider distribution. It is well documented that Romans knew the relationship between marshes and “malaria like” fevers, and used it in the selection of places for new settlements and military camps. Early human history is full of battles that were won with the aid of malaria as an additional platoon attacking enemies [25-31]. Even during the coldest years of the little ice age (1560s to 1730s) reports of malaria outbreaks in England and Scotland were common. In the former, these were associated with years of high famine [32], and the latter having its maxima during the warmest and wettest summers of this period [33]. After World War II malaria endemic transmission was erased from the USA, and since 1973 Europe has been freed of the disease [28,29,34]. More recently, in 1981, Australia, a large subtropical area, also has

been freed of endemic malaria transmission [35]. However, as of today, every 40 seconds a child dies of malaria, and between one to three million people die each year around the world, most of the cases reported for sub-saharan Africa [36].

Based on this background information, I describe next the specific goals and strategic approaches for this dissertation.

Research Agenda

A) The need to define the problem from a wider perspective

To introduce the research agenda for this dissertation, a brief description of the work by Clifford Allchin Gill [37-38] will be discussed. His vision of the malaria problem at a relatively early time, was to some extent above the level of the ecological debate later developed by Andrewartha/Birch and Nicholson, as he was aware of the need for a synthetic approach in order to establish effective warning systems for disease epidemics.

Gill [37] realized that the dynamics of a vector transmitted can be assessed from 4 different perspectives. Each perspective emphasizes a factor, namely: the human host, the parasite, host immunity, and transmission.

Under the human factor, this author included several human activities that can drive the dynamics of the disease and that are modulated primarily by socio-economic reasons. For example, he noticed the importance of human movement for the synchronicity of malaria epidemics in his region of study, the Punjab [37].

He also reported previous findings that demonstrated the exacerbation of epidemics at times of extreme and generalized famine.

For the parasite factor, he basically stated that the patterns seen in epidemics depend on the specific parasite species causing the infections. The example he used was that of significant variation in seasonal patterns for *P. vivax* vs. *P. falciparum*, since the former species can have a delayed onset of clinical manifestations and relapses, i.e., clinical manifestations after a period without symptoms [38].

The consideration of the immunity factor was very innovative and insightful for the little immunological knowledge at that time, 1928. Gill's inspection of data on spleen enlargement in children allowed him to realize that another important factor was the degree of protection against infection by a parasite that the human population has as a whole. His main hypothesis was that when these rates were relatively small the likelihood of a large epidemic increased, because of the lack of protection against the disease.

The transmission factor included the effect of mosquitoes on the seasonality and synchronicity of malaria epidemics in the British Punjab [37], and their relationship to both climate variables and landscape structure.

In his two books, Gill [37,38] emphasized that malaria epidemics result from the interplay of all four factors and this led him to propose a qualitative forecasting system based on proxy-measurements of these four factors. This system was successful in predicting the relative magnitude of epidemic outbreaks for malaria in the Punjab. Finally, the work by Gill [37,38] shows that

there is an established need to pose the problem from a large enough perspective to get a clear picture of the ecology of vector transmitted diseases in changing environments.

B) Research goals

In general, the main goal of this dissertation will be to study the ecological dynamics of vector-borne diseases in changing environments. In the context of this dissertation ecological dynamics will be defined as changes in population size in both space and time.

The dissertation is composed of six additional chapters to this introduction that illustrate more specific goals. Chapters II & III illustrate the importance of climatic variability at the local and global scale as a driver of the dynamics of vector-borne diseases. Chapter II addresses the importance of climate by analyzing a time series for infections of American Cutaneous Leishmaniasis in Costa Rica from 1991 to 2001. It addresses how the cycles in this series reflect changes in environmental variables associated with the natural history of the disease and examines the accuracy of predictions of linear models including and excluding climatic co-variates [39]. Chapter III extends the analysis to non-linear models for forecasting and presents epistemological reflections on how to use the information on the relationship between disease and climatic forces, as well as guidelines for testing their robustness [40].

In Chapter IV the analysis goes one step further by fitting simple models to understand the regulatory mechanisms for the transmission of vector borne diseases, examining the importance of several aspects known to affect their

transmission and the co-occurrence of related parasites in a host population. More specifically, this chapter aims to abstract the forces regulating the dynamics of malaria caused by two co-occurring parasites, *Plasmodium falciparum* and *P. vivax*, on the island of Espirito Santo, Vanuatu, to understand how is the feedback between hosts and population level immunity which would mediate the interactions between the two parasites, as well as the role of climate on the infections by each parasite [41].

Chapters V and VI have the common theme of describing abrupt qualitative changes in populations coping with gradually changing environments. Chapter V describes techniques to find abrupt changes (breakpoints) in time series, and shows how the importance of climatic drivers for malaria dynamics in the archipelago of Vanuatu changed qualitatively after the introduction of bednets, both in mean value and variability [42]. Chapter VI extends these ideas, but emphasizes the need to understand the context of changing environments in human diseases [43, 44] . It also extends the level of analysis to include the spatial component. This chapter presents techniques to identify breakpoints in the relationship between the rate of a disease and predictors when the relationship is non monotonic. These approaches are applied to identify the factors that determine the distribution of the disease at country scale, and how deforestation interacts with climatic forces at the smaller spatial scales where the disease is clustered [45]. Finally, Chapter VII presents the conclusions of this work, its implications for the control of infectious diseases and pathways for future research.

References

1. Nicholson AJ (1957) The self adjustment of populations to change. Cold Spring Harbor Symp. Quant. Biol. 22:153-173.
2. Birch LC (1957) The role of weather in determining the distribution and abundance of animals. Cold Spring Harbor Symp. Quant Biol. 22:203-218.
3. Andrewartha HG (1957) The use of conceptual models in Ecology. Cold Spring Harbor Symp. Quant. Biol. 22:219-236.
4. Royama T (1992) Analytical Population Dynamics. Chapman and Hall: London.
5. Turchin P (2003) Complex population dynamics. Princeton: Princeton University Press.
6. Hotez PJ, Molyneux DH, Fenwick A, Ottesen E, Sachs SE, Sachs JD (2006) Incorporating a rapid impact package for neglected tropical diseases with programs for HIV/AIDS, tuberculosis and malaria. PLoS Med. 3: e102.
7. Faraway JJ (2005) Linear models with R. Boca Raton: Chapman and Hall/CRC
8. Levins R (1995) Toward an Integrated Epidemiology. Trends Ecol. Evol. 10: 304
9. May RM, Oster GF (1976) Bifurcations and Dynamic complexity in simple ecological models. Am. Nat. 110: 573-599.
10. Levins R (1979) Coexistence in a variable environment. Am. Nat. 144: 765-783.
11. Lewontin RC (2000) The Triple Helix: Gene, Organism and Environment. Cambridge: Harvard University Press
12. Levin SA (1992) The problem of pattern and scale in ecology. Ecology. 63: 1943-1967.
13. Levins R (1966) The strategy of model building in population biology. Am. Sci. 52: 421-431.
14. Levins R (1993) A response to Orzack and Sober: Formal analysis and the fluidity of science. Quart. Rev. Biol. 68: 547-555.

15. Levins R, Lewontin RC (1985) *The dialectical Biologist*. Cambridge: Harvard University Press.
16. Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian Paradigm: a critique of the adaptationist program. *Proc. Roy. Soc. London B*. 205: 581-598.
17. Orzack SH, Sober E (1993) A critical assessment of Levins's The strategy of model building on population biology (1966). *Quart. Rev. Biol.* 68: 533-546.
18. Levins R (2006) Strategies of Abstraction. *Biol & Philos* 21:741-755.
19. Committee on Climate, Ecosystems, Infectious Disease, and Human Health (2001) *Under the weather: climate, ecosystems, and infectious disease*. Washington DC: National Academy Press.
20. Ashford RW (1997) The Leishmaniasis as model zoonoses. *Ann Trop Med Parasitol.* 91: 693-710.
21. Chaves LF, Hernandez M-J, Dobson AP, Pascual M (2007) Sources and sinks: revisiting the criteria for identifying reservoirs for American cutaneous leishmaniasis. *Trends. Parasitol.* 23, 311-316.
22. Martens R (1999) Una Aproximación Antropológica a la Enfermedad de la Leishmaniasis en la Cordillera Andina de Mérida *Talleres* 6, 45-73.
23. Pardo IJ (1955) *En esta tierra de Gracia: Imagen de Venezuela en el siglo XVI*. Caracas: Litografía Elite.
24. Silveira FT, Lainson R, Corbett CEP (2004) Clinical and immunopathological spectrum of American Cutaneous Leishmaniasis with special reference to the disease in the Amazonian Brazil – A review. *Mem. Inst. Oswaldo Cruz.* 99: 231-251.
25. Jones WHS (1907) *Malaria a neglected factor in the history of Greece and Rome*. Cambridge: MacMillan & Bowes (Introduction by R. Ross and Conclusion by G.G. Ellet)
26. Celli A (1933) *The history of malaria in the roman campagna from ancient times*. London: John Bale, Sons & Danielsson, Ltd.
27. De Zulueta J (1973) Malaria and Mediterranean History. *Parassitologia.* 15: 1-15
28. De Zulueta J (1987) Changes in the geographical distribution of malaria throughout history. *Parassitologia* 29:193-205.

29. De Zulueta J (1994) Malaria and ecosystems: from prehistory to posteradication. *Parassitologia*. 36: 7-15.
30. Najera JA (1994) The control of tropical diseases and socioeconomic development (with special reference to malaria and its control). *Parassitologia* 36: 17-33.
31. Sallares R (2002) *Malaria and Rome. A history of malaria in ancient Italy*. Oxford: Oxford University Press.
32. Reiter P (2000) From Shakespeare to Defoe: Malaria in England during the Little Ice Age. *Emerg. Inf. Dis.* 6: 1-11.
33. Duncan K (1993) The possible influence of Climate on historical outbreaks of malaria in Scotland. *Proc. R. Coll. Physicians. Edinb.* 23: 55-62.
34. Snowden FM (1999) 'Fields of Death': malaria in Italy, 1861-1962. *Modern Italy*. 4:25-57.
35. Bryan JH, Foley DH, Sutherst RW (1996) Malaria transmission in Australia. *Med. J Australia* 164: 345-347.
36. Sachs J, Malaney P. 2002. The economic and social burden of malaria. *Nature*. 415: 680-685.
37. Gill CA (1928) *The genesis of epidemics and the natural history of disease. An introduction to the science of epidemiology based upon the study of epidemics of Malaria, Influenza and Plague*. London: Bailliere, Tindall and Cox.
38. Gill CA (1938) *The seasonal periodicity of malaria and the mechanism of the epidemic wave*. London: J. & A. Churchill Ltd.
39. Chaves LF, Pascual M (2006) Climate Cycles and Forecasts of Cutaneous Leishmaniasis, a Nonstationary Vector-Borne Disease *PLoS Medicine*. 3: e295
40. Chaves LF, Pascual M (2007) Comparing Models for Early Warning Systems of Neglected Tropical Diseases. *PLoS NTDs* 1, e33
41. Chaves LF, Kaneko A, Björkman A, Pascual M (submitted) Random, top-down or bottom-up co-existence of parasites: explaining the dynamics of malaria in multi-parasitic settings

42. Chaves LF, Kaneko A, Taleo G, Pascual M, Wilson ML (submittedb) Malaria transmission pattern resilience to climatic variability is mediated by insecticide treated nets.
43. Lewontin R, Levins R (2007) *Biology under the influence*. New York: Monthly Review Press.
44. Wilson ML (1994) Developing paradigms to anticipate emerging diseases. *Ann. N.Y. Acad. Sci.* 740: 418-422.
45. Chaves LF, Cohen JM, Pascual M, Wilson ML (2008) Social Exclusion modifies climate and deforestation impacts on a Vector-borne disease *PLoS NTDs* 2, e176

CHAPTER II

CLIMATE CHANGE AND THE ABILITY TO FORECAST DISEASES: I. NON-STATIONARY PATTERNS FOR AMERICAN CUTANEOUS LEISHMANIASIS IN COSTA RICA

Introduction

The leishmaniasis are among the most important emerging and resurging vector-borne protozoal diseases, second only to malaria in terms of the number of affected people. Like malaria, the leishmaniasis can be caused by infection with any of several species of parasites belonging to the genus *Leishmania* (Kinetoplastida: Trypanosomatidae) and transmitted by sand flies (Diptera: Psychodidae) [1–3]. The disease has two main clinical manifestations: visceral and cutaneous/mucocutaneous. The latter is caused by at least 14 different species of parasites belonging to the subgenera *Viannia* and *Leishmania* [4]. However, it has been suggested that the cutaneous manifestation can be due to heterogeneities in the hosts [5]. An interesting aspect of the dynamics of this disease is that it is strongly associated with the presence of reservoirs, animals that act as sources and sinks of infection to sand flies, while humans are considered to be only incidental hosts, i.e., sinks for infections [3,6]. As with other vector-borne diseases, seasonal patterns in cases and vector abundance suggest that cutaneous leishmaniasis (CL) transmission is sensitive

to climatic exogenous factors. More specifically, vector density is correlated to climate variables—producing seasonal patterns that have been widely described [7–10]—vector density is correlated with number of cases [9,11], transmission is restricted to wet and forested areas [12], and vector density diminishes with altitude [13]. Also, interannual climatic events related to the El Niño Southern Oscillation (ENSO) have been shown to be associated with outbreaks of visceral leishmaniasis at the annual level [14]. However, the interannual cycles of CL have not been examined, and disease data have not been considered at the higher temporal resolution of months, more relevant to the seasonal dynamics of transmission. While many studies of climate–disease couplings have emphasized the potential application of associations between climate and disease to early warning systems, the forecasting ability of the resulting models has not been systematically evaluated. This critical step should be carried out with “out of fit” data if such models are to become useful tools to guide public health policy [3]. In addition, analyses of climate–disease relationships must take into account a common property of disease data that can mask the patterns: time series of cases are typically nonstationary, with changes in the mean and/or the variance over time.

In the present paper, CL cycles and their relationship to climate variables are described, and linear statistical models that use climate variables as predictors are used to assess the accuracy of forecasts based on climatic variables.

Methods

Data

Monthly records of CL cases, from January 1991 to December 2001, were obtained from the epidemic surveillance service, Vigilancia de la Salud, of Costa Rica. Data were normalized using a square root transformation. Sea Surface Temperature 4 (SST 4) (also known as the Niño 4 index; <http://www.cpc.ncep.noaa.gov/data/indices>) and the average temperature of the $0.5^{\circ} \times 0.5^{\circ}$ grids corresponding to the Costa Rica land surface (<http://www.cru.uea.ac.uk>; [15]) were used as climatic variables to study the patterns of association between climate and CL. Temperature, SST 4, and the Multivariate ENSO Index (MEI) (<http://www.cdc.noaa.gov/people/klaus.wolter/MEI>; [16]) were used as predictors for forecasting CL cases. All the time series are presented in Figure 2.1.

Statistical Analysis

Seasonality.

The seasonality of CL cases was assessed by using a box diagram (see Figure 2.1E) [17].

Interannual cycles of CL. A time series is stationary if it has a constant mean and variance [17–19]. Therefore, a nonstationary time series is one whose mean and/or variance is nonconstant. As Figure 2.1A shows, CL is nonstationary as it has a changing mean, an observation that can be confirmed by inspecting its autocorrelation function (see Figure 2.S1). Because the disease is nonstationary, we used several methods of analysis to obtain robust results about the

characteristic temporal scales of the cycles. The cycling patterns of a time series, y_t , can be studied in the frequency domain and the time-frequency domain [18]. In the frequency domain, we used two main general approaches to determine the dominant frequencies in the data. The first one consisted of computing the periodogram, which gives the distribution of power (or, equivalently, variance) among different frequencies (see [19] and Appendix S1 for details). Thus, a peak in the periodogram indicates a dominant frequency. However, the periodogram assumes that the time series is stationary (i.e., with a constant mean and variance). Because our data are clearly nonstationary, we detrended the time series using the method known as discrete wavelet shrinkage (described in detail in Appendix S1).

The second method consisted of the maximum entropy spectral density $Y(vk)$ and was computed using the parameters (σ_w^2, ϕ_r) of a p th order autoregressive process fitted to the data (see [20] and Appendix S1 for details).

The above characterizations of the cycles consider the whole temporal extent of the data and therefore provide, as such, only an average picture of dominant frequencies in the data. More recently, the importance of localizing these frequencies in time has been emphasized, particularly for nonstationary data.

The wavelet power spectrum (WPSy) allows us to do so by calculating a measure of power as a function of both frequency and time. In other words, we can determine when in the temporal record a particular frequency is dominant and significant ([21,22]; see Appendix S1 for technical details).

Patterns of association between climate variables and CL.

Besides allowing us to determine how the variability of the data is allocated to different frequencies at different times, the wavelet transform can be used to study the patterns of association between two nonstationary time series [21]. Specifically, with the wavelet coherency analysis, we can determine whether the presence of a particular frequency at a given time in the disease corresponds to the presence of that same frequency at the same time in a climate covariate, and with the cross-wavelet phase analysis we can determine the time lag separating these two series as well.

Linear models and forecasts.

Seasonal autoregressive models were fitted to the data using the Kalman recursions for their state space representation ([23]; see Appendix S1 for technical details). To select the lags for the climate variables the following procedure was used [19,23,24]: (i) a null model was fitted to the square-root-transformed cases, (ii) MEI, SST 4, and temperature were filtered with the coefficients of the null model, and (iii) cross-correlation functions were computed using the residuals of the null model and those of the filtered climatic variables. On the basis of the cross-correlation functions, a full model was fitted that included as predictors all the statistically significant lagged climate variables [19]. This model was then simplified based on the following criteria: (i) the minimization of the Akaike information criterion and (ii) the absence of a significant difference ($p < 0.05$) when comparing the full model with the simpler version through a chi-squared likelihood ratio test [18,19,23].

Forecasts were obtained for time intervals 1, 3, 6, and 12 mo ahead using a total of 24 mo for each time interval. The model was refitted before computing the next prediction in two different ways, by including (i) all the previous months in the series or (ii) only the months from the previous 9 y.

The accuracy of the forecast was measured using the predictive R², which has an interpretation similar to the R² of a linear regression as defined in [25] and is given by $(1 - \text{mean squared error}/\text{variance of the series})$. Forecast accuracy was tested for the model selected as best, as well as for the simpler versions of this model, including a null model without climate covariates.

Results

CL shows, on average, a seasonal peak during May, though epidemic outbreaks happen around the year, as demonstrated by the existence of outliers in February, July, September, October, November, and December (Figure 2.1E). Interannual cycles with a period of approximately 3.2 y were identified by all the methods, with the exception of maximum entropy spectral density (with smoothing splines), which found cycles of 2.7 y (Figure 2.2). In addition, cycles of 8 y were found with the maximum entropy spectral density regardless of the method used to de-noise the data. However, evidence for the 8-y cycles is weak as they are longer than half the longitude of the series, a period above which the methods become unreliable [20]. The interannual cycles of 3.2 y coincide with those present in temperature and SST 4 for the same time period, suggesting a possible association between climate and transmission. In fact, cross-wavelet coherency analysis reveals that the square-root-transformed cases are

significantly ($p < 0.05$) coherent with temperature and SST 4 at periods around 3 y, with SST 4 and temperature leading the disease data, as shown by the cross-wavelet phases (Figure 2.3). The difference in phase is longer for SST 4 than for temperature. A similar pattern of coherence and phase is also seen for MEI, with MEI and cases coherent at periods of 3 y, and MEI leading the dynamics of the disease with a longer phase than that of temperature (Figure 2.S2). For both series, the cases are also significantly coherent at the seasonal scale, with only temperature leading cases, while SST 4 follows, possibly an artifact due to the seasonality of both series [21]. The results of cross-wavelet coherence and phase are consistent with those of the cross-correlation functions: the dynamics of the square-root-transformed CL cases are led by temperature with a 4-mo lag and by MEI with a 13-mo lag (Figure 2.4).

The forecasting accuracy is higher for the model selected by the likelihood ratio tests and Akaike information criterion values (Table 2.1). The best model describes the following process:

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) - \phi_1\phi_2(y_{t-3} - \mu) + \alpha_1 T_{t-4} + \gamma_a MEI_{t-13} + \varepsilon_t \quad (1)$$

where μ is the intercept, ϕ_j are the autoregressive terms, α_1 is a regression parameter for temperature (T), γ_a is a regression parameter for MEI, and the residuals ε_t are $N(0, \sigma_w^2)$ distributed. This model has a predictive accuracy of over 72%. The model with MEI as a predictor outperforms the model with just temperature and the null model with no climate covariates (Figure 2.4). In fact, the MEI model accounts for more than 65% of the variance in the data for prediction times up to 1 y (Figure 2.4). By contrast, the predictive R^2 for the

temperature model is negative for predictions 12 mo ahead, and the mean squared error is two orders of magnitude higher than the original variance of the series (Figure 2.S3). As expected, for all the evaluated models, forecasting accuracy diminishes with time and is slightly improved when all the previous months are employed for fitting the model (Figure 2.4).

Discussion

The description of cycles in any natural phenomenon is relevant to predict its dynamic behavior. The finding of a period close to 3 y for the interannual cycles of CL is robust, as the four applied methods indicate consistently the presence of cycles between 2.7 and 3.2 y in this series. The robustness of this result indicates that its statistical significance is not an artifact of any particular methodology. However, the description of cycles by itself is not sufficient to assess the effects of exogenous drivers or gain insights into the processes driving such oscillations.

A deeper understanding of the relationship between climate and disease dynamics is key for anticipating the potential effects that trends in a changing climate would have on the incidence and distribution of the disease [26]. The results of this study show that the dynamics of CL are strongly associated to those of climate variables, including temperature and ENSO indices, with coherent cycles of around 3 y. Similarly, associations with climate have been found for other vector-transmitted diseases, including dengue [27] and malaria [28–31].

A strong association between climate and CL incidence is further supported here by the finding that linear models can forecast satisfactorily the incidence of this disease, with an accuracy between 72% and 77%. In particular, MEI and temperature are identified as useful variables sustaining predictability for a window of 1 y. Interestingly, MEI is defined as the first principal component of several climate variables that predict ENSO [16]. This type of variable has been known to work well in linear regression because it reduces the number of predictors, avoiding problems of collinearity in the predictor matrix [32,33]. Longer-term data are needed to evaluate forecasting accuracy further in time. In the predictive model the nonstationarity of the CL time series is captured by the climatic covariates and the seasonal autoregressive part of the model. The finding that the model with MEI as the only predictor outperforms the model with just temperature supports the recent proposal that large-scale climate indices may be more useful for forecasting than local climate variables [32]. Climate can affect through several linear and nonlinear pathways the dynamics of infection in a host population. It can affect several biological traits of the organisms involved in the life cycle of the parasites, from individual life histories to population dynamics [34], and modify several factors that determine the context of disease transmission, including food production and the general standard of living of the population under the changing environment [35], both of which are known as important risk factors for other vector-borne diseases [36]. While local climate is more likely to affect only the biological components of

disease transmission, large-scale climate patterns could also influence contextual components of disease dynamics, such as population susceptibility. For CL there are several plausible ways in which climate could affect transmission dynamics. As already pointed out, vector density is sensitive to climate variability, with vector densities varying seasonally [7–10]. Parasite developmental time in vectors is also sensitive to environmental conditions, decreasing with high temperatures [37]. The density of reservoirs might also be sensitive to climate. For example, hantavirus outbreaks have been associated with changes in rodent reservoir densities, and high densities of rodents correlate with the altered production of seeds as the result of climatic conditions [38,39]. We can also expect there to be contextual effects of climate on transmission, such as those mediated by natural disasters, which could increase the risk of acquiring an infectious disease [35].

Future work should compare the forecasting ability of nonlinear models and more mechanistic formulations. While mechanistic models are necessary to propose and evaluate methods of control [3], there may be trade-offs between the complexity of these models and their ability to predict [3,40]. The results obtained here provide a basis for modeling other aspects of CL [3] and for producing forecasts for windows of time as long as 1 y ahead. The approach described in this paper could be applied to evaluating predictability in other vector-transmitted diseases.

Table 2.1. Model Selection and Parameter Values

$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_{12}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\gamma}_a$	$\hat{\gamma}_b$	$\hat{\sigma}_\varepsilon$	p_{χ^2}	d.f.	Akaike Information Criterion
13.32 ± 17.76	0.48 ± 0.08	0.50 ± 0.08	-0.55 ± 0.28	0.27 ± 0.27	0.37 ± 0.26	0.39 ± 0.22	-0.26 ± 0.36	1.89	—	—	480.17
-2.21 ± 16.07	0.47 ± 0.08	0.47 ± 0.08	—	0.36 ± 0.27	0.54 ± 0.25	0.30 ± 0.21	-0.42 ± 0.35	1.95	0.04	1	482.13
24.09 ± 14.37	0.50 ± 0.08	0.50 ± 0.08	-0.60 ± 0.27	—	0.32 ± 0.26	0.48 ± 0.21	-0.32 ± 0.37	1.90	0.31	1	479.19
29.59 ± 13.54	0.49 ± 0.08	0.51 ± 0.08	-0.67 ± 0.26	0.20 ± 0.27	—	0.42 ± 0.22	-0.34 ± 0.36	1.91	0.16	1	480.09
12.65 ± 19.09	0.50 ± 0.08	0.49 ± 0.08	-0.45 ± 0.28	0.42 ± 0.27	0.39 ± 0.27	—	-0.47 ± 0.37	1.94	0.07	1	481.37
5.56 ± 13.99	0.47 ± 0.08	0.49 ± 0.08	-0.59 ± 0.27	0.30 ± 0.27	0.40 ± 0.26	0.44 ± 0.21	—	1.90	0.46	1	478.70
-17.80 ± 9.59	0.45 ± 0.08	0.46 ± 0.08	—	0.43 ± 0.27	0.61 ± 0.24	0.35 ± 0.21	—	1.98	0.02	2	481.65
15.81 ± 10.59	0.48 ± 0.08	0.50 ± 0.08	-0.66 ± 0.26	—	0.34 ± 0.26	0.54 ± 0.19	—	1.93	0.18	2	477.96
21.01 ± 9.62	0.47 ± 0.08	0.50 ± 0.08	-0.75 ± 0.25	0.23 ± 0.26	—	0.48 ± 0.21	—	1.93	0.09	2	478.99
-3.59 ± 13.84	0.48 ± 0.09	0.48 ± 0.08	-0.50 ± 0.27	0.52 ± 0.26	0.46 ± 0.26	—	—	1.97	0.03	2	481.08
-5.96 ± 6.14	0.48 ± 0.08	0.47 ± 0.08	—	—	0.56 ± 0.25	0.49 ± 0.20	—	2.01	<0.01	3	482.15
27.28 ± 6.23	0.48 ± 0.08	0.50 ± 0.08	-0.78 ± 0.25	—	—	0.55 ± 0.20	—	1.93	0.06	3	477.71
13.95 ± 10.83	0.54 ± 0.08	0.48 ± 0.08	-0.60 ± 0.27	—	0.37 ± 0.26	—	—	2.02	<0.01	3	482.95
26.08 ± 6.44	0.54 ± 0.08	0.49 ± 0.08	-0.72 ± 0.26	—	—	—	—	2.05	<0.01	4	483.89
7.93 ± 0.43	0.48 ± 0.08	0.47 ± 0.08	—	—	—	0.49 ± 0.20	—	2.09	<0.01	4	485.21
8.13 ± 0.46	0.52 ± 0.08	0.46 ± 0.08	—	—	—	—	—	2.18	<0.01	5	488.85

The process for the full model is defined as: $y_t = \mu + \phi_1(y_{t-1} - \mu) + \phi_{12}(y_{t-12} - \mu) - \phi_1\phi_{12}(y_{t-13} - \mu) + \alpha_1 T_{t-4} + \gamma_a MEI_{t-13} + \varepsilon_t$

and the process for the null model as: $y_t = \mu + \phi_1(y_{t-1} - \mu) + \phi_{12}(y_{t-12} - \mu) - \phi_1\phi_{12}(y_{t-13} - \mu) + \varepsilon_t$

In both cases it is assumed that the error is independent and normally distributed:

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. The parameters for the full model are in the first data row, and for the null model are in the last data row. Parameters are described in the text and are given as value \pm standard error. p_{χ^2} is the significance of the chi-squared likelihood ratio test between each model and the full model, and d.f. its degrees of freedom.

Figure 2.1. Time Series (A) CL cases in Costa Rica. (B) Mean temperature in Costa Rica. (C) SST 4. (D) MEI. (E) Box plot with monthly square-root-transformed CL cases. (F) The fits of (1) the Daubechies discrete wavelet (green lines), used to detrend the series so that the resulting data can then be analyzed for their dominant frequencies with a periodogram (a filter number 5 and eight levels of decomposition were used for this wavelet; the dashed line corresponds to periodic edges, and the dotted line to symmetric ones); (2) smoothing splines (blue solid line) and the first four reconstructed components of singular spectrum analysis (black dashed line; 60 orders). These methods were used to de-noise the signals so that dominant frequencies could be identified (with the maximum entropy spectral density method).

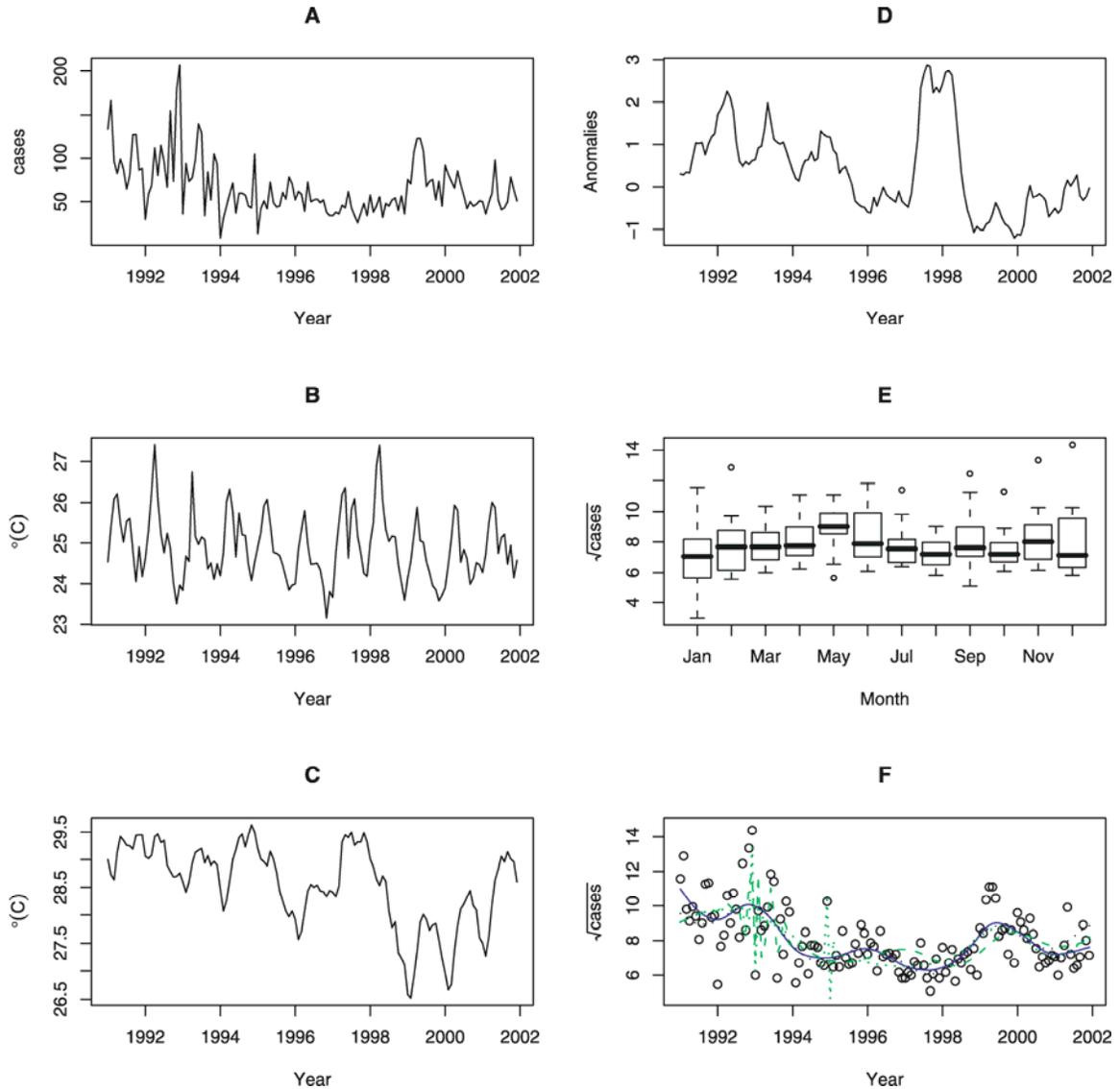


Figure 2.2. Dominant Frequencies in the Data(A and B) Smoothed periodograms for (A) the detrended series (using Daubechies discrete wavelet with filter number 5 and periodic edges) and (B) the detrended series (using Daubechies discrete wavelet with filter number 5 and symmetric edges). In the periodograms, the blue lines are the 95% point confidence intervals [19]. (C and D) Maximum entropy spectral density for (C) the de-noised series (with smoothing splines) and (D) the de-noised series (with singular spectrum analysis). For the periodograms and the maximum entropy spectral density, frequencies are in cycles per year. (E) Wavelet power spectrum. The solid line is the cone of influence indicating the region of time and frequency where the results are not influenced by the edges of the data and are therefore reliable. The dashed line corresponds to the 95% confidence interval for white noise based on the variance of the square-root-transformed incidence series. The intervals were obtained using a chi-squared distributed statistic with one degree of freedom (see [22] for details). The Morlet wavelet was used [18,21,22]. In all analyses, the cases are square-root-transformed. Maximum entropy spectral densities were computed using the software described in [20]. For the maximum entropy spectral density an autoregressive process of order $p = 40$ was used, i.e., AR(40).

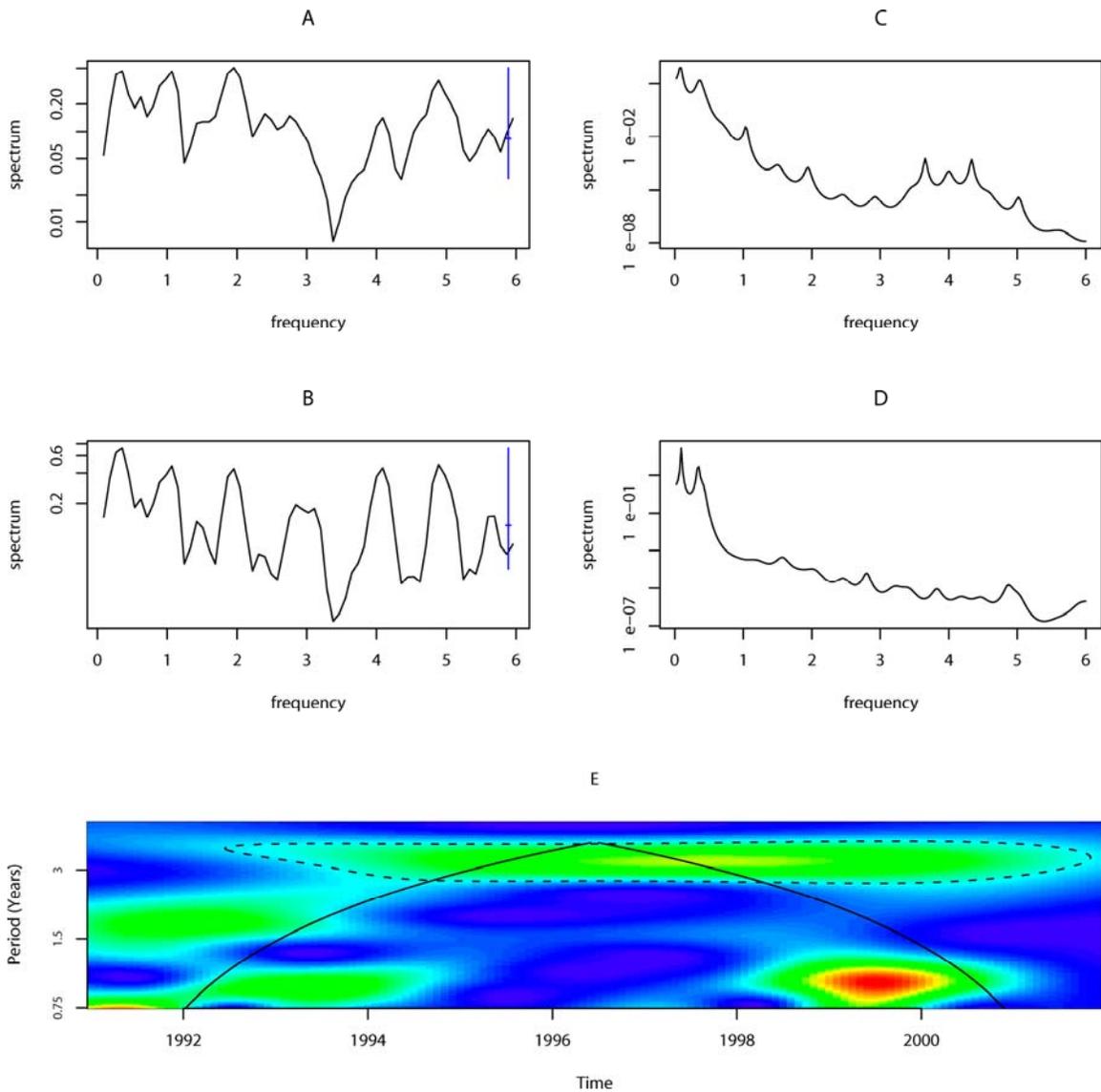


Figure 2.3. Cross-Wavelet Coherency and Phase The coherency scale is from zero (blue) to one (red). Thus, red regions indicate frequencies and times for which the two series share variability. The cone of influence (within which results are not influenced by the edges of the data) and the significant ($p < 0.05$) coherent time-frequency regions are indicated by solid lines. The colors in the phase plots correspond to different lags between the variability in the two series for a given time and frequency, measured in angles from $-\pi$ to π . A value of π corresponds to a lag of 16 mo. Cases are square-root-transformed. The procedures and software are those described in [21]. A smoothing window of 15 mo (2 wk + 1 d = 31 d) was used to compute the cross-wavelet coherence.

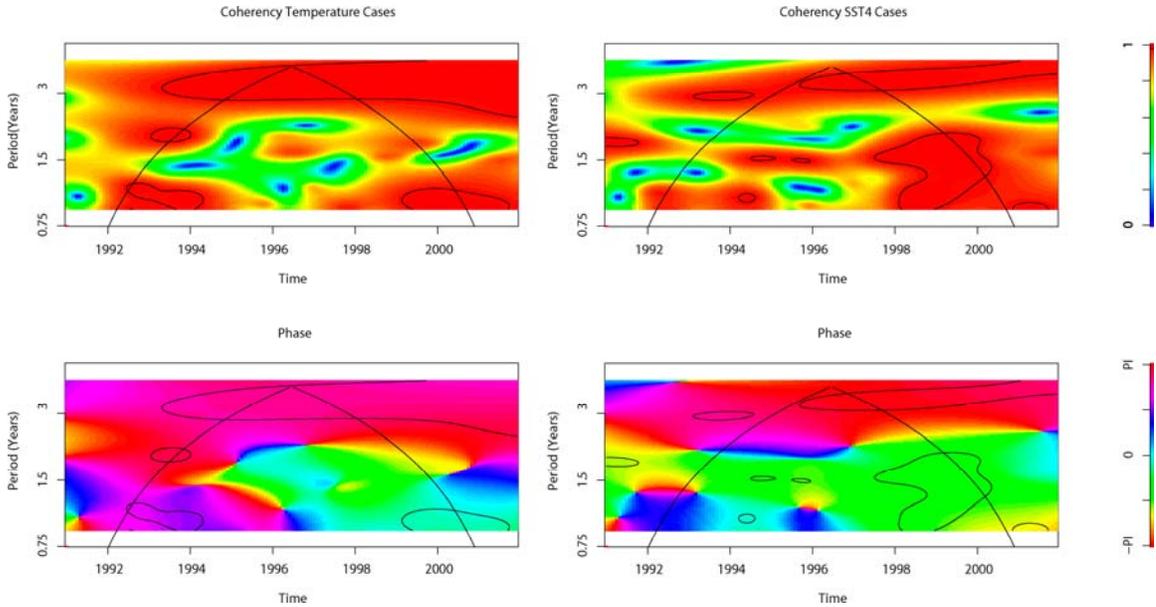


Figure 2.4. Cross-Correlation Functions of Square-Root-Transformed Cases with SST 4, Temperature, and MEI (A–C) Cross-correlation functions (CCF) with (A) SST 4, (B) temperature, and (C) MEI. The blue dashed lines are the 95% point confidence intervals for the cross-correlation between two series that are white noise [23]. (D) Predictive R^2 measuring the accuracy of the predictions. Blue is for predictions with only 9 y of training data (used to fit the model) and black for predictions generated with all months preceding the prediction. (The value for 12-mo predictions with temperature is not shown, because it was negative.)

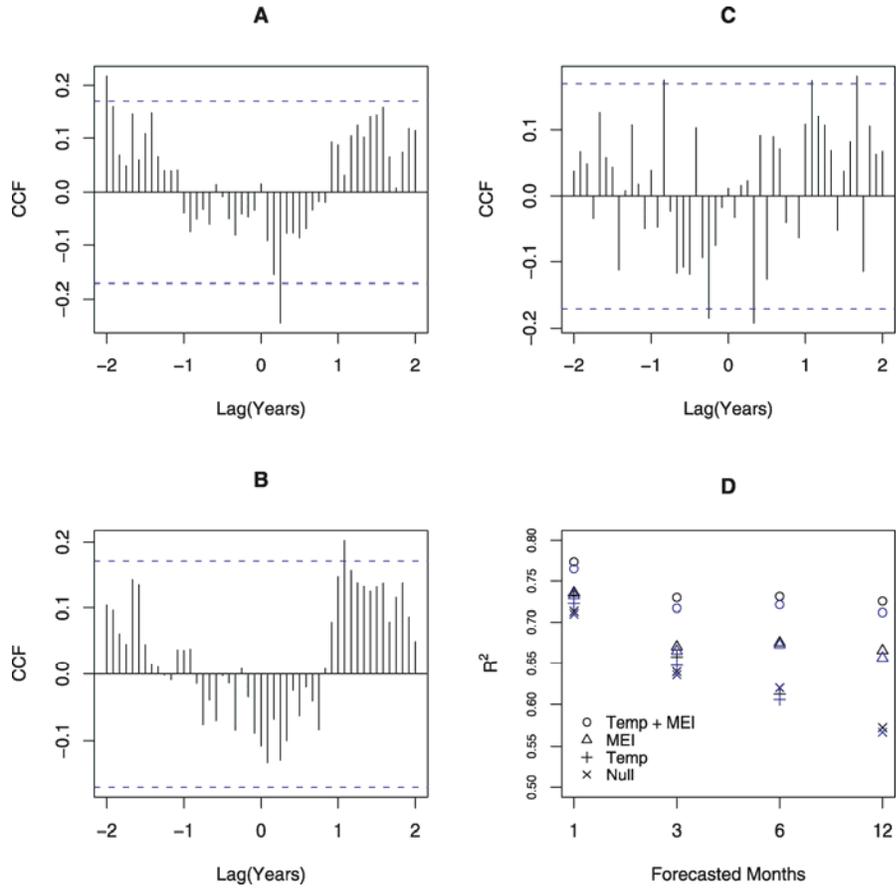


Figure 2.S1. The Autocorrelation Function of the Square-Root-Transformed CL Incidence Time Series. This series is nonstationary, because the autocorrelation function is statistically significant for lags different from zero and decays over time, a pattern that is superimposed on that resulting from seasonality, which produces a significant autocorrelation at a lag of 1 y.

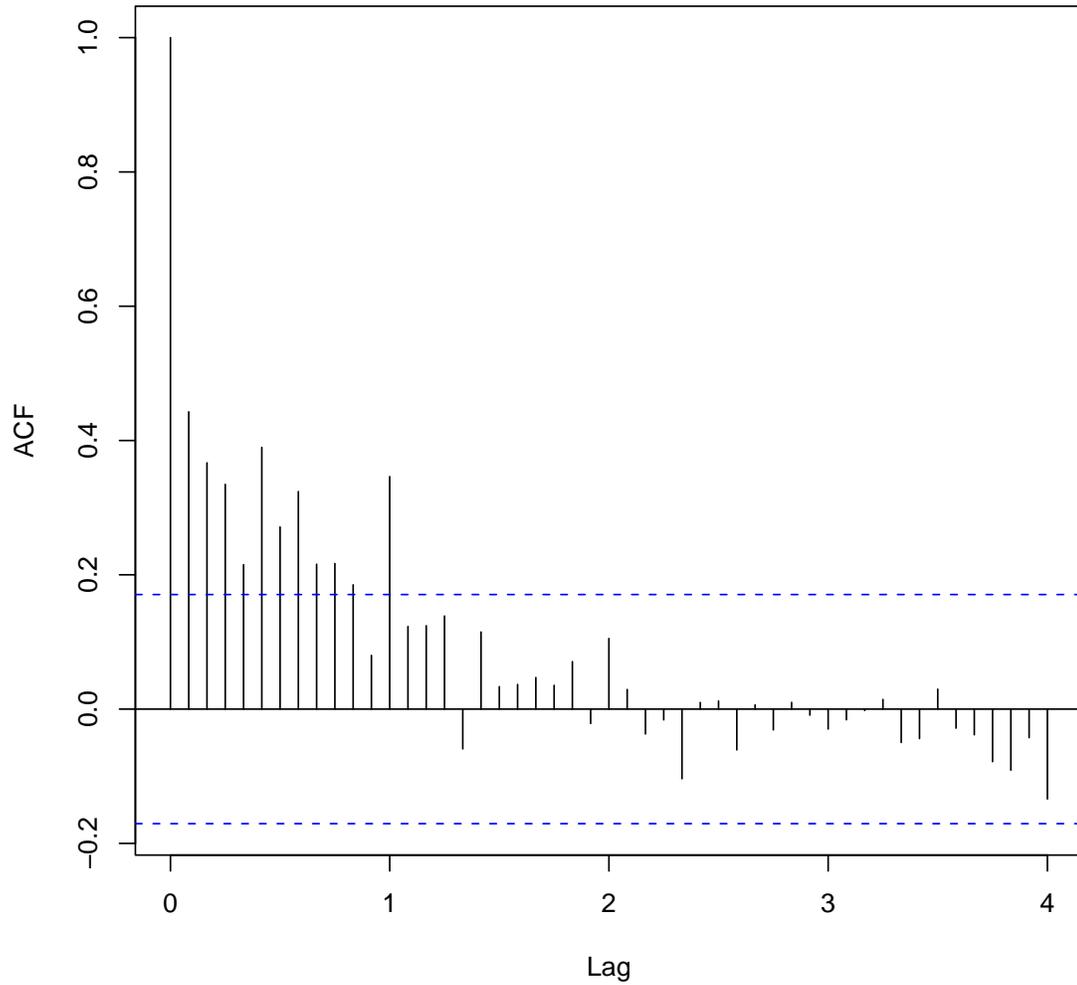


Figure 2.S2. Wavelet Coherence and Phase for the Square-Root-Transformed Incidence of CL and MEI For technical details see caption of Figure 3.

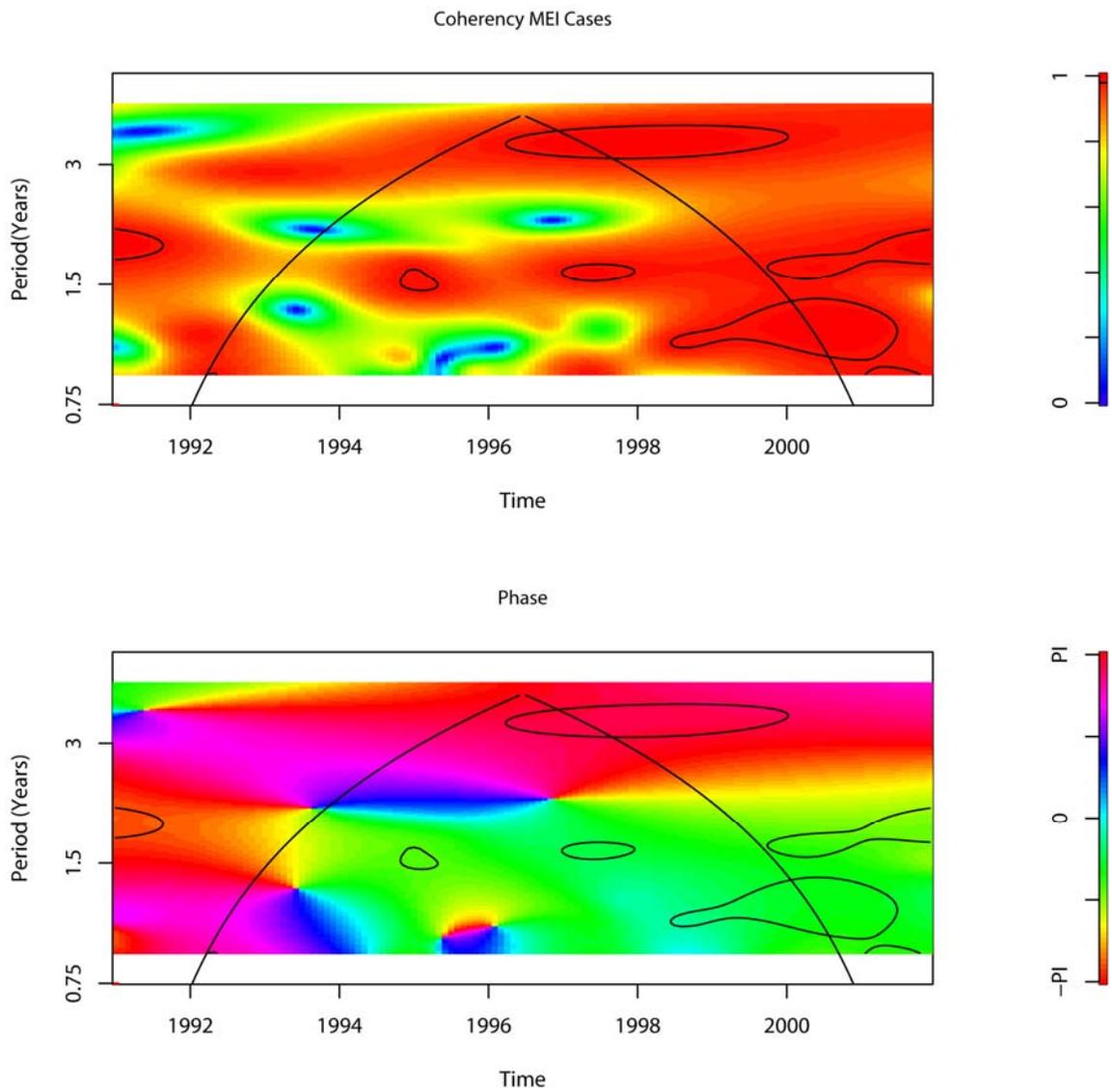
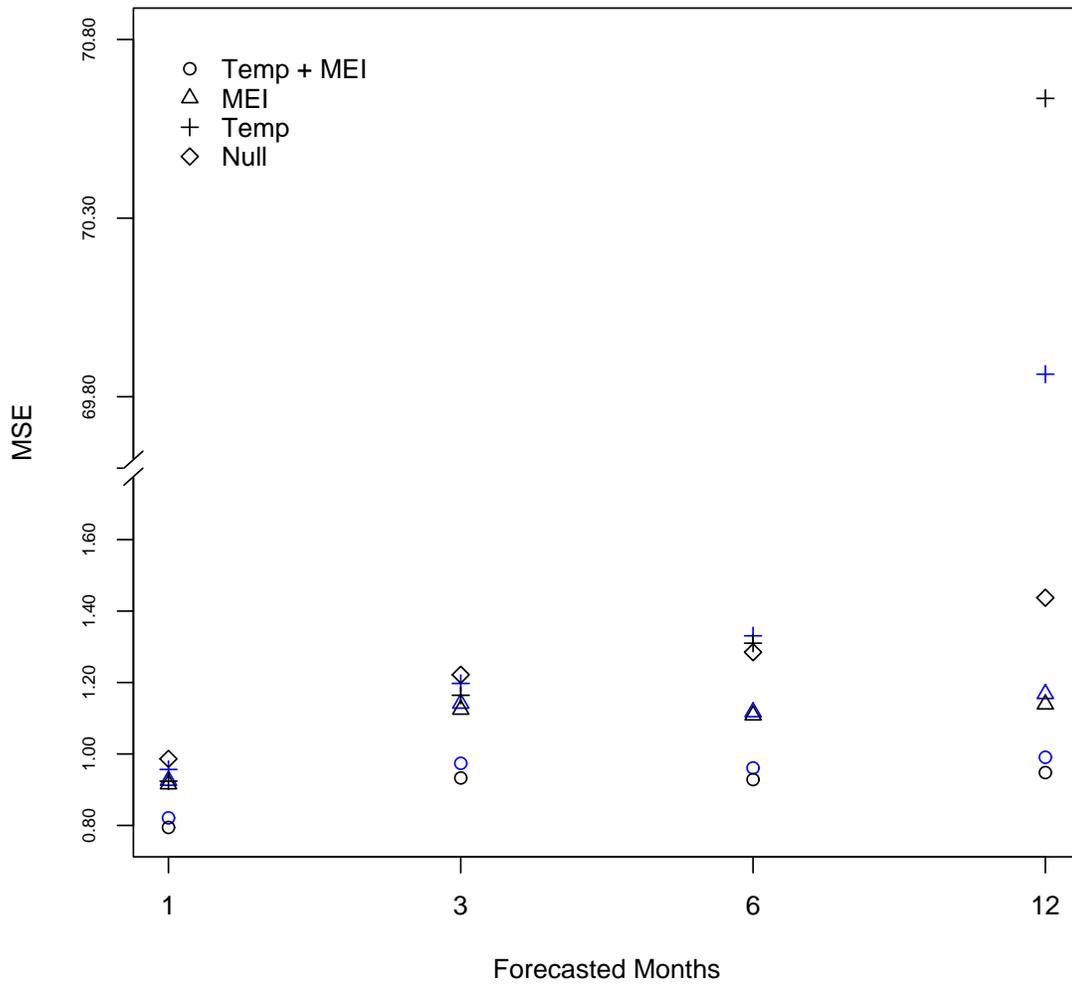


Figure 2.S3. Mean Squared Error of the Forecasts. Blue is for predictions with only 9 y of training data, and black, for all previous months preceding the prediction interval.



References

1. Gratz NG (1999) Emerging and resurging vector borne diseases. *Annu Rev Entomol* 44: 51–75.
2. Lainson R, Shaw JJ (1978) Epidemiology and ecology of leishmaniasis in Latin-America. *Nature* 273: 595–600.
3. Chaves LF, Hernandez MJ (2004) Mathematical modelling of American cutaneous leishmaniasis: Incidental hosts and threshold conditions for infection persistence. *Acta Trop* 92: 245–252.
4. Silveira FT, Lainson R, Corbett CE (2004) Clinical and immunopathological spectrum of American cutaneous leishmaniasis with special reference to the disease in Amazonian Brazil—A review. *Mem Inst Oswaldo Cruz* 99: 231–251.
5. Zeledon R (1991) Cutaneous leishmaniasis and *Leishmania infantum*. *Trans R Soc Trop Med Hyg* 85: 557.
6. Ashford RW (1997) The leishmaniasis as model zoonoses. *Ann Trop Med Parasitol* 91: 693–701.
7. Scorza JV, Gomez I, McLure MT, Ramirez M (1968) Sobre las condiciones microclimaticas prevalentes en los microhabitats de flebotomos. *Acta Biol Venez* 6: 1–27.
8. Marquez JC, Scorza JV (1984) Dinamica poblacional de *Lutzomyia townsendi* (Ortiz, 1959) (Diptera: Psychodidae) y su paridad en Trujillo, Venezuela. *Bol Dir Malariol San Amb* 24: 8–20.
9. Feliciangeli MD, Rabinovich J (1998) Abundance of *Lutzomyia ovallesi* but not *Lu. gomezi* (Diptera: Psychodidae) correlated with cutaneous leishmaniasis incidence in north-central Venezuela. *Med Vet Entomol* 12: 121–131.
10. Salomon OD, Wilson ML, Munstermann LE, Travi BL (2004) Spatial and temporal patterns of phlebotomine sand flies (Diptera: Psychodidae) in a cutaneous leishmaniasis focus in northern Argentina. *J Med Entomol* 41: 33–39.
11. Rabinovich JE, Feliciangeli MD (2004) Parameters of *Leishmania braziliensis* transmission by indoor *Lutzomyia ovallesi* in Venezuela. *Am J Trop Med Hyg* 70: 373–382.

12. Marrano NN, Mata LJ, Durack DT (1989) Cutaneous leishmaniasis in rural Costa Rica. *Trans R Soc Trop Med Hyg* 83: 340.
13. Añez N, Nieves E, Cazorla D, Oviedo M, De Yarbuh AL, et al. (1994) Epidemiology of cutaneous leishmaniasis in Merida, Venezuela. III. Altitudinal distribution, age structure, natural infection and feeding behavior of sandflies and their relation to the risk of transmission. *Ann Trop Med Parasitol* 88: 279–287.
14. Franke CR, Ziller M, Staubach C, Latif M (2002) Impact of the El Niño/Southern Oscillation on visceral leishmaniasis, Brazil. *Emerg Infect Dis* 8: 914–917.
15. New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. *Clim Res* 21: 1–25.
16. Wolter K, Timlin MS (1998) Measuring the strength of ENSO—How does 1997/98 rank? *Weather* 53: 315–324.
17. Venables WN, Ripley BD (2002) *Modern applied statistics with S*. New York: Springer. 495 p.
18. Gençay R, Selçuk F, Whitcher B (2002) *An introduction to wavelets and other filtering methods in finance and economics*. San Diego: Academic Press. 359 p.
19. Shumway RH, Stoffer DS (2000) *Time series analysis and its applications*. New York: Springer. 572 p.
20. Ghil M, Allen RM, Dettinger DM, Ide K, Kondrashov D, et al. (2002) Advanced spectral methods for climatic time series. *Rev Geophys* 40: 1–41.
21. Maraun D, Kurths J (2004) Cross wavelet analysis: Significance testing and pitfalls. *Nonlinear Proc Geophys* 11: 505–514.
22. Torrence C, Compo G (1998) A practical guide to wavelet analysis. *Bull Am Meteor Soc* 79: 61–78.
23. Brockwell PJ, Davis RA (2002) *Introduction to time series and forecasting*, 2nd ed. New York: Springer. 434 p.
24. Durbin J, Koopman SJ (2001) *Time series analysis by state space methods*. Oxford: Oxford University Press. 253 p.

25. Ellner SP, Bailey BA, Bobashev GV, Gallant AR, Grenfell BT, et al. (1998) Noise and nonlinearity in measles epidemics: Combining mechanistic and statistical approaches to population modeling. *Am Nat* 151: 425–440.
26. Patz JA, Hulme M, Rosenzweig C, Mitchell TD, Goldberg RA, et al. (2002) Regional warming and malaria resurgence. *Nature* 420: 627–628.
27. Cazelles B, Chavez M, McMichael AJ, Hales S (2005) Nonstationary influence of El Niño on the synchronous dengue epidemics in Thailand. *PLoS Med* 2: e106.
28. Abeku TA, De Vlas SJ, Boorsboom GJJM, Tadege A, Gebreyesus Y, et al. (2004) Effects of meteorological factors on epidemic malaria in Ethiopia: A statistical modelling approach based on theoretical reasoning. *Parasitology* 128: 585–593.
29. Teklehaimanot HD, Schwartz J, Teklehaimanot A, Lipsitch M (2004) Weather-based prediction of *Plasmodium falciparum* malaria in epidemic-prone regions of Ethiopia II. Weather based prediction systems perform comparably to early detection systems in identifying times for interventions. *Malar J* 3: 44.
30. Zhou G, Minakawa N, Githeko AK, Yan G (2004) Association between climate variability and malaria epidemics in the east African highlands. *Proc Natl Acad Sci U S A* 101: 2375–2380.
31. Thomson MC, Mason SJ, Phindela T, Connor SJ (2005) Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana. *Am J Trop Med Hyg* 73: 214–221.
32. Stenseth NC, Ottersen G, Hurrell JW, Mysterud A, Lima M, et al. (2003) Studying climate effects on ecology through the use of climate indices: The North Atlantic Oscillation, El Niño Southern Oscillation and beyond. *Proc Biol Sci* 270: 2087–2096.
33. Faraway JJ (2005) *Linear models with R*. Boca Raton: Chapman and Hall/CRC. 229 p.
34. Hallet TB, Coulson T, Pilkington JG, Clutton-Brock TH, Pemberton JM, et al. (2004) Why large-scale climate indices seem to predict ecological processes better than local weather. *Nature* 430: 71–75.
35. Collier M, Webb RH (2002) *Floods, droughts and climate change*. Tucson: University of Arizona Press. 153 p.

36. Bouma MJ (2003) Methodological problems and amendments to demonstrate effects of temperature on the epidemiology of malaria. A new perspective on the highland epidemics in Madagascar, 1972–1989. *Trans R Soc Trop Med Hyg* 97: 133–139.
37. Leaney AJ (1977) Effect of temperature on *Leishmania* in sand flies. *Parasitology* 75: R28–R29.
38. Levins R, Epstein PR, Wilson ME, Morse SS, Slooff R, et al. (1993) Hantavirus disease emerging. *Lancet* 342: 1292.
39. Stone R (1993) The mouse-piñon connection. *Science* 262: 833.
40. Levins R (1966) The strategy of model building in population biology. *Am Sci* 54: 421–431.

CHAPTER III

CLIMATE CHANGE AND THE ABILITY TO FORECAST DISEASES:II. NON-LINEAR TOOLS APPLIED TO AMERICAN CUTANEOUS LEISHMANIASIS FORECASTS IN COSTA RICA

Introduction

One of the best documented patterns in the dynamics of vector transmitted diseases is their periodicity at seasonal and interannual temporal scales [1-7]. These periodicities are the basis for the proposal that Early Warning Systems (EWS) are feasible and useful tools for planning and decision making [2]. EWS are alert systems whose objective is to predict either epidemic outbreaks in regions where disease transmission is unstable or large outbreaks where the disease is endemic. From the early 1910s, when Captain S.R. Christophers of the British army developed a system to predict malaria in India using climatic and socioeconomic data [8,9], to present times when systems are based on indoor resting densities of vectors [10], climate, land use, and satellite imagery [11], EWS have been regarded as useful tools to help the development of poor and disease stricken nations [2,11]. The early experience by Christophers was highly successful, and his system was in use until the 1940's, when the importance of malaria as a public health issue in the Indian subcontinent

diminished [9,11]. However, recent results have demonstrated that the blind use of EWS can lead to unreliable forecasts especially when models are used in regions where the connection between climate and disease is not well understood [12].

Despite the possible caveats of climate-based EWS, especially because of the complexity of human diseases for which social components can be as important as natural forces [13-15], there are successful examples of prediction of “out of fit” data based on the known association between climate and disease [6]. Although most of the effort in developing EWS has been focused on malaria [1,2,16], similar efforts would be valuable for neglected diseases that represent a large burden for developing countries and whose transmission is sensitive to climate variability [6,17]. The leishmaniasis in particular represent the fourth most important neglected disease, with a burden of at least 2.1 million of infected people per year, second to malaria in terms of the number of people affected by a protozoan vector transmitted disease [17,18]. Like many other diseases the infections are caused by protozoa, belonging to any of several different species of *Leishmania* spp (Kinetoplastida: Trypanosomatidae), transmitted by sand flies (Diptera: Psychodidae). The clinical manifestation encompasses visceral and cutaneous/mucocutaneous cases, and is associated with a certain parasite species [6]. Our previous results indicate that American cutaneous leishmaniasis (ACL) is a good candidate for the use of climate-based EWS, because predictions with seasonal autoregressive (SAR) models can have an accuracy of over 75 % [6]. In the present paper, our objective is to illustrate a protocol for the

development of EWS, including the evaluation of different linear and non-linear techniques for time series modeling and prediction, as well as the assessment of the robustness of the relationship between the disease and climate that is the basis for building EWS.

Methods

Data

Leishmaniasis Monthly records of ACL cases from January 1991 to December 2001, were obtained from the epidemic surveillance service Vigilancia de la Salud, of Costa Rica. The data were normalized using a square root transformation.

Climatic Covariates The temperature (T) data are those used in [6] consisting of the average temperature in the $0.5^\circ \times 0.5^\circ$ grids composing the Costa Rica land surface [<http://www.cru.uea.ac.uk>,19]. The monthly average of these temperature records, T , and the multivariate ENSO index, MEI , [<http://www.cdc.noaa.gov/people/klaus.wolter/MEI>, 20] were used as predictors for modeling the transformed ACL cases. For all the models below, except for the non-linear forecasting (NLF) and the basic structural model (BSM), the lags for the introduction of climate covariates T and MEI were chosen based on our previous results using cross-correlation functions [6], with a fixed delay (i.e., months preceding the cases series) of 13 months for MEI and 4, 8, and 20 months for T . All time series are shown in Figure 1. Other climatic covariates, precipitation and relative humidity, were ignored since they did not show a strong

association with the case data using non-stationary tools like wavelet cross-coherence [6].

Statistical Analyses

Forecasting models: Several linear and non-linear models were fitted to the square root transformed case data. Brief descriptions follow of: (1) the approach to handle seasonality, (2) the types of models used, and (3) their classification as linear or non-linear.

Seasonality To introduce seasonality, the strategy for all models was to include lags 12 and 13 of the transformed case data. This approach was chosen because the autoregressive treatment of seasonality is known to be the best approximation to the asymptotic cyclical structure of a time series [21]. This approach specifically allows a better minimization of the error variance when compared to a fixed seasonality implemented with a standard cyclical function (such as sines or cosines) that leads to a symmetrical cyclical structure [21].

Linear. In this class of models, parameters have a linear relationship with the response variable [22], in this case the transformed number of cases. This definition should not be confused with that of a linear dynamical system where the relationship of the dependent variables or covariates is linear with that of the independent variable [23]. In fact linear models can be used to fit the parameters of non-linear dynamical systems provided that the relationship between a response (which can be a transformation of the independent variable in the non-linear dynamical system) and the covariates (which also can be transformed) is

linked by a parameter linearly. Linear models used in this paper include: SAR and BSM.

Non-Linear. In these models the relationship between the response and the parameters for the predictors is not constrained to be linear. Models include: NLF, generalized additive models (GAM) and feed-forward neural networks (FNN). A description of the methods (linear and non-linear) and of the fitted models can be found in the Appendix S2.

Forecasts For all models, forecasts were obtained for prediction time intervals of 1, 3, 6, and 12 months ahead for a total of 24 months each. Each model was refitted recurrently before computing the next prediction by including all the previous months in the series [6]. The accuracy of the forecast was measured using the predictive R^2 , which has an interpretation similar to the R^2 of a linear regression by definition [23] and it is obtained as: $R^2 = 1 - (\text{mean square error}/\text{variance of the series})$. Thus, the errors are normalized by the variance of the time series; an R^2 of 1 indicates perfect forecasts while a value close to 0 or negative indicates poor predictability. Forecasting accuracy was tested for all the fitted models. To establish a baseline for comparison, the predictive R^2 was also computed when the prediction is the monthly mean value of the transformed time series.

Robustness of the exogenous forcing by climate

Once the best modeling approach was selected, the robustness of the association between the cases and the exogenous forces T and MEI was assessed with a non-parametric bootstrap approach based on 10000

randomisations. The idea of the non-parametric bootstrap is to reconstruct an experimental dataset based on the fitted values of a model plus the residuals sampled with replacement from such a model [24]. To generate the bootstrap samples, the model with the highest predictive R^2 was used. The bootstrap was initially used to see the frequency (%) of times the model from which we generated the bootstrap samples was actually selected as the best model, using the Akaike Information criterion [25,26]. Then, using the sub-sample of models selected as best that also have the highest probabilities in the above bootstrap test, we constructed confidence intervals for the parameters. We further refitted the model without the last 24 points to make forecasts and get their confidence intervals.

Results

Figure 2 shows the square root transformed cases plotted against their lagged values (1, 12 and 13 months) and the lagged covariates T (4 months) and MEI (13 months). In all cases, no obvious non-linearity is apparent in the relationship among the three variables. As expected, all models but FNN were most successful for predictions of 1 month ahead. However, for prediction steps larger than one month only NLF, SAR and GAM models with environmental covariates, MEI and T (4 months lag), did better than predictions based on the average of the time series (Table 1). The models with the worst performance were FNNs, followed by BSM and the null SAR (i.e., without covariates). For NLF, the best results were found with $E=2$ and $E=3$, with the latter embedding dimension providing slightly better results for a 12 months ahead prediction.

The predictive R^2 was highest for the SAR model with T (4 months lag) and MEI (13 months lag) as covariates. Thus, the fitted values and residuals used for the bootstrap were those of the model in the first equation of (1) in the Appendix S2. The bootstrap results show that the best model is the one used to generate the data (for 67.40% of the simulated time series, the model was selected as best). The confidence intervals for this model show that the parameters for T and MEI are statistically significant, a result that holds even if the intervals are constructed using the values for this parameter when the model was not selected as best (Figure 3A). The autoregressive terms, however, are not significant as the confidence intervals include 0. The variance of the residuals obtained from the real data is significantly shorter than the one from the simulations, probably because of the destruction of the autoregressive structure by the re-sampling of residuals [25]. Finally, the results also show (Figure 3B) that the maximum forecasting ability for these models is 80%, and can be as low as 50% probably because of the sensitivity of the models to a lack of a well defined SAR structure.

Discussion

The need for forecasts by policy makers goes well beyond the development of EWS for diseases. In a world where large scale changes are happening at a rapid pace, from increased average temperatures to extensive land use changes, major alterations in biogeochemical cycles, water availability, food production, biodiversity and diseases are already occurring and likely to be

exacerbated in the future [27,28]. Although the imperative need for predictions that can inform policy has been repeatedly emphasized [11,28], the common practice regarding diseases is to evaluate models by their ability to fit the data [29-35] and only in very few instances, tests have been conducted based on data that have not been used to fit the models [6]. Consideration of “out-of-fit” data is critical if we are to evaluate the ability of the models to predict the future.

In this paper, we have presented several methods to study seasonal time series, and used a simple measure, the predictive R^2 , to compare models based on their ability to predict future dynamics and not their goodness of fit of the past. By comparison with modeling results for other infectious diseases on the predictability of NLF methods [36], our results demonstrate a very high predictability for ACL. An important element that might explain this difference is the association of this disease to climate, since models that incorporated climatic covariates performed generally better than those that only considered previous disease levels. Another explanation might be the robustness of the association between the disease and climatic covariates as demonstrated by the bootstrap results. While the parameters for the covariates are statistically significant, the autoregressive parameters are not consistently so, and the variance of the residuals significantly increases.

One of the main lessons from the study of populations is that non-linear dynamics are common in nature but often satisfactorily captured by linear approximations [37,38]. This has been demonstrated by the analysis of time series from a wide variety of animals and diseases. While chaos is present in a

small sample of the populations considered periodicities are common, particularly in infectious diseases, that can be explained by either the effect of exogenous forces, like climate, or endogenous ones, like recruitment of new individuals and the concurrent changes in densities [39-42]. Our results indicate that ACL is another example of a population phenomenon whose dynamics can be satisfactorily described by linear statistical models, provided that appropriate covariates and transformations of the data are used. Thus, though linear models do best, functional forms underlying the influence of covariates are likely to be non-linear as indicated by the transformations used. This result is further supported by the observation that the predictive R^2 for NLF with E=3 does not vary with the prediction time step, while this value for the SAR model without covariates decreases abruptly, as expected in systems where the dynamics are non-linear [36-43]. Linear models were also used successfully for other vector-borne diseases, Malaria [43] and Ross river virus [4,35], and for cutaneous leishmaniasis in other regions of the new world [45]. For ACL, the usefulness of linear models (after appropriate transformation) might also follow from the fact that humans are only sinks for the pathogen and therefore, provide no feedback to transmission [46,47]. This conjecture would not necessarily apply to other vector-transmitted diseases where infected humans provide sources of new infections within the population.

This result also highlights two open questions that need to be addressed when modeling infectious diseases transmitted by vectors: first, the appropriate functional form to introduce climate variables into the dynamics [46, 48]; second,

the best approach for modeling seasonality [8, 49] . Mathematically the relationship between climatic co-variates and the numbers of the disease can be non-linear, described by simple non-linear functions, like those of the functional responses in consumer-resource interactions [50] or modeled by linear models with self-excited thresholds [51]. This is especially relevant, since a saturating non-linear functional form can lead to very different scenarios in the dynamics of the disease under altered environmental conditions. In the case of ACL, however, no apparent need for non-linear functions describing the relationship to climate was evident. In general seasonality has been modeled using fixed structures, i.e., values are assumed to be constant [e.g., 9,49] or approximated by the sum of sine and cosine functions [e.g., 41,52]. The introduction of SAR seasonality in mechanistic models should be further investigated.

A factor that deserves further consideration in developing EWS, is the understanding of the role of space. Predictability at more local scales was not addressed here because half of the series was only available at levels below that of the whole country, and because Costa Rica encompasses a small area for which temperature variability is quite homogenous, as seen in the very small variability between temperature grids. However, for larger spatial scales heterogeneities in the landscape for disease transmission would need to be considered [53].

EWS are a feasible ecological application for neglected diseases, as illustrated for ACL. Available models have good levels of predictability up to one year ahead for the number of cases. Predictability strongly depends on the use of

an appropriate structure for the different components of the model, including seasonality and exogenous drivers such as climatic variables. Depending on the model, predictability can range from poor, with approximately 50% accuracy, to high, with 80% accuracy, significantly better than that of seasonal averages (about 65%). Forecasts can be useful in planning services for the populations affected, allowing estimates of approximate number of hospital beds, vaccine shots, drug doses and vector control measures. If EWS need to incorporate the spatial spread of the disease, they should do so dynamically and in relation to different landscapes, such as the geopolitical unit of this study or regions with similar climatic patterns [53], otherwise predictions are likely to fail as illustrated by [12]. While there is no unique early warning system for a given disease, there should be a general approach for the development of EWS. Our work illustrates three key components of such an approach for vector-borne diseases: (i) the evaluation of predictability with “out-of-fit” data and not simply goodness of fit [6,40,41]; (ii) the comparison of a suite of possible models in terms of predictability [55,56] and (iii) the robustness of the relationship with covariates in the selected model. Here, robustness is used following [55], to identify covariates that are useful to predict disease numbers even when the skeleton of the model changes. Finally, none of these efforts are possible without the invaluable role of sustained surveillance and monitoring efforts. A historical retrospective reinforces this point: the success of Christophers was possible because of data availability and a deep knowledge of malaria biology; from parasites to mosquitoes and humans, realizing the influence of factors as diverse as weather and wheat

prices in rendering the epidemics of malaria predictable [8]. Time series sufficiently long for developing and evaluating forecasting models around the world are countable; their number pales by comparison to the data available for weather forecasting. It is imperative that on-going efforts are sustained and new ones are initiated whose long-term planning includes EWS as a specific goal.

Table 3.1 Models and predictive R^2 .

Model	1 month	3 months	6 months	12 months
NLF (E=2)	0.69	0.62	0.61	0.66
NLF (E=3)	0.67	0.60	0.59	0.67
NLF (E=4)	0.66	0.59	0.58	0.66
FNN (2 Layers)	0.55	0.53	0.44	0.44
FNN (3 Layers)	0.62	0.58	0.61	0.60
SAR (null)	0.71	0.64	0.62	0.57
SAR (MEI)	0.73	0.67	0.67	0.66
SAR (MEI + T)	0.77	0.73	0.73	0.72
BSM	0.69	0.59	0.52	0.65
GAM (MEI)	0.66	0.59	0.56	0.57
GAM (MEI + T)	0.73	0.68	0.67	0.68
MEAN	0.64	0.64	0.64	0.64

For model identification see common abbreviations. Months indicate the number of months predicted ahead. Mean indicates the results that could be obtained by just using the monthly average number of cases.

Figure 3.1. Time Series: **A** Square root Transformed ACL Cases in Costa Rica **B** Mean Temperature in Costa Rica **C** MEI

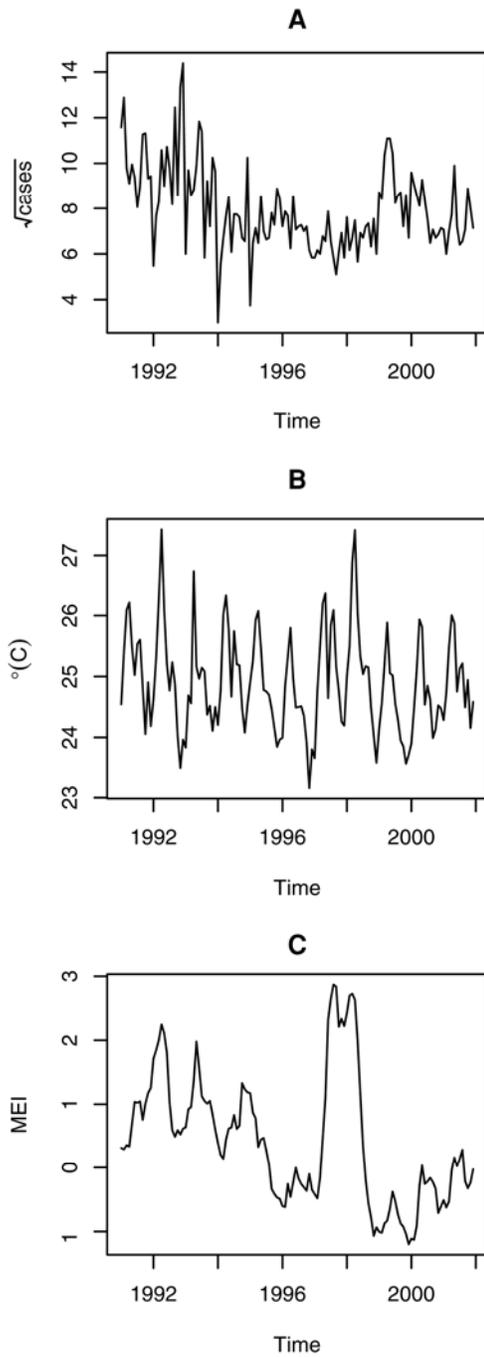


Figure 3.2. Multidimensional plots for the square root transformed ACL cases (y_t) as function of: **A** Autoregressive (y_{t-1}) and Seasonal (y_{t-12}) components **B** Seasonal (y_{t-12}) and Autoregressive Seasonal (y_{t-13}) components **C** Autoregressive component (y_{t-1}) and Temperature (lag 4, T_{t-4}) **D** Autoregressive component (y_{t-1}) and MEI (lag 13, MEI_{t-13}).

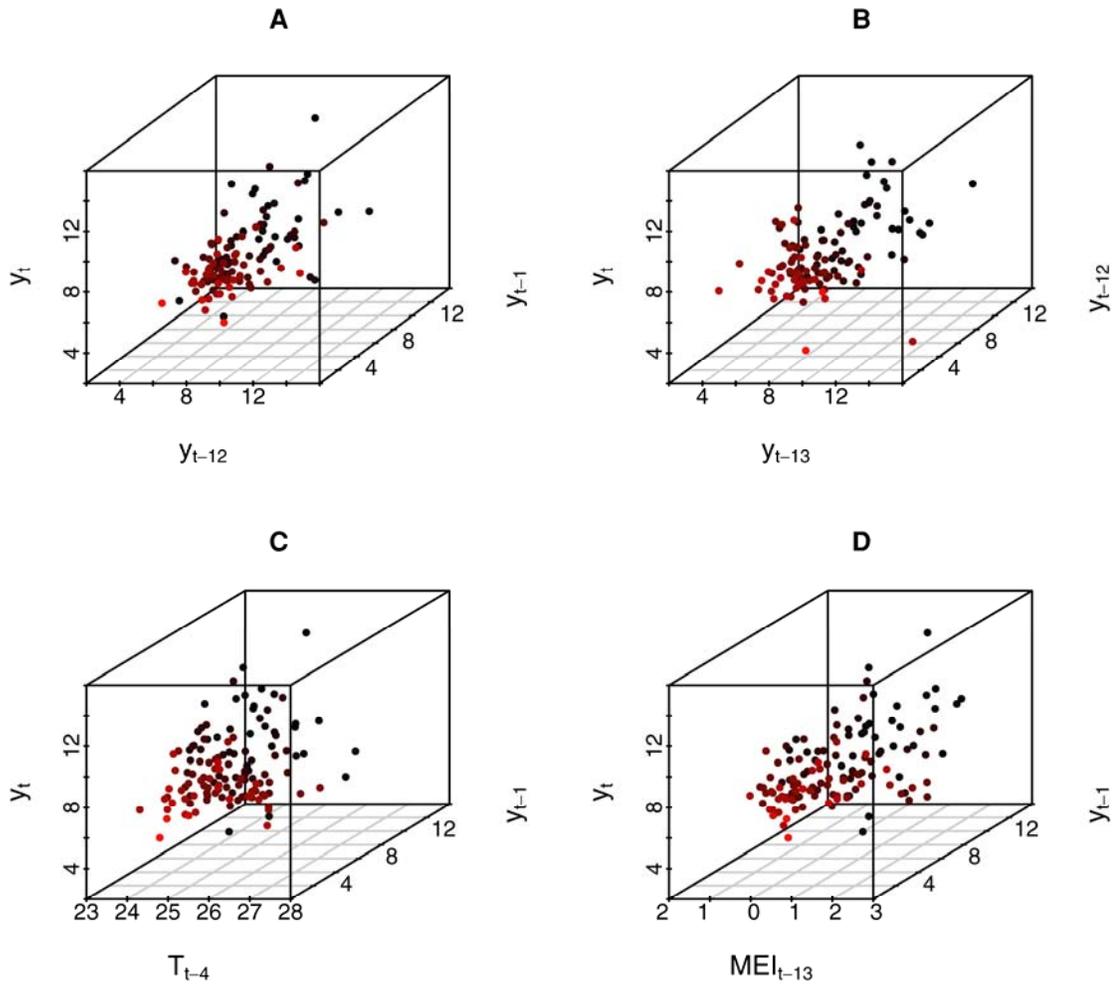
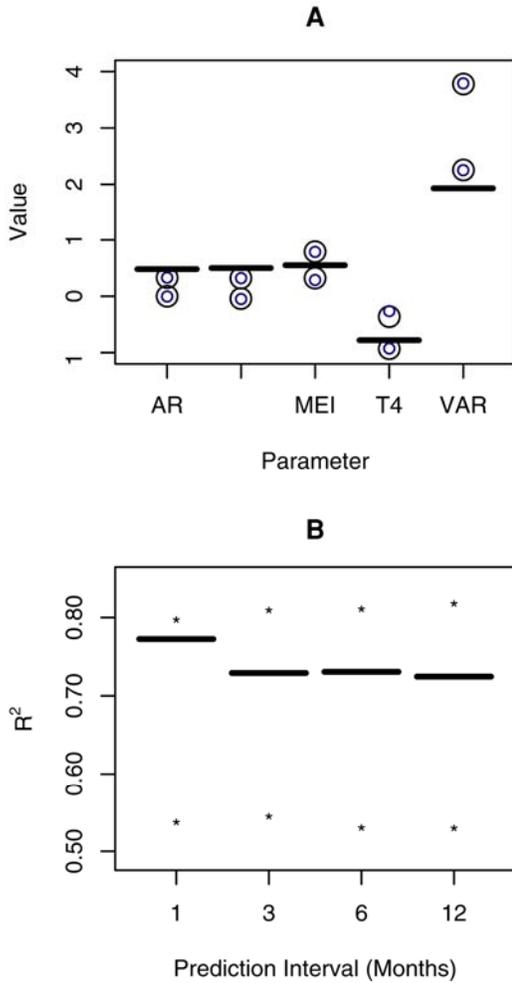


Figure 3.3. Bootstrap Experiment: **A** 95% Confidence intervals for the parameters of the best model. AR stands for the autoregressive component of the model (ϕ_1); ARseas for the seasonal autoregressive component of the model (ϕ_{12}); VAR for the variance of the residuals (σ_ε^2); MEI and T4 for the parameter for MEI at lag 13 (α) and Temperature at lag 4 (γ) respectively. Black signs are 95% CI's using values from the sub-sample when the model is selected as best, and blue including all the bootstrap samples. The structure of the best model can be seen in the Appendix S2 **B** Predictive R^2 and the 95% confidence intervals, indicated by stars, for the bootstrapped best model and prediction interval.



References

1. Connor SJ, Thomson MC, Molyneux DH (1999) Forecasting and prevention of epidemic malaria: new perspectives on an old problem. *Parassitologia* 41: 439-448.
2. Thomson MC, Connor SJ (2001) The development of malaria early warning systems for Africa. *Trends Parasitol* 17: 438-445.
3. Hay SI, Myers MF, Burke DS, Vaughn D, Endy T, et al. (2000) Etiology of interepidemic periods of mosquito-borne disease. *Proc Natl Acad Sci USA*, 97: 9335-9339.
4. Tong SL, Hu WB, McMichael AJ (2004) Climate variability and Ross River Virus transmission in Townsville region, Australia, 1985-1996. *Trop Med Int Health* 9: 298-304.
5. Patz JA, Campbell-Lendrum D, Holloway T, Foley JA (2005) Impact of regional climate change on human health. *Nature* 438: 310-317.
6. Chaves LF, Pascual M (2006) Climate cycles and forecasts of cutaneous leishmaniasis, a nonstationary vector-borne disease. *PLoS Med* 3: e295.
7. Woodruff RE, Guest CS, Garner MG, Becker N, Lindsay M (2006) Early warning of Ross virus epidemics: combining surveillance data on climate and mosquitoes. *Epid* 17: 569-575.
8. Gill CA (1928) *The genesis of epidemics and the natural history of disease. An introduction to the science of epidemiology based upon the study of epidemics of malaria, influenza and plague.* London: Bailliere, Tindall and Cox.
9. Gill CA (1938) *The seasonal periodicity of malaria and the mechanism of the epidemic wave.* London: J. & A. Churchill Ltd
10. Lindblade KA, Walker ED, Wilson ML (2000) Early warning of malaria epidemics in African highlands using *Anopheles* (Diptera: Culicidae) indoor resting density. *J Med Entomol* 37: 664-674.
11. Rogers DJ, Randolph SE, Snow RW, Hay SI (2002) Satellite imagery in the study and forecast of malaria. *Nature* 415: 710-715.
12. Hay SI, Were EC, Renshaw M, Noor AM, Ochola SA, et al. (2003) Forecasting, warning, and detection of malaria epidemics: a case study. *Lancet* 361: 1705-1706.

13. Wilson ML (1994) Developing paradigms to anticipate emerging diseases – transmission cycles and a search for pattern. *Ann N Y Acad Sci* 740: 418-422.
14. Levins R, Lopez C (1999) Toward an ecosocial view of health. *Int J Health Serv* 29: 261-293.
15. Chaves LF (2007) *Casas Muertas and Oficina No.1* : internal migrations and malaria trends in Venezuela 1905-1945. *Parasitol. Res.* 101:19-23.
16. Thomson MC, Mason SJ, Phindela T, Connor SJ (2005) Use of Rainfall and sea surface temperature monitoring for malaria early warning system in Botswana. *Am J Trop Med Hyg* 73: 214-221.
17. Hotez PJ, Molyneux DH, Fenwick A, Ottesen E, Erlich-Sachs S, et al. (2006) Incorporating a rapid impact package for neglected tropical diseases with programs for HIV/AIDS, tuberculosis and malaria. *PLoS Med* 3: e102.
18. Lainson R, Shaw JJ (1978) Epidemiology and ecology of leishmaniasis in Latin-America. *Nature* 273: 595-600.
19. New M, Lister D, Hulme M, Makin I. (2002) A high-resolution data set of surface climate over global land areas. *Clim Res* 21:1-25.
20. Wolter K, Timlin MS (1998) Measuring the strength of ENSO—How does 1997/98 rank? *Weather* 53: 315-324.
21. Priestley MB (1988) *Non-linear and non-stationary time series analysis*. London: Academic Press.
22. Faraway JJ (2005) *Linear models with R*. Boca Raton: Chapman & Hall. CRC.
23. Kaplan M, Glass D (1995) *Understanding Nonlinear Dynamics* New York: Springer.
24. Efron B, Tibshirani R. (1993) *An introduction to the bootstrap*. London: Chapman & Hall.
25. Brockwell PJ, Davis RA (2002) *Introduction to time series and forecasting*. 2nd ed. New York: Springer.
26. Shumway RH, Stoffer DS (2000) *Time series analysis and its applications*. New York: Springer.

27. Levins R, Awerbuch T, Brinkmann U, Eckardt I, Epstein P, et al. (1994) The emergence of new diseases Am. Scien. 82: 52–60.
28. Clark JS, Carpenter SR, Barber M, Collins S, Dobson A, et al. (2001) Ecological forecasts: an emerging imperative. Science 293: 657-661.
29. Abeku TA, De Vlas SJ, Boorsboom GJJM, Tadeqe A, Gebreyesus Y, et al. (2004) Effects of meteorological factors on epidemic malaria in Ethiopia : a statistical modeling based approach based on theoretical reasoning. Parasitol 128: 585-593.
30. Abeku TA, De Vlas SJ, Boorsboom GJJM, Teklehaimanot A, Kebede A, et al. (2004) Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia: a simple seasonal adjustment method performs best. Trop Med Int Health 7: 851-857.
31. Teklehaimanot HD, Lipsitch M, Teklehaimanot A, Schwartz J (2004) Weather-based prediction of *Plasmodium falciparum* malaria in epidemic-prone regions of Ethiopia I. Patterns of lagged weather reflect biological mechanisms. Malar J 3: 41.
32. Teklehaimanot HD, Schwartz J, Teklehaimanot A, Lipsitch M (2004) Weather-based prediction of *Plasmodium falciparum* malaria in epidemic-prone regions of Ethiopia II. Weather-based prediction systems perform comparably to early detection systems in identifying times for interventions. Malar J 3: 44.
33. Zhou G, Minakawa N, Githeko AK, Yan G (2004) Association between climate variability and malaria epidemics in the east African highlands. Proc Natl Acad Sci USA 101: 2375-2380.
34. Hu WB, Nicholls N, Lindsay M, Dale P, McMichael AJ, et al. (2004) Development of a predictive model for Ross River virus disease in Brisbane, Australia. Am J Trop Med Hyg 71: 129-137.
35. Hu WB, Tong SL, Mengersen K, Oldenburg, B (2006) Rainfall, mosquito density and the transmission of Ross River: a time series forecasting model. Ecol Model 196: 505-514.
36. Grenfell BT, Kleczkowski A, Ellner SP, Bolker BM (1994) Measles as a case study in nonlinear forecasting and chaos. Phil Trans Roy Soc London A. 348: 515-530.
37. Royama T. (1992) Analytical Population Dynamics. London: Chapman and Hall.

38. Turchin P (2003) *Complex Population Dynamics*. Princeton: Princeton University Press.
39. Hassell MP, Lawton JH, May RM (1976) Patterns of dynamical behavior in single species populations. *J. Anim. Ecol.* 45: 471-486.
40. Turchin P, Taylor AD (1992) Complex dynamics in ecological time series. *Ecology*. 73: 289-305.
41. Ellner S, Turchin P (1995) Chaos in a noisy world new methods and evidence from time series analysis. *Am. Nat.* 145: 343-375.
42. Kendall BE, Prendergast J, Bjørnstad ON (1998) The macroecology of population dynamics: taxonomic and biogeographic patterns in population cycles. *Ecol. Lett.* 1: 160-164.
43. Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*. 344: 734-741.
44. Loevinsohn ME (1994) Climatic warming and increased malaria incidence in Rwanda. *Lancet*. 343: 714-718.
45. Feliciangeli MD, Rabinovich J (1998) Abundance of *Lutzomyia ovallesi* but no *Lu. gomezi* (Diptera: Psychodidae) correlated with cutaneous leishmaniasis incidence in north-central Venezuela. *Med. Vet. Entomol.* 12: 121-131.
46. Chaves LF, Hernandez MJ (2004) Mathematical modelling of American cutaneous leishmaniasis: incidental hosts and threshold conditions for infection persistence. *Acta Trop.* 92: 245-252.
47. Chaves LF, Hernandez MJ, Dobson AP, Pascual M (2007) Sources and sinks: revisiting the criteria for identifying reservoirs for American cutaneous leishmaniasis. *Trends Parasitol.* 23: 311-316.
48. Cazelles B, Hales S. (2006) Infectious diseases, climate influences and nonstationarity. *PLoS Med.* 3: e328.
49. Altizer S, Dobson AP, Hosseini P, Hudson PJ, Pascual M, et al. (2006) Seasonality and the dynamics of infectious diseases. *Ecol Lett.* 9: 467-484.
50. Murdoch WW, Briggs CJ, Nisbet RM (2003) *Consumer-Resource Dynamics*. Princeton: Princeton University Press.

51. Grenfell BT, Wilson K, Finkenstädt BF, Coulson TN, Murray S, et al. (2004) Noise and determinism in synchronized sheep dynamics. *Nature*. 394: 674-677.
52. Pascual M, Rodo X, Ellner SP, Colwell R, Bouma MJ (2000) Cholera Dynamics and El Niño- Southern Oscillation. *Science*. 289: 1766-1769.
53. Levin SA (1992) The problem of pattern and scale in ecology. *Ecology* 73: 1943-1967.
54. Grover-Kopec E, Kawano M, Klaver RW, Blumenthal B, Ceccato P, et al. (2005) An online operational rainfall-monitoring resource for epidemic malaria early warning system in Africa. *Malar J.* 4: 6.
55. Levins R (1966) The strategy of model building in population biology. *Am. Scien.* 54: 421-431.
56. Levins R (2006) Strategies of abstraction. *Biol. Philos.* 21:741-755.

CHAPTER IV

MODELLING DISEASES FROM FIRST PRINCIPLES: SELF REGULATORY FEEDBACKS, BIOLOGICAL INTERACTIONS AND CLIMATIC FORCING, THE ECOLOGICAL DYNAMICS OF VIVAX AND FALCIPARUM MALARIA

Introduction

Malaria, one of the most devastating infectious diseases in humans is widely distributed across the tropics. It is caused by four different species: *Plasmodium vivax*, *P. ovale*, *P. malariae* and *P. falciparum*. In most places 2 or more parasite species co-occur [e.g., 1], as well as several strains of any given species [e.g., 2]. This diversity poses a challenge to our understanding of the population dynamics of the disease, and several scenarios have been proposed to understand how the infection by one parasite species or strain determines the fate of an infection by another.

Classical views on the problem considered patterns of infection random, and ultimately determined by the action of climatic forces [3,4]. The development of ecological theory for competition [e.g., 5,6], in addition to a growing body of knowledge on the human immune system, led to the proposal of parasite cross-immunity (a.k.a., heterologous immunity) as a force regulating the infection by closely related parasites [1]. A major emphasis was placed on the specific immune response [7], which recognizes and neutralizes specific pathogens

through the response to parasite specific signals or antigens and the selection of T and B cells. This within-host mechanism implies a top-down regulation in co-infections because it leads to competition for susceptibles in the process of transmission between hosts. The sign of the interaction between parasites (co-infection) is negative, since the feedbacks between hosts and their immune system is negative both ways (see Figure 4.1). By using a phylogenetic argument, Cohen [1] proposed that species with common descent are likely to share the antigens that trigger an immune response. Therefore infection by one species would generate the “memory” to fight new infections by a closely related species. This idea was widely appealing because it led to a statistically testable hypothesis. The existence of cross-immunity predicts a number of cross-infections smaller than expected by random. Following Cohen [1], several authors supported the existence of cross-specific immunity by analyzing cross-sectional studies and finding the number of cross-infections to be below the one expected by random [8-12]. This idea also underlies more dynamical approaches to the subject of malaria immunity and the role of multiple strains of a given parasite species [13,14].

However, as pointed out by Molineaux et al [15], Cohen’s [1] method, requires high quality data and assumes that all individuals in a population are homogeneous in the way they manage infections. Molineaux et al [15] found the number of cross-infections to be higher than expected by random, a pattern that does not rule out cross-immunity, but can arise when all individuals in a population do not raise a proper immune response against the parasites.

Although largely ignored, the pattern described by Molineaux et al [15] was shown to be more common than originally suspected with the advent of molecular techniques, both at the intra-specific [2,16-18] and inter-specific level [19-21]. These findings suggest the limited use of parasite prevalence, or more generally data from any static diagnosis as a reliable measure of co-infection by closely related parasite species, as further supported by simulations on individual course of infections [22]

Advances in immunology also showed that immune responses can act through non-specific or innate mechanisms that do not lead to the generation of memory cells necessary for the development of heterologous immunity[7]. Before the discovery of many non-specific mechanisms of action for the immune system, dynamic transmission models for malaria, developed in the context of the Garki project included hosts whose qualitative behavior after infection differed [23]. In this model, some hosts develop temporary immunity, while others can be readily infected after clearing their parasites. Further analysis of the course of individual infections [24] strongly supported individual heterogeneity in immune responses, with no evidence for cross-specific or heterologous immunity at the population level. There was evidence, however, for a reduced amount of cross-strain or homologous immunity (at the cost of an increased tolerance to circulating parasites). Non-specific immunity implies a bottom-up regulation mechanism for co-infections (see Figure 4.1), in the sense that hosts can be seen as a self-regulating resource for the parasites (who act as consumers) through the action of the immune system, or alternatively we can consider hosts feeding their

immune system when infected, which implies a cost for immunity, that results in a positive interaction among parasites.

The model by Dietz et al [23] captured this diversity of mechanisms by subdividing the population according to the handling of infection and the development of immunity. It has been the most successful approximation to explain the dynamics of the populations in the Garki project [25-27]. However, the data on which the model was based is unique, since most epidemiological data are not as detailed as the sources for Garki. Also the model does not consider the possible effects of environmental forcing on the dynamics of the disease [28]. In the present paper we propose a time series modeling approach based on conceptual ideas from population dynamics [29-31]. This approach allows to combine the statistical framework of seasonal autoregressive models with mechanistic elements of a simple transmission model, where climatic forcing and the fraction of parasite free individuals in the population as limiting resource, can be considered explicitly. This model is used test three main mechanisms (randomness, top-down or bottom-up regulation) for parasite co-existence in a population of hosts.

Methods

Loop analysis for qualitative understanding of species interactions

Expectations for biological interactions among species can be studied through the analysis of a community matrix that describes the interactions of all its members around an equilibrium [32]. The negative inverse of this matrix shows the direction of change in abundance of the community members

following a small perturbation, where indirect effects product of the web of interconnections can be visualized [33,34]

Model and Theoretical considerations

Consider a general equation for the dynamics of infected hosts I :

$$\frac{dI}{dt} = I(b - \delta) \quad (1)$$

where b accounts for the recruitment of new infected individuals and δ for the recovery of those already infected [35]. I is the state that is detected and recorded as disease incidence.

We can equate $b - \delta$ to the instantaneous per capita growth rate of infections, $r(t)$, and use the approximate integration method presented by Turchin [30; pages 53-54]. By assuming that $r(t) = \ln(I_t/I_{t-1})$ remains constant for a discrete time step equation (1) can be written in discrete time as follows:

$$I_t = I_{t-1} e^{r(t)} \quad (2)$$

Exogenous forcing, $exoF(t)$, can be added within $r(t)$ [36,37]. To define $r(t)$, we consider a general mass action transmission [38] and let $b = \beta S/N$, (where β is the transmission rate, N the total population size, S the susceptible population size). With $S = N - I$ equation (1) becomes:

$$\frac{dI}{dt} = I \left(\beta' - \beta \frac{I}{N} \right) \quad (3)$$

where $\beta' = \beta - \delta$.

The above model can be generalized to let I represent new cases, with the number of parasitemic hosts represented at any given time step with a new class

(I'), which is a function of infected individuals, $I'=f(I)$ to handle the difference between the time for parasite clearance to the length of the modeled time step. The class for parasitemic hosts is also necessary to account for the regulation in the recruitment of new infected individuals by controlling the number of susceptible hosts in the following manner: $S = N-I'$. Therefore $r(t)$ can be defined as follows:

$$r(t) = \beta' - \beta \frac{I'_t}{N_t} + \alpha \text{exo} F_{t-\tau} + w_t \quad (4)$$

where w_t is an i.i.d. normal random variable accounting for unexplained variation.

The basic model in discrete time becomes:

$$I_t = I_{t-1}^\theta \exp\left(\beta' - \beta \frac{I'_t}{N_t} + \alpha \text{exo} F_t + w_t\right) \quad (5)$$

where $0 < \theta < 1$ is included to consider deviations from mass-action in transmission [39]. It also can be seen as a density dependence factor in the generation of new infections [40,41], or to account for the spatial clustering among hosts during the process of infection [42].

Notice that under the abstraction behind this model the growth of infections, or new cases, is tracked. The growth is limited by the number of hosts that can be infected. Therefore, this formulation also can be seen as one where parasites colonize hosts in a metapopulation fashion, and the interpretation of I_{t-1} on the right side should not be restricted to the one of models based on the concept of force of infection, where cases at I_{t-1} are the only individuals contributing to transmission [42].

Seasonality: Classical work on the dynamics of infectious diseases has recognized the meaning of seasonality as deviations from a mean annual value, that occur every year approximately after a fixed period, σ , that corresponds to the number of annual subdivisions (months, weeks) used to accumulate cases [3]. $r(t)$ can be defined as a seasonal function:

$$r(t) = r_{\text{seas}}(k) + \beta' - \beta \frac{I'_t}{N_t} + \alpha \text{exo}F_t + w_t \quad (6)$$

where $r_{\text{seas}}(k)$ accounts for the seasonal contribution during the k^{th} period of any given season.

Most models incorporate seasonal variation in transmission using fixed seasonalities, either resorting to mean values for the temporal subdivisions of data collection (weeks, months, etc) [e.g., 42] or symmetrical functions, like sines and cosines [e.g., 43-44]. This approach can lead to a symmetrical cyclical behaviour that lacks the inherent seasonal variability of a time series, limiting the ability of models to reduce the variability due to unknown factors [45]. Seasonal autoregressive forms [46,47] overcome this limitation by letting the value of a variable x at time t is a function of its previous seasonal value ($t-\sigma$). Therefore $r_{\text{seas}}(k)$ can be defined as a function of $r(t-\sigma)$. When this function can be linearized (using a parameter φ), $r_{\text{seas}}(k)$ becomes:

$$r_{\text{seas}}(k) = \varphi \log(I_{k-\sigma}) \quad (7)$$

Then the seasonal version of the model in equation (5) can be written as:

$$I_t = I_{t-1}^{\theta} I_{t-\sigma}^{\varphi} \exp\left(\beta' - \beta \frac{I'_t}{N_t} + \alpha \text{exo}F_t + w_t\right) \quad (8)$$

Parasite Clearance and Immunity: we consider the variation among individuals in the duration of parasite clearance (with or without the generation of long lasting immunity) by introducing a random variable d whose distribution describes the length of individual parasitemia. Then:

$$I'_t = \sum_{i=0}^{\infty} (1 - cmf(i)) * I_{t-i} \quad (9)$$

where cmf is the cumulative mass function of the random variable d at time i , and I_{t-i} , the number of new cases at time t . At equilibrium in (3): $1/\delta = E[d]$.

Randomness, Top-Down, Bottom-Up mechanisms: the model presented in (8) can be extended to account for possible inter-specific interactions between 2 parasites, by computing a coefficient (β_2) for the second species as follows:

$$I_t = I_{t-1}^0 I_{t-\sigma}^{\sigma} \exp\left(\beta_1 - \beta \frac{I_t}{N_t} \pm \beta_2 \frac{I_2}{N_t} + \alpha exo F_t + w_t\right) \quad (10)$$

The following possibilities correspond to the different mechanisms we wish to test for. If $\beta = \beta_2 = 0$, then random interactions are supported. If $\beta < 0; \beta_2 < 0$ (both coefficients with negative signs) the most likely mechanism is one where cross-specific (heterologous) immunity is present, since the interaction of the parasites have the signs expected under feedback loops of top-down regulation, i.e., the effect of the second species on the first is negative (see Figure 4.1). If $\beta < 0; \beta_2 > 0$ (the focal species with a negative coefficient, the second species with a positive one) the most likely mechanism is one where non-specific immunity is at play, since the interaction of the parasites have the signs expected under

feedback loops of bottom-up regulation, i.e., the effect of the second species on the first is positive (see Figure 4.1).

Data

Monthly records of malaria in Espirito Santo, Vanuatu were obtained from people attending government health centers (free of charge) who presented with fever or a recent history of fever, and whose standard blood slide analysis indicated infection with either *Plasmodium vivax* or *P. falciparum*, from January 1983 to December 1997. Additional data on distributed insecticide treated nets, ITNs, with permethrin and population growth were available for the same period and obtained from the Malaria and other Vector Borne Diseases Control Unit, Ministry of Health, Port Vila, Vanuatu (Figure 4.2,4.3). The seasonal patterns for the malaria cases are presented in Figure 4.2.

Model Fitting

Estimation of d and I' : to estimate d we used the data from *P. falciparum* malariotherapy for neurosyphilis patients published by Collins and Jeffery [48], and analyzed by Sama et al [49]. We considered the datasets from Georgia and South Carolina (Figure 4.3), as examples of a population with minimum use of drugs and some use of drugs respectively. Unlike Sama et al [49], who used continuous distributions, we fitted a discrete distribution to the data, given the discrete nature of the data and the way we computed I' . Negative Binomial distributions minimized the likelihood when compared with other discrete

distributions (e.g., poisson). I' was obtained with equation (8) and the series of I for each species, with month length was based on each calendar year, the cmf was truncated at the end of a calendar year (365 days). Since no similar data is available for *P. vivax* we assumed the clearance time distributions to be the same as for *P. falciparum*.

Inapparent infections: We considered the effect of inapparent infections on the series by multiplying the I' series by the ratios published by Maitland et al (1996) of 1:1 in the wet season (November-May) and 1:4 in the dry season (June-October) for *P. falciparum*, and 1:1 in the wet season and 1:2 in the dry one for *P. vivax*.

Effect of bednets: To account for the effects of bednets we considered two possibilities; bednets either diminish the size of N , the total host population, or they increase the feedback, i.e., the number of available hosts for the generation of new infections via I' . In both cases the effects were assumed to be additive, i.e., $N_b = N - \text{No. bednets}$, $I_b' = I' - \text{No. bednets}$. We assume the effect of bednets was cumulative or transient (just for the month when they were delivered). We also computed the % population covered with Insecticide treated bednets for each delivery (Figure 4.3) and assumed the effects of the bednets lasted from 1 to 6 months, the lifespan of Permethrin [50].

Exogenous Forcing: to find appropriate lags for the introduction of climatic variables we used the pre-whitening method described in [51]. We found that temperature (T) with 5 months of lag was statistical significant for both species. We used records from the airport of Pekoia, a close-by island, available at

<http://weather.noaa.gov/>. Missing data were imputed from a time series for the political area of Vanuatu [52]. For the analysis, the temperature series was demeaned.

Parameter estimation: For parameter estimation, we linearized the model in (10) as follows:

$$\log(I_t) = \theta \log(I_{t-1}) + \varphi \log(I_{t-12}) + \beta_1 \frac{I_t'}{N_t} \pm \beta_2 \frac{I_{2t}'}{N_t} + \alpha T_{t-5} + w_t \quad (11)$$

and used a fitting procedure for negative binomial generalized linear models (NB-GLM) for the time series of each species. To avoid confusion in the parameters, the finding of parameters with incorrect signs or magnitude, a problem to be expected because of the linear algebra behind linear models [53,54], the estimation of parameter β was restricted between $(-\infty, -1]$, using the Nelder-Mead algorithm [55]. The parameter β needs to be larger than the recovery rate, given the endemicity of the disease and recovery rates constrained to be smaller than 1, since average duration of infections is longer (>3 months) than the modeled time step (1 month). Since likelihood ratio tests for NB-GLMs are only reliable for a fixed over-dispersion parameter (k), we compared the most complex model with simplified ones fixing the overdispersion parameter of the most complex. To ensure robustness, we also make comparisons the other way around. The models were fitted only to the data from January 1985 to December 1997, given the need for burning values in I' .

Results

Figure 4.3 shows the distribution of clearance times in neurosyphilis patients treated with malariotherapy. For Georgia, long-time clearance patients, the mean of the negative binomial distribution was 169.15 days, which is close 6 months, with an over-dispersion parameter of 2.10. For South Carolina, short-time clearance patients, the mean of the negative binomial distribution was 78.40 days (~3 months) with an over-dispersion parameter of 3.51. Figure 4.3 also shows the proportion of people locally covered with bednets, which was on average about 5% of the island population.

Table 4.1 shows the search for the best model under the scenarios considered and gives an idea of the robustness of the findings. For both species, consideration of inapparent infections increases the likelihood (minimizes the AIC) of the models. A similar result was found for the long-time clearance, which is partially shown for the best models in Table 4.2. For *P. falciparum*, all models without *P. vivax* outperformed the corresponding models with it. For *P. vivax* the opposite result was found, with models that consider the parasitemic individuals for *P. falciparum* have the highest likelihood. For both species, the likelihood was also maximized when the effect of bednets was considered transient and short, only one month (Table 4.1).

Temperature was an important driver for the dynamics of both species with a lag of 5 months. For both species the magnitude of this forcing is very small, about 10%, when compared with the parameter β' , the difference between transmission and recovery (Table 4.2). In absolute terms the effect of

temperature is slightly larger, about 25%, for *P. falciparum* ($\hat{\alpha} \cong 0.11$), than for *P. vivax* ($\hat{\alpha} \cong 0.08$).

Finally, Figure 4.4 shows the success of the best models in fitting the data. The correlation between observed and fitted values is 0.77 for the best *P. falciparum* model, and 0.74 for the best model for *P. vivax*.

Discussion

Malaria models have been evolving [56-58] since the first proposals of Ross [59]. As knowledge on intra-host dynamics has accumulated, and regulatory mechanisms for parasitic infections are being elucidated, it is becoming clear that innate (non-specific) immunity plays a key role, since its direct action or mediation in regulating adaptive (specific) responses are fundamental for the proper development of immunity [60-64]. Although several models have considered these new findings explicitly [9,11,58,65-67], as well as the effect that drugs may have on the dynamics of the disease [68-69], the question of parasitic infections regulation has been not addressed at the population level. Results of our time series model support the regulation of malaria infections by the parasites present in the population. The process of model selection showed that the likelihood of models is minimized when a short-time of clearance (3 months) is used instead of a longer time (6 months). This suggests that long lasting population immunity is not a major force regulating the population dynamics of infection. However, this result also indicates that either some kind of immunity to clear parasites is developed or population level use of

antimalarials have an effect in shortening clearance times, since the longer estimate corresponds to immunologically naïve populations [49]. In fact, if we consider the difference between the fixed values of β to β' , i.e., the recovery rates, and compute the inverse, we get values of 5-6 months for the amount of time individuals are removed from susceptibility, a time longer than the average clearance of 3 months. However, this point can be elucidated by resorting to other tools for modeling where a more clear distinction between immunity and parasite clearance can be separated, and by having estimates of population level use of antimalarial drugs. The regulation of transmission by parasite abundance was originally proposed by Ross [59] and it was the basis for future developments of control based on reducing mosquito population. Strategies for mosquito control, primarily through mosquito larval habitat reduction [70-71] and bednets [50,72-75], have been very successful, by contrast to vaccine trials [76-77].

The transient effect of bednets at the population level can be explained by the low local coverage associated with deliveries, which was about 5% of the population, when compared to the average of 80% seen at other localities [50,72,73], indicating that the threshold for population effectiveness of bednets was never reached in this island. Although the effect of bednets seems to be primarily one of reducing the total population at risk of infection, more sophisticated models are needed to understand their regulatory function.

The effect of climatic forcing in regulating transmission (or growth of infections) was very small when compared to the endogenous factors regulating

the population, in accordance with similar findings in other populations [29,78,79], and slightly larger for *P. falciparum*, in accordance with the effects seen for the whole archipelago of Vanuatu [74].

Differences in seasonal patterns and age specific prevalence have been used as evidence for a heterologous (cross-specific) immunity for these two malaria parasites in the setting we studied [80,81]. However, the seasonal patterns are very similar for the two species over a longer time horizon (Figure 4.2), the infection with one species does not seem to reduce the infection with the other, and in the biogeographical region where Vanuatu is located, multi-species infections are very common, although only easily detected by very sensitive molecular based techniques [20,21,82]. More generally, as mentioned in the introduction, prevalence is a static measure that can blur parasites' interactions in the dynamics, and even co-infection, since the sampling of parasites on blood slides is sensitive to their intra-host density, as demonstrated by sampling longitudinal data on the same individuals with co-infections at random times during fever episodes [22]. By contrast our results are consistent with observations from the malariatherapy patients [83,84], where following a co-infection, *P. falciparum* reached higher densities first, in some cases suppressing the growth of *P. vivax*, and with evidence on the absence of heterologous immunity at the population level [24,85]. The lack of heterologous immunity may be explained by differences in resource exploitation by the two parasites inside the hosts, since *P. falciparum* is able to parasitize all erythrocytes while *P. vivax* only parasitizes young ones, leading to an enhanced parasitemia for the latter

[86]. This difference can also explain the dynamic sufficiency of *P. falciparum* dynamics, since it can reach peak densities in the presence of *P. vivax*, whereas the latter reaches its maximum only after the former parasite is cleared [84].

For the interaction between the parasites to be positive at the population level, a mechanism of self-regulation that is density dependent is likely to be at play as proposed by Bruce and Day [82]. This is supported both by observations and theoretical results. The results by Bruce et al [20] suggest that parasites are likely to be regulated only when they reach large densities, since fluctuating densities of parasites through time were reported together with the development of tolerance to malaria parasites in the malariotherapy patients [24]. Our results with loop analysis show that when the two parasites co-occur the positive effect of one species on the other is plausible because of the self-regulation of the former as expected under a bottom-up regulation mechanism (Figure 4.1).

Another explanation for the observed dynamics is the switch of immune strategy of individuals with age, since there is evidence for changes from non-specific to specific immune responses in hyper-endemic settings elsewhere [60,87]. The demographic profile of Vanuatu, Espirito Santo included, shows that at least 30% of the total population is under 15, and most of the cases are concentrated in this age-group [50,73]. Although adults may be an important source of infections [88] children are the main source of gametocytes (infecting stage to mosquitoes) in hyperendemic settings, where prevalence is above 20%, like Espirito Santo [89]. This would explain why long-lasting immune responses

that can lead to heterologous immunity do not play a major role in regulating the dynamics of transmission.

Finally, our study emphasizes the need to understand the factors regulating the dynamics of infection before formulating strategies of control at the population level. The failure of strategies that target infection management through immunity may ultimately be determined by their irrelevance to the regulation of disease dynamics at the population level.

Table 4.1 Model Search. Species 1 indicates the species studied, Species 2 indicates whether a second species was considered, Inapparent Infections whether the ratios for inapparent infections were considered, Bednets indicates how the bednet effect was considered and AIC, the Akaike Information Criterion value for the models (highlighted values indicate the minimum for each species). Models are not directly comparable because they have different overdispersion parameters. NC indicates models whose parameters did not converge in the iterative process for fitting the NB-GLMs. When considering the second species, its I' was the best estimate for the species alone.

Species 1	Species 2	Inapparent Infections	Bednets	AIC	
<i>Plasmodium falciparum</i>	<i>P. vivax</i>	No	N_b (Constant 1 month)	1916.3	
			N_b (Cumulative)	NC	
			I_b (Constant 1 month)	1770.1	
		Yes	No	I_b (Cumulative)	1838.9
				No effect	1772.9
				N_b (Constant 1 month)	1769.9
	Yes		N_b (Cumulative)	1763.4	
			I_b (Constant 1 month)	1768.7	
			I_b (Cumulative)	1813.0	
	None	No	No effect	1771.7	
			N_b (Constant 1 month)	1908.1	
			N_b (Cumulative)	NC	
			I_b (Constant 1 month)	1772.0	
		Yes	I_b (Cumulative)	1814.6	
No effect			1766.4		
N_b (Constant 1 month)			1764.9		
N_b (Cumulative)			1765.4		
<i>Plasmodium vivax</i>	<i>P. falciparum</i>	No	I_b (Constant 1 month)	1770.7	
			I_b (Cumulative)	1813.0	
			No effect	1765.2	
		Yes	No	N_b (Constant 1 month)	1633.2
				N_b (Cumulative)	1707.0
				I_b (Constant 1 month)	1567.9
	Yes		I_b (Cumulative)	1627.4	
			No effect	1562.1	
			N_b (Constant 1 month)	1561.1	
	None	No	N_b (Cumulative)	1565.4	
			I_b (Constant 1 month)	1568.2	
			I_b (Cumulative)	1627.2	
			No effect	1573.9	
		Yes	N_b (Constant 1 month)	1623.4	
N_b (Cumulative)			1717.0		
I_b (Constant 1 month)			1563.4		
I_b (Cumulative)			1620.5		
None	No	No effect	1574.3		
		N_b (Constant 1 month)	1565.7		
		N_b (Cumulative)	1562.2		
	Yes	I_b (Constant 1 month)	1564.2		
		I_b (Cumulative)	1620.3		
		No effect	1562.2		

Table 4.2 Parameter estimates for the best models (see methods section for definitions). Parameter values are assuming short-times for parasite clearance (South Carolina patients). AIC values inside parenthesis are for the same model assuming long-times for parasite clearance (Georgia patients for malariotherapy).

Species	$\hat{\theta}$	$\hat{\phi}$	$\hat{\beta}'$	$\hat{\beta}_2$	$\hat{\alpha}$	\hat{k}	AIC
<i>P. falciparum</i>	0.757 ± 0.059	0.114 ± 0.054	0.785 ± 0.381	1.33 ± 3.00	-0.108 ± 0.028	6.42 ± 0.75	1769.6 (1772.9)
<i>P. falciparum</i>	0.737 ± 0.050	0.121 ± 0.050	0.842 ± 0.343	-	-0.106 ± 0.028	6.53 ± 0.76	1764.9 (1766)
<i>P. vivax</i>	0.648 ± 0.058	0.155 ± 0.048	0.820 ± 0.240	3.66 ± 1.55	-0.0796 ± 0.0270	7.58 ± 0.96	1561.1 (1566.6)

Figure 4.1. Feedback loops for top-down and bottom-up regulation of multi-species malaria infections. Negative effects are indicated by open circles, positive by arrows, unknown effects are indicated by dotted lines. H is hosts, I is immune system, Pf stands for *Plasmodium falciparum*, and Pv for *P. vivax*. Under top-down regulation (A) the interaction of hosts and their immune systems is negative both ways, the interaction of hosts with the parasites control their immunity and the immunity regulate the hosts available for the parasites at the population level. Therefore each parasitic infection has a negative effect on the other (B). Under a bottom-up regulation (C), the hosts feed their immune system when parasitized, with a cost indicated by the negative effect of I on H, resulting in a positive interaction among parasites (D) and self regulation for each species. (C) and (D) were obtained through a loop analysis (Puccia and Levins, 1985).

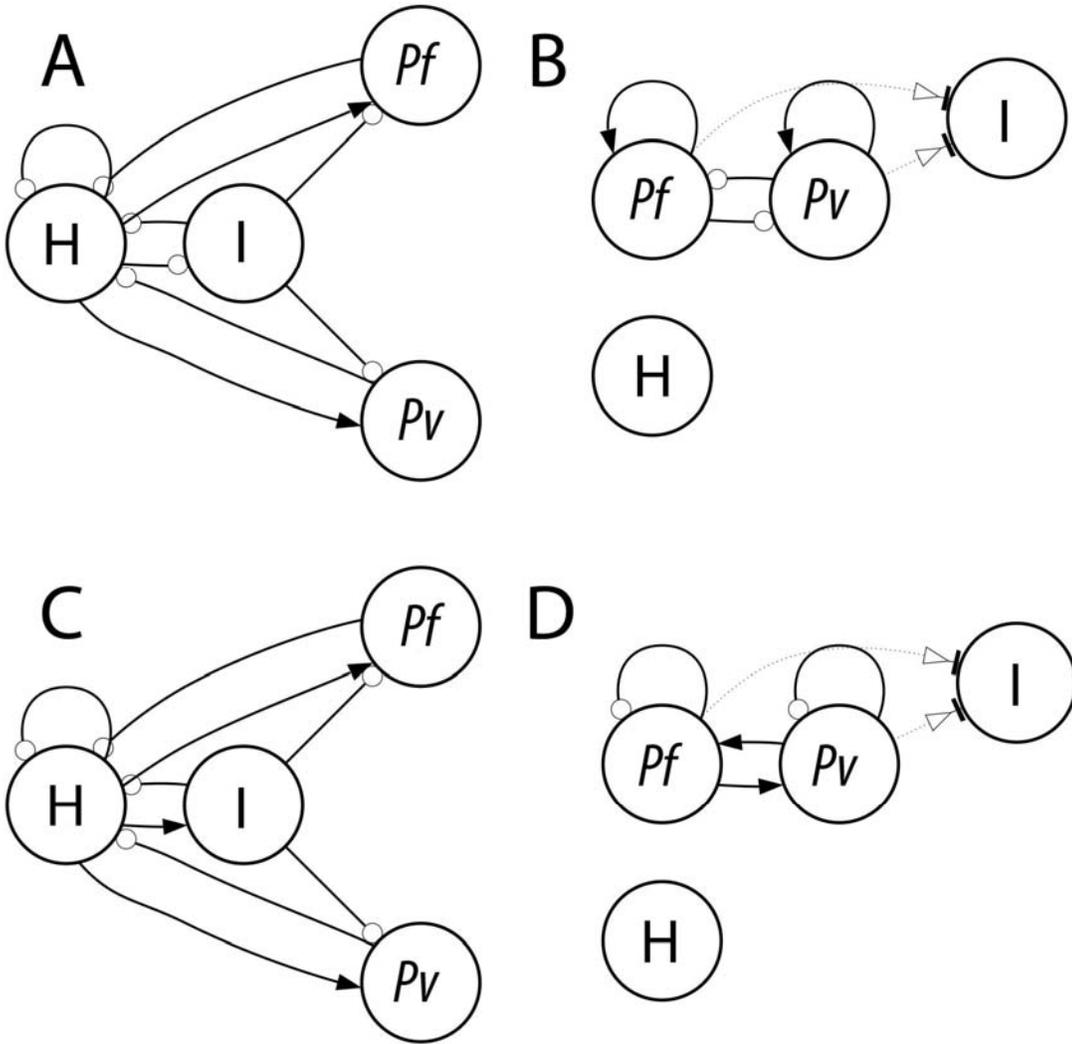


Figure 4.2. Time series: *Plasmodium falciparum* malaria: monthly incidence (A) and its seasonality (B). *P. vivax* malaria monthly incidence (C) and its seasonality (D). Population size (E) and temperature from Peko Airport (F).

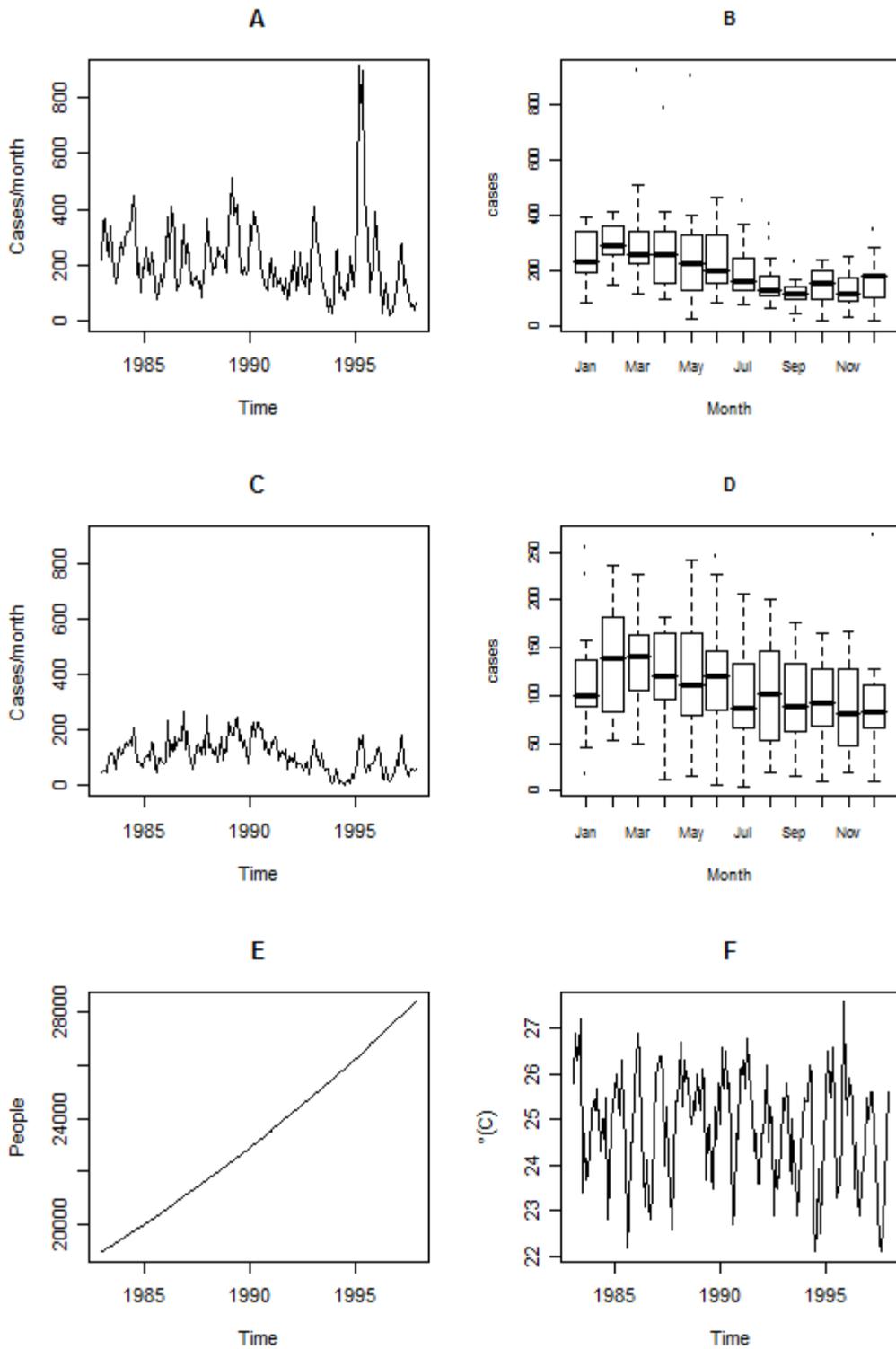


Figure 4.3. Bednets and Parasite Clearance. Monthly number of bednets distributed (A), percent of people locally covered with each bednet distribution (B), long-time parasite clearance for Georgia patients (C), Short-time parasite clearance for South Carolina patients (D).

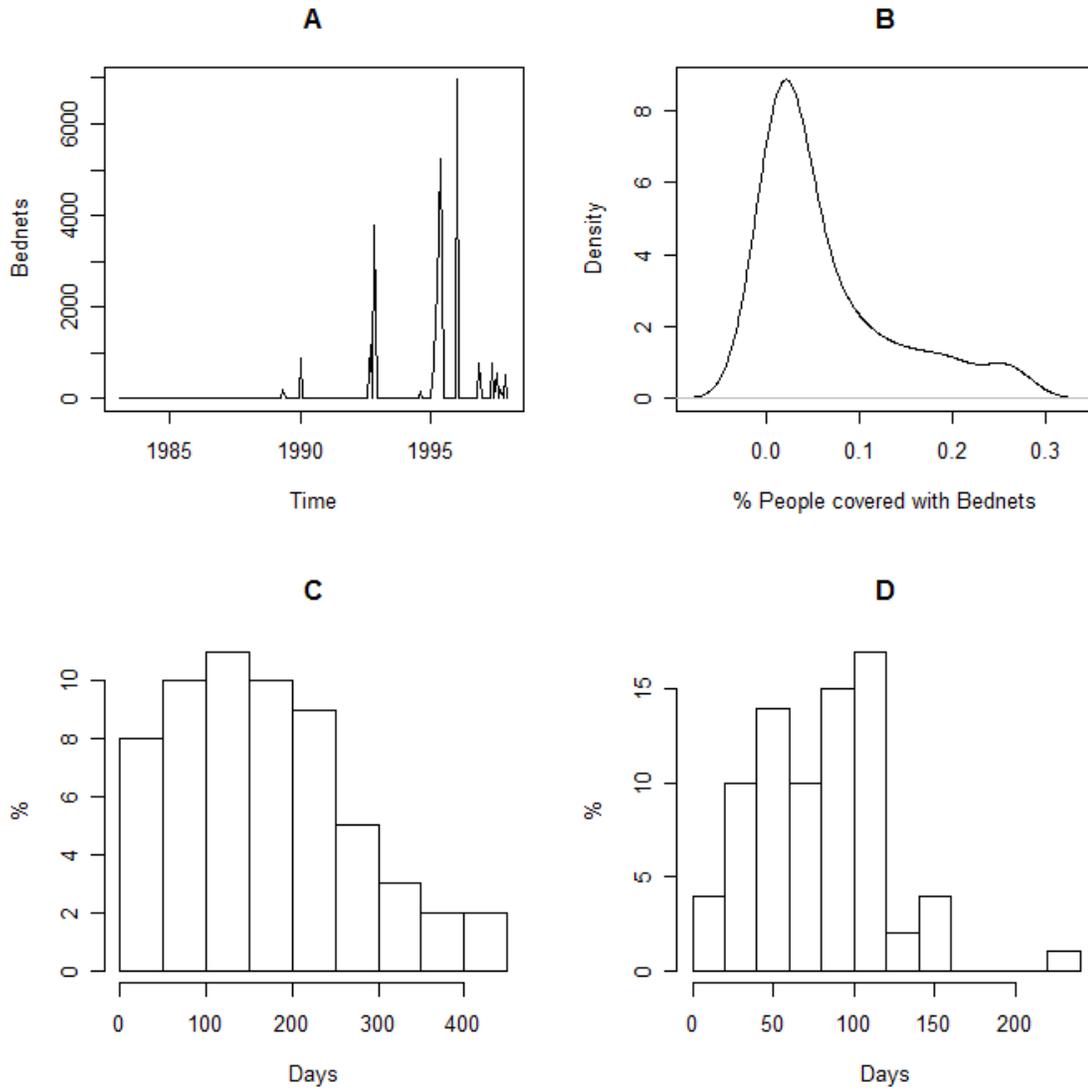
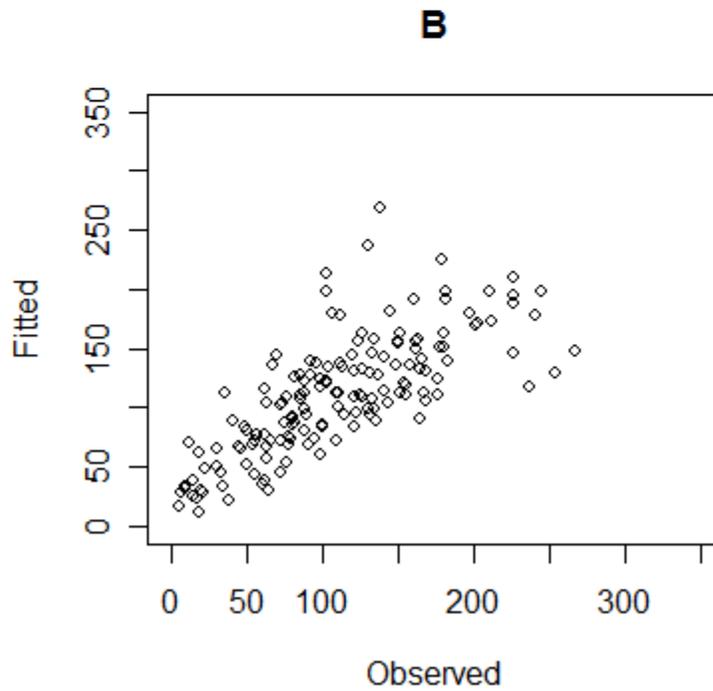
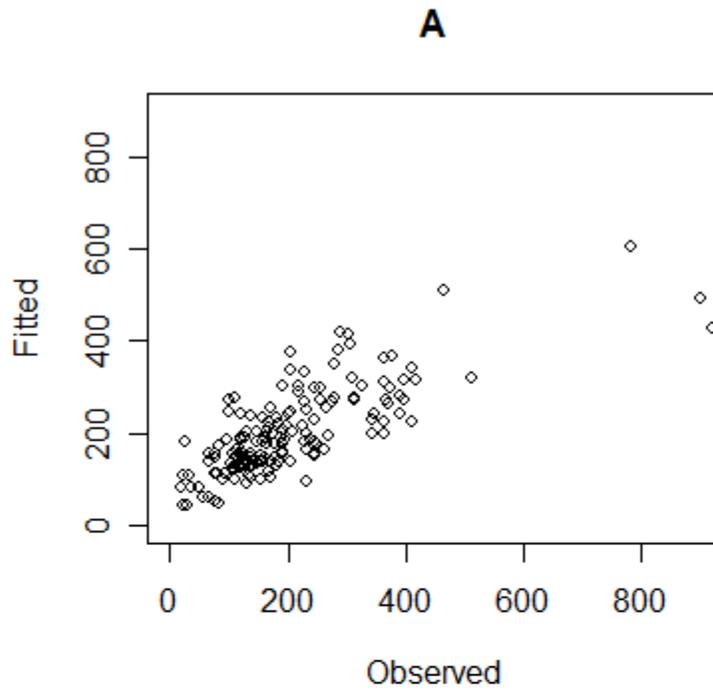


Figure 4. Fitted vs Observed values for the best models. *Plasmodium falciparum* malaria, Pearson's $r=0.77$, $CI=[0.70,0.83]$ (A); *P.vivax* malaria, Pearson's $r=0.74$, $CI=[0.67,0.81]$ (B)



References

1. Cohen JE (1973) Heterologous immunity in human malaria. *Quart Rev Biol* 48:467-489.
2. Farnert A, Snounou G, Rooth I, Bjorkman A (1997) Daily dynamics of *Plasmodium falciparum* subpopulations in asymptomatic children in a holoendemic area. *Am J Trop Med Hyg* 56:538-547.
3. Gill CA (1928) The genesis of epidemics and the natural history of disease. An introduction to the science of epidemiology based upon the study of epidemics of malaria, influenza and plague. London: Bailliere, Tindall and Cox.
4. Gill CA (1938) The seasonal periodicity of malaria and the mechanism of the epidemic wave. London: J. & A. Churchill Ltd.
5. Gause GF (1934) The struggle for existence. Baltimore: Williams and Wilkins.
6. Vandermeer JH (1969) Competitive structure of communities - an experimental approach with protozoa. *Ecology* 50:362-&.
7. Kuby J (1997) Immunology. New York : W.H. Freeman
8. Richie TL (1988) Interactions between malaria parasites infecting the same vertebrate host. *Parasitology* 96:607-639.
9. McKenzie FE, Bossert WH (1997) Mixed-species *Plasmodium* infections of humans. *J Parasitol* 83:593-600.
10. McKenzie FE, Bossert WH (1999) Multispecies *Plasmodium* infections of humans. *J Parasitol* 85:12-18.
11. Smith T, Genton B, Baea K, Gibson N, Narara A, Alpers MP (2001) Prospective risk of morbidity in relation to malaria infection in an area of high endemicity of multiple species of *Plasmodium*. *Am J Trop Med Hyg* 64:262-267.
12. Howard SC, Donnelly CA, Chan MS (2001) Methods for estimation of associations between multiple species parasite infections. *Parasitology* 122:233-251.
13. Gupta S, Day KP (1994) A strain theory of malaria transmission. *Parasitol Today* 10:476-481.

14. Gupta S, Day KP (1994). A theoretical framework for the immunoepidemiology of *Plasmodium falciparum* malaria. *Parasite Immunology* 16:361-370.
15. Molineaux L, Storey J, Cohen JE, Thomas A (1980) A longitudinal-study of human malaria in the west-african savanna in the absence of control measures - relationships between different plasmodium species, in particular *P. falciparum* and *Plasmodium-malariae*. *Am J Trop Med Hyg* 29:725-737.
16. Farnert A, Rooth I, Svensson A, Snounou G, Bjorkman A (1999) Complexity of *Plasmodium falciparum* infections is consistent over time and protects against clinical disease in Tanzanian children. *J Inf Dis* 179:989-995.
17. Felger I, Smith T, Etoh D, Kitua A, Alonso P, Tanner M, Beck HP (1999) Epidemiology of multiple *Plasmodium falciparum* infections - 6. Multiple *Plasmodium falciparum* infections in Tanzanian infants. *Trans Roy Soc Trop Med Hyg* 93:S29-S34.
18. Smith T, Beck HP, Kitua A, Mwankusye S, Felger I, Fraser-Hurt N, Irion A, Alonso P, Teuscher T, Tanner M (1999) Epidemiology of multiple *Plasmodium falciparum* infections - 4. Age dependence of the multiplicity of *Plasmodium falciparum* infections and of other malariological indices in an area of high endemicity *Trans Roy Soc Trop Med Hyg* 93:S15-S20.
19. Mehlotra RK, Lorry K, Kastens W, Miller SM, Alpers MP, Bockarie M, Kazura JW, Zimmerman PA (2000) Random distribution of mixed species malaria infections in Papua New Guinea. *Am J Trop Med Hyg* 62:225-231.
20. Bruce MC, Donnelly CA, Alpers MP, Galinski MR, Barnwell JW, Walliker D, Day KP (2000) Cross-species interactions between malaria parasites in humans. *Science* 287:845-848.
21. Mayxay M, Pukrittayakamee S, Newton PN, White NJ (2004) Mixed-species malaria infections in humans. *Trends Parasitol* 20:233-240.
22. O'Meara WP, Collins WE, McKenzie FE (2007) Parasite prevalence: A static measure of dynamic infections. *Am J Trop Med Hyg* 77:246-249.
23. Dietz K, Molineaux L, Thomas A (1974) Malaria model tested in african savannah. *Bull W H O* 50:347-357.
24. Molineaux L, Trauble M, Collins WE, Jeffery GM, Dietz K (2002) Malaria therapy reinoculation data suggest individual variation of an innate

immune response and independent acquisition of antiparasitic and antitoxic immunities. *Trans Roy Soc Trop Med Hyg* 96:205-209.

25. Nedelman J (1984) Inoculation and recovery rates in the malaria model of Dietz, Molineaux, and Thomas. *Math Biosci* 69:209-233.
26. Nedelman J (1985) Some new thoughts about some old malaria models - introductory review. *Math Biosci* 73:159-182.
27. Struchiner CJ, Halloran ME, Spielman A (1989) Modeling malaria vaccines .1. new uses for old ideas. *Math Biosci* 94:87-113.
28. Pascual M, Cazelles B, Bouma MJ, Chaves LF, Koelle K (2008) Shifting patterns: malaria dynamics and rainfall variability in an African highland. *Proc Roy Soc B* 275:123-132.
29. Royama T (1992) *Analytical Population Dynamics*. London: Chapman and Hall.
30. Turchin P (2003) *Complex Population Dynamics*. Princeton: Princeton University Press.
31. Berryman AA (2002) Population: a central concept for ecology? *Oikos* 97:439-442.
32. Levins R (1968) *Evolution in Changing Environments*. Princeton University Press, Princeton, NJ, USA.
33. Puccia CJ, Levins R (1985) *Qualitative Modeling of Complex Systems*. Harvard University Press, Cambridge, MA, USA.
34. Zavaleta JO, Rossignol PA (2004) Community level analysis of risk of vector-borne disease. *Trans Roy Soc Trop Med Hyg* 98:610-618.
35. Anderson RM, May RM (1992) *Infectious Diseases of Humans*. Oxford University Press.
36. Levins R (1969) Effect of random variations of different types on population growth. *Proc Nat Acad Sci USA* 62:1061-&.
37. Lewontin RC, Cohen D (1969) On population growth in a randomly varying environment. *Proc Nat Acad Sci USA* 62:1056-&.
38. De Jong MCM (1995) How does transmission of infection depend on population size? Epidemic models, their structure and relation to data (D. Mollison Ed), pp, 84-94. Cambridge: Cambridge University Press.

39. Liu WM, Hethcote HW, Levin SA (1987) Dynamic behavior of epidemiologic models with nonlinear incidence rates. *J Math Biol* 25:359-380.
40. Hochberg ME (1991) Nonlinear transmission rates and the dynamics of infectious-disease. *J Theor Biol* 153:301-321.
41. Fenton A, Fairbairn JP, Norman R, Hudson PJ (2002) Parasite transmission: reconciling theory and reality. *J Anim Ecol* 71:893-905.
42. Finkenstadt BF, Grenfell BT (2000) Time series modelling of childhood diseases: a dynamical systems approach. *J Roy Stat Soc C-App Stat* 49:187-205.
43. Kot M, Schaffer WM, Truty GL, Graser DJ, Olsen LF (1988) Changing criteria for imposing order. *Ecol Model* 43:75-110.
44. Pascual M, Rodo X, Ellner SP, Colwell R, Bouma MJ (2000) Cholera dynamics and El Nino-Southern Oscillation. *Science* 289:1766-1769.
45. Priestley MB, (1988) Non-linear and non-stationary time series analysis. London: Academic Press.
46. Shumway RH, Stoffer DS (2000) Time series analysis and its applications. New York: Springer.
47. Brockwell PJ, Davis RA (2002) Introduction to time series and forecasting. 2nd ed. New York: Springer.
48. Collins WE, Jeffery GM (1999) A retrospective examination of secondary sporozoite- and trophozoite-induced infections with *Plasmodium falciparum*: Development of parasitologic and clinical immunity following secondary. *Am J Trop Med Hyg* 61:20-35.
49. Sama W, Dietz K, Smith T (2006) Distribution of survival times of deliberate *Plasmodium falciparum* infections in tertiary syphilis patients. *Trans Roy Soc Trop Med Hyg* 100:811-816.
50. Kaneko A, Taleo G, Kalkoa M, Yaviong J, Reeve PA, Ganczakowski M, Shirakawa C, Palmer K, Kobayakawa T, Bjorkman A (1998) Malaria epidemiology, glucose 6-phosphate dehydrogenase deficiency and human settlement in the Vanuatu Archipelago. *Acta Tropica* 70:285-302.

51. Chaves LF, Pascual M (2006) Climate cycles and forecasts of cutaneous leishmaniasis, a nonstationary vector-borne disease. *Plos Medicine* 3:1320-1328.
52. Mitchell TD, Hulme M, New M (2002) Climate data for political areas. *Area* 34:109-112.
53. Faraway JJ (2004) *Linear Models with R*. Boca Raton: Chapman CRC Press.
54. Ellner SP, Seifu Y, Smith RH (2002) Fitting population dynamic models to time-series data by gradient matching. *Ecology* 83:2256-2270.
55. Nelder JA, Mead R (1965) A simplex-method for function minimization. *Computer J* 7:308-313.
56. Bailey NTJ (1982) *The biomathematics of Malaria*. London: Griffin.
57. Awerbuch T (1994) Evolution of mathematical models of epidemics. *Ann N Y Acad Sci* 740:232-241
58. Smith T, Maire N, Ross A, Tanner M (2006) A platform for stochastic modeling of malaria epidemiology and control. *Icopa XI: Proceedings of the 11th International Congress of Parasitology*:189-193.
59. Ross R (1911) *The prevention of Malaria*. London: John Murray.
60. Artavanis-Tsakonas K, Tongren JE, Riley EM (2003) The war between the malaria parasite and the immune system: immunity, immunoregulation and immunopathology. *Clin Exp Immunol* 133:145-152.
61. McIntosh RS, Shi JG, Jennings RM, Chappel JC, de Koning-Ward TF, Smith T, Green J, van Egmond M, Leusen JHW, Lazarou M, van de Winkel J, Jones TS, Crabb BS, Holder AA, Pleass RJ (2007) The importance of human Fc gamma RI in mediating protection to malaria. *Plos Pathogens* 3:647-658.
62. Roetynck S, Baratin M, Johansson S, Lemmers C, Vivier E, Ugolini S (2006) Natural killer cells and malaria. *Immunol Rev* 214:251-263.
63. Stevenson MM, Riley EM (2004) Innate immunity to malaria. *Nature Rev Immunol* 4:169-180.
64. Urban BC, Ing R, Stevenson AM (2005) Early interactions between blood-stage *Plasmodium* parasites and the immune system. *Immunol Immunopath Malar* 297:25-70.

65. McKenzie FE, Bossert WH (2005) An integrated model of *Plasmodium falciparum* dynamics. *J Theor Biol* 232:411-426.
66. Gurarie D, Zimmerman PA, King CH (2006) Dynamic regulation of single- and mixed-species malaria infection: Insights to specific and non-specific mechanisms of control. *J Theor Biol* 240:185-199.
67. Filipe JAN, Riley EM, Drakeley CJ, Sutherland CJ, Ghani AC (2007) Determination of the processes driving the acquisition of immunity to malaria using a mathematical transmission model. *PLoS Comput Biol* 3:e255.
68. Gurarie D, McKenzie FE (2006) Dynamics of immune response and drug resistance in malaria infection. *Malaria Journal* 5.
69. Gurarie D, McKenzie FE (2007) A stochastic model of immune-modulated malaria infection and disease in children. *Math Biosci* 210: 576-597.
70. Kitron U, Spielman A (1989) Suppression of transmission of malaria through source reduction - antianopheline measures applied in Israel, the United States, and Italy. *Rev Infect Dis* 11:391-406.
71. Keiser J, Singer BH, Utzinger J (2005) Reducing the burden of malaria in different eco-epidemiological settings with environmental management: a systematic review. *Lancet Infect Dis* 5:695-708.
72. Chaves LF, Kaneko A, Taleo G, Pascual M, Wilson ML (2008) Malaria transmission pattern resilience to climatic variability is mediated by insecticide treated nets. *Malar J* 7: Accepted
73. Kaneko A, Taleo G, Kalkoa M, Yamar S, Kobayakawa T, Bjorkman A (2000) Malaria eradication on islands. *Lancet* 356:1560-1564.
74. Lengeler C, Sharpe B (2003) Indoor Residual Spraying and Insecticide-treated Nets. Washington: Global Health Council. 17-24 p.
75. Lengeler C (2004) Insecticide-treated bed nets and curtains for preventing malaria. *Cochrane Database Syst Rev*: CD000363.
76. Druilhe P, Barnwell JW (2007) Pre-erythrocytic stage malaria vaccines: time for a change in path. *Curr Opin Microbiol* 10:371-378.
77. O'Meara WP, Hall BF, McKenzie FE (2007) Malaria vaccine efficacy: the difficulty of detecting and diagnosing malaria. *Malaria Journal* 6.

78. Sinclair ARE (1989) The regulation of animal populations. Pages 197–241 in J. M. Cherrett, ed. Ecological concepts. Oxford: Blackwell.
79. Berryman AA, Lima M, Hawkins BA (2002) Population regulation, emergent properties, and a requiem for density dependence. *Oikos* 99:600-606.
80. Maitland K, Williams TN, Bennett S, Newbold CI, Peto TEA, Viji J, Timothy R, Clegg JB, Weatherall DJ, Bowden DK (1996) The interaction between *Plasmodium falciparum* and *P. vivax* in children on Espiritu Santo island, Vanuatu. *Trans Roy Soc Trop Med Hyg* 90:614-620.
81. Maitland K, Williams TN, Newbold CI (1997) *Plasmodium vivax* and *P. falciparum*: Biological interactions and the possibility of cross-species immunity. *Parasitol Today* 13:227-231.
82. Bruce MC, Day KP (2002) Cross-species regulation of malaria parasitaemia in the human host. *Curr Opin Microbiol* 5:431-437.
83. Boyd MF, Kitchen SF (1937) Simultaneous Inoculation with *Plasmodium vivax* and *Plasmodium falciparum*. *Am J Trop Med Hyg.* 17: 855 - 861.
84. Boyd MF, Kitchen SF (1938) Vernal Vivax Activity in Persons Simultaneously Inoculated with *Plasmodium vivax* and *Plasmodium falciparum*. *Am J Trop Med Hyg.* 18:505-514.
85. Collins WE, Jeffery GM (1999) A retrospective examination of sporozoite- and trophozoite-induced infections with *Plasmodium falciparum*: Development of parasitologic and clinical immunity during primary infection. *Am J Trop Med Hyg* 61:4-19.
86. McQueen PG, McKenzie FE (2006) Competition for red blood cells can enhance *Plasmodium vivax* parasitemia in mixed-species malaria infections. *Am J Trop Med Hyg* 75:112-125.
87. Rogier C, Ly AB, Tall A, Cisse B, Trape JF (1999) *Plasmodium falciparum* clinical malaria in Dielmo, a holoendemic area in Senegal: No influence of acquired immunity on initial symptomatology and severity of malaria attacks. *Am J Trop Med Hyg* 60:410-420.
88. Bousema JT, Drakeley CJ, Kihonda J, Hendriks JCM, Akim NIJ, Roeffen W, Sauerwein RW (2007) A longitudinal study of immune responses to *Plasmodium falciparum* sexual stage antigens in Tanzanian adults. *Parasite Immunol* 29:309-317.

89. Maitland K, Williams TN, Peto TEA, Day KP, Clegg JB, Weatherall DJ, Bowden DK (1997) Absence of malaria-specific mortality in children in an area of hyperendemic malaria. *Trans Roy Soc Trop Med Hyg* 91:562-566.

CHAPTER V

RESILIENCE TO THE EFFECTS OF ENVIRONMENTAL VARIABILITY: BEDNETS AND MALARIA IN VANUATU

Introduction

Qualitative changes in the dynamics of populations, or regime shifts, are common phenomena across all living organisms [1,2]. Originally defined in fisheries science [3], the concept that at some time (termed a "breakpoint") there are disturbances that push a biological system beyond its normal dynamical pattern and can qualitatively change its behavior. Recently, this has become a major concern for vector-borne diseases in the context of global climatic change [4,5,6]. Such "breakpoints" derive from ecological analysis that has come to be known as Schmalhausen's law [2] which states that systems at the border of their limits of tolerance to one factor become more sensitive to small changes along any other dimension of its existence [2]. Schmalhausen's law implies that if a system is pushed away from a state of exacerbation, its mean value and variability should decrease. This principle is strongly connected with the idea of resilience [7], the robustness of an ecological system before changing to a qualitatively different state, which in principle should be more susceptible to the effects of climatic variability as populations become less vulnerable to infection [8]. Malaria in the archipelago of Vanuatu historically has been a major public

health problem as shown by the early entomological surveys of Buxton and Hopkins [9], followed by the extensive work of Bastien [10], where an increase in the burden of the disease in the early 1980s was reported [11], as well as its possible association to the evolution of quinine resistant parasites [12,13], numerous studies have shown this disease to be a major burden for Vanuatu inhabitants. Although occasionally hyperendemic in some areas of sub-Saharan Africa, malaria patterns are very different from this region in several aspects. There, the frequency of fatal cases is greatly diminished [14,15], the number of inapparent infections changes seasonally, disease depends on *Plasmodium* species [16], the diversity of parasites is also reduced [17], and the genetic make-up of the native populations presents signatures of evolutionary changes driven by malaria. The latter is expressed in an increased frequency of α -thalassemia associated with mild cases of malaria [18], and an increased frequency of G6PDH enzyme deficiency [19] which is different from sickle cell anemia, the most common one seen in Africa [18,19].

Malaria control efforts also are important to analysis of this time pattern. In 1988, a major control intervention was launched, with a massive distribution of Insecticide Treated Nets (ITNs), following the abandonment of indoor residual spraying for controlling malaria [20]. Although focused studies have demonstrated the use of ITNs to be very effective on small islands of this archipelago, as demonstrated by the elimination of the disease in Aneytium [21], another study analyzing the effects of this policy at the level of the whole country

has not been undertaken. In the present study, we evaluated the dynamics of malaria before and after the introduction of ITNs into the archipelago in an attempt to determine whether there were breakpoints where dynamics shifted transmission patterns, and quantified the effects of climate on these patterns before and after this intervention took effect.

Methods

Malaria data and Monitored Population at risk

Monthly records of malaria were obtained from health centers of people who presented with fever or a recent history of fever, and whose standard blood slide analysis indicated infection with either *Plasmodium vivax* or *P. falciparum*, from January 1983 to December 1999. Malaria cases detected by this passive surveillance were the basis of our analysis. During this period total population increased (Figure 5.1). Data on distributed ITNs with permethrin and re-impregnations were available for the same period (Figure 5.3 A). All data were obtained from the Malaria and other Vector Borne Diseases Control Unit, Ministry of Health, Port Vila, Vanuatu.

This passive case detection system changed in January 1991, as slide examination in small rural health posts was discouraged by the central government of Vanuatu [19]. This policy change reduced the number of people being monitored, however, it remained representative of the whole population [19]. To account for the possible effects of this policy change, we measured changes in the rate of slide examination before and after the breakpoint obtained

for the rate of slide examination, and assumed it to be linearly correlated to changes in the population monitored (see Figure 5.1F). That is, we multiplied the population at risk (corresponding to the population in districts where malaria was present) by the fraction obtained by dividing the average rate of examination before and after the breakpoint to evaluate this possible source of error. We found a 50% reduction in average rate of slide examination during 1990-1991 (Figure 5.1 F), like in [19].

Environmental Data

Weather data included Sea Surface Temperature (SST) indexes: 1+2, 3, 3.4 and 4 (also known as the Niño 1+2, 3, 3.4 and 4; <http://www.cpc.ncep.noaa.gov/data/indices>, Figure 5.1S), and precipitation and temperature data from the climate database for political areas [22, <http://www.tyndall.ac.uk>]. These data were used as predictors in models to assess changes in the magnitude of forcing by climatic variables in the dynamics of malaria incidence.

Statistical Analysis

Breakpoints & Regime Shifts. Tests of structural changes in time series can be undertaken using at least three different strategies: F tests that compare the null hypothesis of no regime shift to the presence of a regime shift, generalized fluctuation tests that do not assume any particular pattern of deviation from the absence of regime shifts [23,24] and adaptive filtering of signals [25]. We used

these three approaches in the present study to assess the robustness of the findings. The F statistic is obtained by comparing the residuals $\hat{\varepsilon}(i)$ of a segmented regression at time i with the residuals $\hat{\varepsilon}$ from an unsegmented regression using the following expression:

$$F_i = \frac{\hat{\varepsilon}^T \hat{\varepsilon} - \hat{\varepsilon}(i)^T \hat{\varepsilon}(i)}{\hat{\varepsilon}(i)^T \hat{\varepsilon}(i) / (n - 2k)} \quad (1)$$

Where n is the time series length and k the number of parameters. The null hypothesis is rejected when the supremum of the statistic is larger than the value of a distribution SupF derived by Hansen [26,27]. This approach has been generalized for l breaks, with arbitrary but fixed l [28,29]; where the number of breaks can be selected using conventional tools for model selection like the Akaike Information Criterion (AIC) [30].

The other two approaches, the generalized fluctuation test and the adaptive filtering, include formal significance tests, yet reveal regime shifts graphically instead of assuming specific types of departure in advance. For the generalized fluctuation test a parametric model is fitted to the data and an empirical process (EFP) is derived that captures the fluctuation either in residuals or parameter estimates [23, 24]. Under the null hypothesis the fluctuations are governed by central limit theorems while under the alternative (regime shifts) the fluctuation is increased [24]. In the present analysis we used the ordinary least squares (OLS) based CUSUM tests introduced in [31]. This test is based in cumulative sums of residuals from a linear regression:

$$W_n^0(t) = \frac{1}{\hat{\sigma} \sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \hat{\varepsilon}_i \quad (0 \leq t \leq 1) \quad (2)$$

where a regime shift is evidenced by a single peak around the breakpoint, provided that the limiting process for $W_n^0(t)$ is the standard Brownian bridge $W^0(t) = W(t) - tW(1)$, where $W(\cdot)$ denotes Brownian motion. Significance for the CUSUM was tested using the derivations presented in [23, 24]. For equation (1) and (2) the residuals $\hat{\varepsilon}$ came from a linear seasonal autoregressive [30] model fitted by OLS:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_{12} y_{t-12} + \varepsilon_t \quad (3)$$

The third approach is totally non-parametric, and is based on recovering a signal and its breaks. The Kolmogorov-Zurbenko adaptive filter (KZAF) [25] is based on filtering the time series y using:

$$z_t = \frac{1}{q_H(t) + q_T(t)} \sum_{i=-q_T(t)}^{q_H(t)} x_{t+i} \quad (4)$$

Where

$$q_H(t) = \begin{cases} q & \text{If } D'(t) < 0 \\ f(D(t))q & \text{If } D'(t) \geq 0 \end{cases} \quad (5)$$

$$q_T(t) = \begin{cases} q & \text{If } D'(t) > 0 \\ f(D(t))q & \text{If } D'(t) \leq 0 \end{cases}$$

And q is half-length of a k iterative moving average (x_t) applied to the original time series y_t . The term $f(D(t))$ is defined by:

$$f(D(t)) = 1 - \frac{D(t)}{\max[D(t)]} \quad (6)$$

And $D(t)$ is the absolute difference defined by:

$$D(t) = |x_{t+q} - x_{t-q}| \quad (7)$$

And $D'(t)$ as:

$$D'(t) = D(t+1) - D(t) \quad (8)$$

Once z_t is obtained quantitative estimates of discontinuity can be based on an analysis of the sample variances of z_t , defined by:

$$\hat{\sigma}_i^2 = \frac{\sum_{t=q_T}^{q_H} \{z_t - \bar{z}\}^2}{q_T + q_H} \quad (9)$$

When there are no breaks, maxima in the estimated variance of (10) are approximately independent and exponentially distributed with a expected number of peaks of about $n/(2qk^{0.5})$, allowing to consider a breakpoint when the $\hat{\sigma}_i^2$ value exceeds the 95% upper tail of the exponential distribution with such parameter.

Regime shift analyses were carried out on: (i) the monthly rate of slide examination (No. Slides examined*1000/Total population at risk); (ii) the monthly rate of the two malaria parasites (No. slides examined*1000/Monitored population at risk) and (iii) weather variables (rainfall and temperature).

Threshold for ITN coverage: Time series for total number of bednets distributed per month were accumulated and divided by the total population at risk estimated from the annual population data. We assumed that the annual data corresponded to December, and interpolated the rest of the months using a smoothing splines regression as explained in [32]. We also studied the probability density [30] of the percentage of people locally covered with the distributed ITNs.

Seasonality: the seasonality of vivax and falciparum malaria rates (cases/population size) were assessed by using box diagrams before and after the regime shift [30].

Non-stationary patterns of association: The wavelet transform can be used to study the patterns of association between two nonstationary time series [33,34]. Specifically, the wavelet coherency analysis can determine whether the presence of a particular frequency at a given time in the disease corresponds to the presence of that same frequency at the same time in a covariate (e.g., Rainfall and Temperature). The cross-wavelet phase analysis can determine the time lag separating these two series as well.

Changes in the effects of climate on the dynamics: Once breakpoints for the regime shift were identified in the falciparum and vivax malaria rate series, the splitted series around the breakpoints were studied using seasonal autoregressive (SAR) models [30]. The procedure for model building was similar to the one described in [34]: (i) a null model was fitted to the rate of the falciparum and vivax malaria (ii) temperature and rainfall were filtered with the coefficients of the null model, and (iii) cross-correlation functions were computed using the residuals of the null model and those of the filtered climatic variables.

The full model for *P. falciparum* considered precipitation (P) with lags of 2 and 29 months, and temperature (T) with lags of 3 and 12 months, as follows:

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-12} - \mu) + \beta_1 P_{t-2} + \beta_2 P_{t-29} + \alpha_1 T_{t-3} + \alpha_2 T_{t-12} + \varepsilon_t \quad (10)$$

For *P. vivax* the full model considered precipitation (P) a lag 9 months, and temperature (T) with a lag 10 months, as follows:

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-12} - \mu) + \beta_1 P_{t-9} + \alpha_1 T_{t-10} + \varepsilon_t \quad (11)$$

In both cases the error was assumed as independent and normally distributed: $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. After the initial fitting models were simplified using a process of backward elimination[34]: (i) taking out one predictor at a time, (ii) finding the minimum AIC for models with similar complexity, i.e., number of parameters, (iii) comparing the likelihood of the best model (minimum AIC) for each level of complexity with the full model, and simplifying while differences were not statistically significant. For the analyses the climatic covariates were demeaned in order to not affect the intercept value [30].

Results

Temporal patterns of Malaria in Vanuatu present a clear shift in the incidence rate by the end of 1993 and beginning of 1994, for both parasite species (Figure 5.2 A,D).

Breakpoints were confirmed by all three different methods (Figure 5.2S). For the incidence rate in both malaria species, breakpoints were statistically significant according to the F statistic and the variance of the KZAF. Even though the EFP estimates were not significant, peaks were detectable in both cases in January 1992 (Figure 5.3S). During that same time period no significant changes were found for climatic time series (Figure 5.4S). By the time changes were detected, Bednet Coverage (Figure 5.3 B) was as low as 6% (EFP estimate) or slightly above 20% of the population at risk (KZFA). At a more local scale, villages where bednets were distributed mostly had ~80% of the population covered (Figure 5.3 C).

Plasmodium falciparum seasonality was qualitatively very similar before and after the breakpoint (Figure 5.2 B,C), showing maximum incidence during the first quarter of the year (January-March), and minimum incidence during the third quarter of the year (July-September). For *P. vivax* (Figure 5.2 E,F) a similar change was observed, although the patterns were not so clear as for *P. falciparum*, due to greater seasonal variability. With the exception of a brief period during 1992-1996 where cases due to both parasites were synchronous (i.e., with peaks at the same time), the dynamics of the infections were mainly asynchronous and not coherent (i.e., not associated in the frequency domain) at the seasonal scale.

However, both diseases were significantly cross-coherent at an interannual scale, with the dynamics of *P. falciparum* cases being mostly synchronous with that of *P. vivax* (Figure 5.5S).

Regarding the effects of climate during the studied period, the cross-coherence wavelet analysis showed malaria to be correlated with temperature at the seasonal scale; both *P. falciparum* and *P. vivax* incidence rates were led by temperature (Figure 5.4). A similar pattern was seen between the two parasites and rainfall at the seasonal scale, despite the presence of some gaps. A significant coherence with rainfall at interannual scales was also found. For *P. vivax*, coherence was statistically significant for periods between 2 and 4 years,

during 1992-1996. No evidence that El Niño indices were leading the dynamics of the disease was identified.

Finally, Table 5.1 presents the parameter estimates for rate models of *P. falciparum* and *P. vivax*, including exogenous forcing by temperature, before and after the breakpoint. Model selection by backward elimination showed that rainfall was not a significant covariate (detailed values in Appendix S3). Following the qualitative change in the dynamics, *P. falciparum* had a proportional (~66%) and absolute (7.7 cases/1000 population) decline in incidence that was greater than that for *P. vivax* (~52% and 2.6 cases/1000 population, respectively). The importance of temperature in driving the dynamics also declined after the breakpoint for both species, between 31% and 49% for *P. falciparum* and 80% for *P. vivax* (the coefficient after the breakpoint became statistically non-significant). This suggests that the average effect of 1°C increase in temperature will increase incidence in a reduced amount when compared with its effect before the breakpoint. For example, preceding the shift each degree Celsius above the 3-month lagged mean temperature value used to increase the rate by 1.43 cases/1000 people at risk. In contrast, after the breakpoint this change only increased the rate by half of its previous magnitude, i.e., 0.72/1000 people at risk (Table 5.1). A similar phenomenon was also seen for the variability that was not explained by the models, which also was reduced by 33% and 54% in *P. vivax* and *P. falciparum*, respectively, as shown by the decrease in the error variance of the models after the breakpoint (Table 5.1).

Discussion

Following a disturbance, biological systems can either return to their normal state of variability or can move far away from such a state [1, 2, 35, 36]. Transients, i.e., the anomalous behavior between regimes or basins [37, 38], can obscure the qualitative changes of a system, because jumps from one state to another are not always instantaneous, complicating our ability to identify regime shifts [37,39]. This is likely one of the main differences between the dynamics of *P. falciparum* and *P. vivax*, since a consistent estimate for the breakpoint was easy to find for the former, while the estimates for the latter differed significantly. This was especially true for KZAF, which identified a later breakpoint. Assumptions underlying the employed techniques [23-31] might favor the estimate from KZAF, since the F statistic is quite sensitive to the stationarity (i.e., constant mean) of the time series, while the CUSUM EFP may be too sensitive given the quality of the data examined, identifying the change of policy in slide examination. By contrast, the KZAF is an adaptive technique that allows control of the time scale at which changes may be occurring [25]. This a very useful characteristic for addressing one of the major recurrent problems in the study of ecological systems, i.e. finding the appropriate temporal scale of a natural phenomenon [40]. In this study, the adaptive ability of KZAF allowed for breaks to be distinguished from natural cycles associated with exogenous factors (i.e., climate). The fact that the basin (or regime) shift in the time series can be attributed to the effects of bednet use appears robust. During the study period no other major changes in control strategies, landscape cover, medication or drug

resistance were reported [10, 11, 19] after controlling for the policy change in data collection [19].

Our analysis identified a major difference between *P. falciparum* and *P. vivax*, namely the earlier breakpoint for *P. falciparum*. This pattern would not be expected under conditions of cross or heterologous immunity [41], and its evaluation with cross-infection studies is limited because quality data that are necessary to make such inferences [42] are lacking [19,21]. However, this pattern should be studied further, because it might reflect the dynamics of immunity in the population, where a generalized density-dependent immunity may be triggered by the within-host density of each parasite species [43]. Alternatively, if *P. falciparum* was the first species to be cleared, as shown in the classical co-infection neuro-syphilis malariotherapy experiments of Boyd and Kitchen [44], temporal patterns can only be appreciated when studying the dynamics of the within-host parasitic infection [45]. In addition, the pattern simply could arise by the ability of *P. vivax* to relapse [19, 21], possibly in conjunction with the immunity dynamics described above.

Although regime shifts tend to be thought of in terms of increased variability as the best diagnostic condition [46], they can occur in the opposite direction, with systems becoming more stable. For both *P. falciparum* and *P. vivax* not only did the mean value of incidence decrease, but also the variance of the models decreased, which is a more robust measure of stability [36] than just looking at mean values [1] in dynamical systems. The patterns seen for the two species differed: *falciparum* malaria declined more abruptly, in total and relative

terms, than in vivax malaria. Perhaps there are differences in the life history strategies of the parasites under different scenarios for transmission, with the most virulent parasite (*P. falciparum*) being more successful in environments with high transmission rates and the least virulent (*P. vivax*) being less sensitive to the intensity of transmission.

A surprising result was that the breakpoint occurred after just 20% of the population was covered with bednets, which is half that predicted for *Anopheles gambiae* transmission by Killeen et al [47]. Perhaps *An. farauti*, the main vector in Vanuatu [9,48] is less efficient. Regardless, the fact that such ITN coverage could explain the decrease has a robust theoretical explanation as presented in the groundbreaking work of Becker and Dietz [49], later confirmed using field data as the 80/20 rule for several infectious diseases [50,51] where the control, which targets 20% of the population, could benefit the other 80% of people.

Interestingly, this rule has been derived by looking at local populations, but the pattern seen in Vanuatu is more likely to arise from the subdivided nature of the population in villages, or patches if seen from the perspective of metapopulations [52]. The coverage per patch was high enough (80% with a very low dispersion around this value) to guarantee the local interruption of transmission according to mechanistic models of bednet action in settings with a higher entomological inoculation rate [47,53] than that observed in Vanuatu [16, 48].

As a control strategy, ITNs outperform similar strategies aimed at reducing vectorial capacity, such as the indoor residual spraying, mainly because of its

cost-effectiveness, as well as for its ease of implementation and distribution [54, 55]. Several studies have shown that bednets reduce total infant mortality in endemic areas [56, 57], are a sustainable option for control in terms of the reduction of relative risk of malaria death in the medium- to long-term time scales [58], and are successful across several cultural settings [55,59-62]. The advantages of bednets also go beyond the immediate effects, since so far there is no evidence for selection of insecticide-resistant mosquitoes [63], and they are protective even in areas where mosquito resistance to the insecticides used for bednet impregnation has been reported [64]. This result also has been theoretically reinforced by models that consider the use of bednets in conjunction with other control strategies, such as zooprophylaxis [65], provided that both measures in conjunction are likely to counteract any selective pressure for the development of insecticide resistance, since mosquito fitness would not be under a selective pressure, and may even be under selection for feeding preferences in non-human hosts [66,67]. However, urban settings pose a major challenge since effective zooprophylaxis might be diminished because of higher human densities. Behavioral changes in mosquitoes and decreased bednet effectiveness have been documented in urban areas [68]. From a wider perspective, bednets are also a more ecologically-sound strategy since they reduce impacts on natural enemies of vectors via positive feedbacks loops that can be generated by large scale insecticide spraying [66,69,70]. A large body of literature supports that idea that in relatively undisturbed environments mosquito abundance is regulated by interactions with other animals, e.g., tadpoles, fish and other insects [e.g., 69-74],

however such natural control is diminished anthropogenic disturbances of food webs.

The success of the Vanuatu malaria control program also stems from the strategy of bednet distribution, where large fractions of the population were locally covered at the village level, ensuring the reduction in transmission, even leading to local elimination in some islands [21]. As stressed by Killeen et al [47] and Ilboudo-Sanogo et al [63], an efficient bednet program needs to cover a large proportion of the population in order to ensure that both sources [e.g., asymptomatic people] and sinks [e.g., pregnant women and young children] of infection are effectively covered. The erroneous targeting of transmission groups for control can exacerbate the conditions for transmission [75]. Additionally, as suggested by Mathanga et al [76], for ethical and humanitarian reasons the goal should be to cover as much of the population present in the endemic setting as possible, retaining traditional practices (e.g., voluntary work) for the exchange of goods when mainstream means of commercialization are not enough to achieve such a goal. In Vanuatu, special care was taken to address these factors by implementing a strategy where children under 5 years old, their mothers and pregnant women received free nets. Cost was half price for school children and other adults were charged the full price, ensuring an equitable coverage of the population [21] and an equitable distribution of this valued resource.

A factor that deserves further study is the role that concomitant knowledge transfer associated to the distribution of bednets have on the awareness of the population about the risk leading to malaria transmission. Unlike insecticide

residual spraying whose effectiveness depends mostly on being applied correctly, the effective use of bednets requires knowledge for its proper use. In Vanuatu, parents' awareness was likely to play a role in diminishing incidence among young children (<5 years), because of the free distribution to this age group and training to parents about the benefits of using the nets [21]. But, the positive effects of knowledge transfer are likely to be more comprehensive. For example, Mathanga et al [76] showed that even though children didn't regularly use bednets, those in communities where malaria transmission plummeted after the introduction of widespread bednet use were aware of the benefits. Similar knowledge transfers are known to be present among some Native American tribes whose mythology has associated malaria risk with the blossoming of water-retaining flowers where vector larvae develop [77]. Changes in collective behavior in villages that were stricken by malaria have been seen before community-based educational campaigns were implemented [78-80] and more generally, traditional knowledge has been shown to be a robust strategy to handle issues of pest management by native populations in Meso-America [81].

The association between climatic forces and malaria dynamics in Vanuatu presents features that make it unusual when compared to other settings where the climate and ecological dynamics have been studied [e.g., 34, reviewed in 82]. None of the ENSO indices led the dynamics of malaria, yet clear signals of association at interannual time scales were found with local climatic variables. This may be a result of the relationship of ENSO with the local climate in the area [83], and less likely because of a demographic effect of small insular population

size as suggested in [84]. Mechanisms for the action of rainfall across a wide range of landscapes have been very well described, it increases the rate of a disease when new mosquito habitats are created by increased precipitation [85], and the additional weakening of inter-specific interactions regulating mosquito populations [86]. We may similarly understand why hotter temperatures can increase the transmission of vector-borne diseases, because of known effects of temperature on the rate of insect and parasite development [83,87]. However, increased resilience to the effects of climate in an infectious disease as a result of control measures, in our knowledge, has not been reported before. The fact that such a measure also decreases the incidence of malaria under changing climatic conditions is a remarkable fact strengthening the usefulness of this strategy.

Finally, a precautionary note on bednets should be posed. Even though they are a very robust strategy to control malaria from evolutionary, ecological, conservation and cost-effectiveness perspectives [54,55,63,76], the use of bednets should not be viewed as an exhaustive solution if the long-term goal of population health is to be pursued. As shown in [88] a fraction of the death toll that was avoided by controlling malaria through the use of insecticide treated curtains in areas of Burkina Faso was shifted to meningococcal meningitis. Evidence also suggests that in urban settings, for a series of factors that go from the absence of alternative hosts to behavioral shifts in humans, insecticide treated nets are not going to be a sufficient strategy to keep malaria under control [68]. To achieve this goal, a wide research agenda, fully integrated with

policies beyond disease control is a path that needs to be taken [32,89-92], where ultimate goals are aimed at pushing out the stressful contextual conditions that make human populations vulnerable to infectious diseases [2], especially malaria.

Table 5.1 Parameter values and % reduction for *Plasmodium falciparum* and *Plasmodium vivax* rate before and after the breakpoint obtained by using the variance of the Kolmogorov Zurbenko adaptive filter.

Species	Parameter	Before	After	% Reduction	P
<i>P.falciparum</i>	$\hat{\mu}$	11.56 ± 1.02	3.90 ± 1.15	66.26	B/A
	$\hat{\alpha}_1$	1.43 ± 0.42	0.72 ± 0.35	48.59	B/A
	$\hat{\alpha}_2$	1.36 ± 0.44	0.94 ± 0.33	30.88	B/A
	$\hat{\sigma}_\varepsilon^2$	5.76	3.84	33.31	-
<i>P.vivax</i>	$\hat{\mu}$	4.83 ± 0.62	2.33 ± 0.43	51.76	B/A
	$\hat{\alpha}_1$	0.55 ± 0.18	0.11 ± 0.16	80.00	B
	$\hat{\sigma}_\varepsilon^2$	1.17	0.53	54.31	-

. % Reduction is defined as 1- (parameter value before breakpoint/ parameter value after breakpoint). For *P. falciparum* the final model was:

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \phi_{12}(y_{t-12} - \mu) + \alpha_1 T_{t-3} + \alpha_2 T_{t-12} + \varepsilon_t \text{ and for } P.vivax:$$

$y_t = \mu + \phi_1(y_{t-1} - \mu) + \phi_{12}(y_{t-12} - \mu) + \alpha_1 T_{t-10} + \varepsilon_t$ Model selection process and all parameter values can be seen in Table S1. Column P (<0.05) indicates the significance of any parameter B (before breakpoint) / A(after breakpoint)

Figure 5.1. Time Series: **A** *Plasmodium falciparum* malaria cases, **B** *P. vivax* malaria cases, **C** Temperature (°C), **D** Precipitation (mm), **E** Population at risk (solid), **F** Monthly slide examination rate (slides*1000/population at risk), the dashed line corresponds to the breakpoint, August 1990, estimated using the F statistic, and the solid lines at the bottom of the graph to the confidence intervals (February 1990, May 1992) the thick-black solid line is the Kolmogorov-Zurbenko adaptive filter implemented with a half window size, q , of 36 months, the breakpoint is December 1991, the blue line corresponds to the breakpoint obtained using the CUSUM (march 1990). The mean rate (\pm S.D.) of slide examination before the breakpoint (August 1990) was (51.22 ± 11.40) being reduced to (25.92 ± 11.56) after it. Statistical tests of significance can be seen in Figure 5.2S

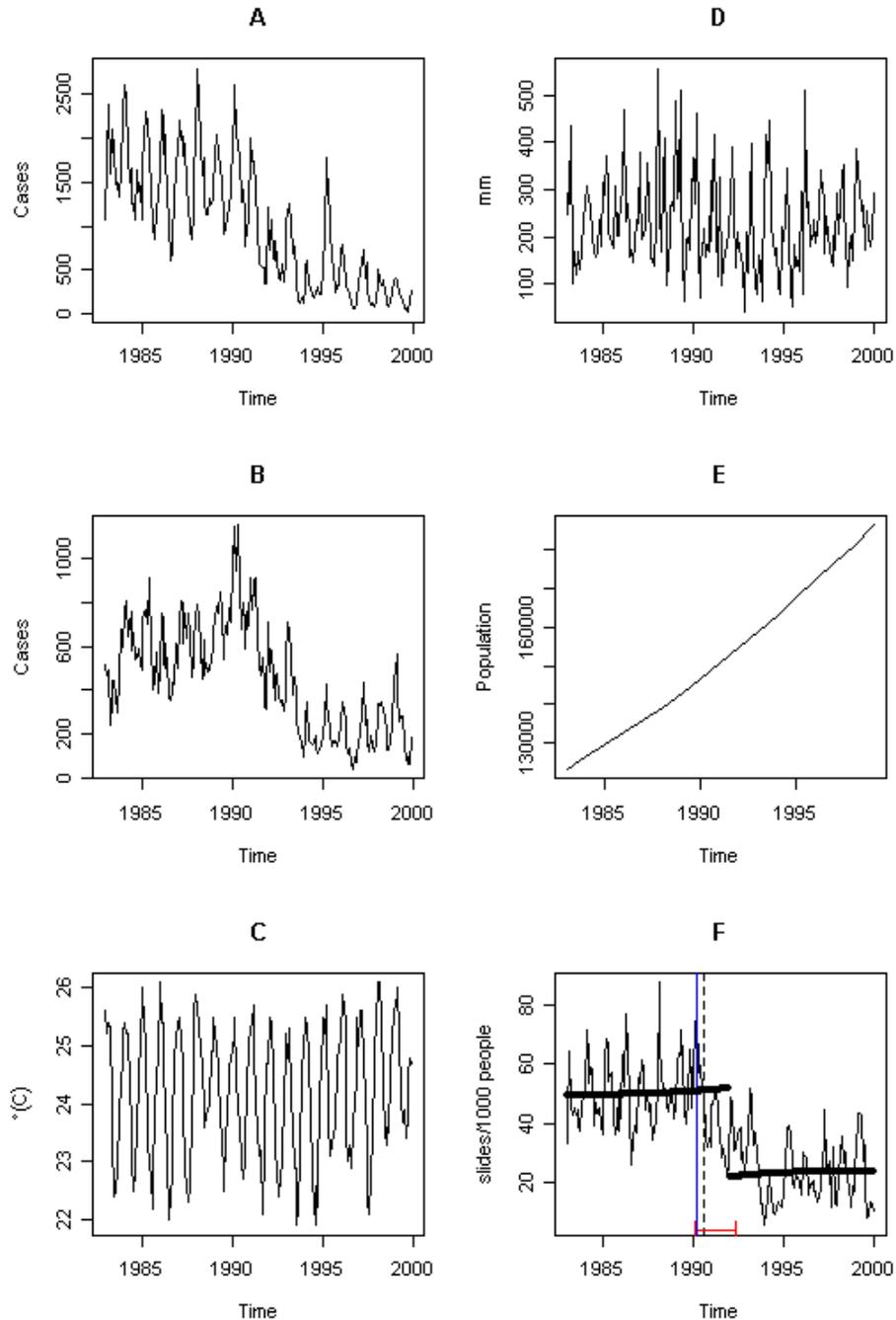


Figure 5.2. Regime Shift for falciparum and vivax malaria: **A** falciparum malaria rate, the dashed line corresponds to the breakpoint, January 1992, estimated using the F statistic, and the solid lines at the bottom of the graph to the confidence intervals (June 1989, June 1994), the thick-black solid line is the Kolmogorov-Zurbenko adaptive filter implemented with a half window size, q , of 36 months, the breakpoint corresponds to August 1993, the blue line corresponds to the breakpoint obtained using the CUSUM (January 1992). **B & C** seasonal falciparum malaria rate before and after breakpoint (January 1992) **D** vivax malaria rate, the dashed line corresponds to the breakpoint, May 1991, estimated using the F statistic, and the solid lines at the bottom of the graph to the confidence intervals (June 1989, November 1992), the black solid line is the Kolmogorov-Zurbenko adaptive filter implemented with a half window size, q , of 36 months, the breakpoint corresponds to February 1994, the blue line corresponds to the breakpoint obtained using the CUSUM (January 1992) **E & F** seasonal vivax malaria rate before and after breakpoint. For the F statistics the 30% percent of the data belonging to the extremes (15% each) was left out. For the Kolmogorov-Zurbenko adaptive filter q was set to 36, in order to avoid the misidentification of cycles shorter than 6 years.

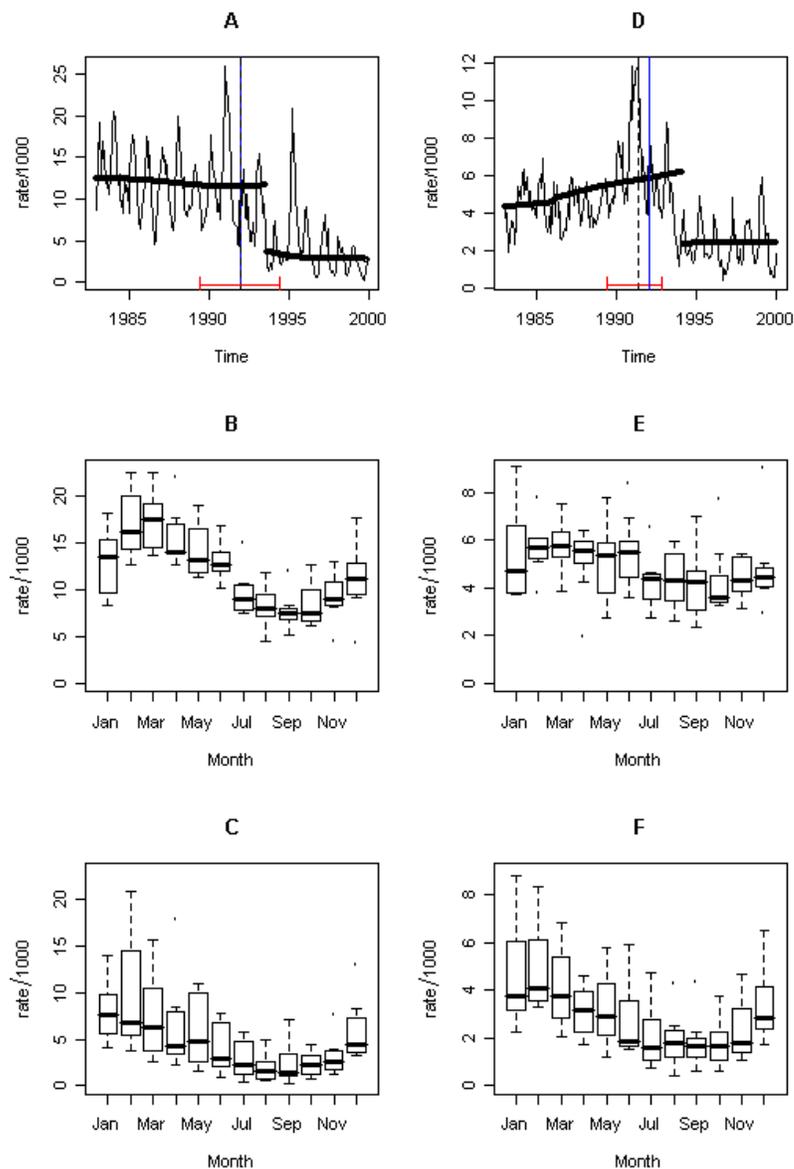


Figure 5.3 Bednets **A** Monthly number of distributed bednets (black line) and number Re-impregnated bednets (green line) **B** Probability density of the percentage of people locally covered with bednets between 1988 and 1997, bandwidth of 0.027 **C** Percent (%) of population covered by bednets for the lower and upper time limit for the breakpoints, the green-blue line corresponds to January 1992 (*Plasmodium falciparum* and *P. vivax*), the green line to September 1992 (*P. falciparum*) and the blue line to December 1992 (*P. vivax*)

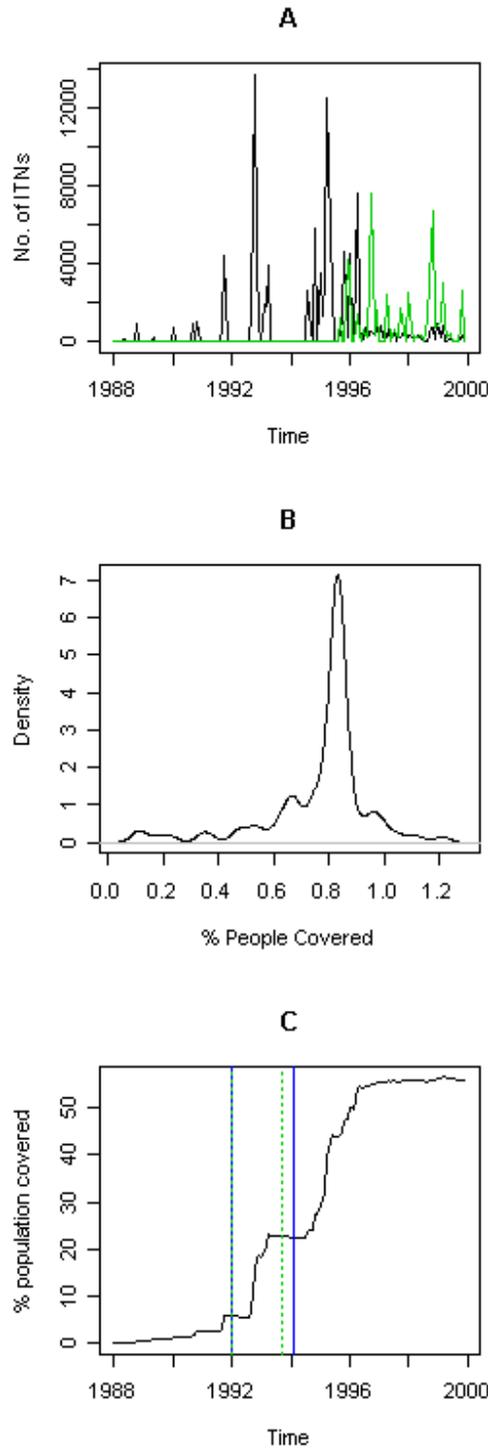


Figure 5.4 Cross-Wavelet Coherency and Phase of *Plasmodium falciparum* malaria rate with **A** temperature and **B** rainfall and of *P. vivax* malaria rate with **C** temperature and **D** rainfall. The coherency scale is from zero (blue) to one (red). Red regions in the upper part of the plots indicate frequencies and times for which the two series share variability. The cone of influence (within which results are not influenced by the edges of the data) and the significant ($p < 0.05$) coherent time-frequency regions are indicated by solid lines. The colors in the phase plots correspond to different lags between the variability in the two series for a given time and frequency, measured in angles from $-\pi$ to π . A value of π corresponds to a lag of 17 mo. The procedures and software are those described in [31,32]. A smoothing window of 15 mo ($2w + 1 = 31$) was used to compute the cross-wavelet coherence.

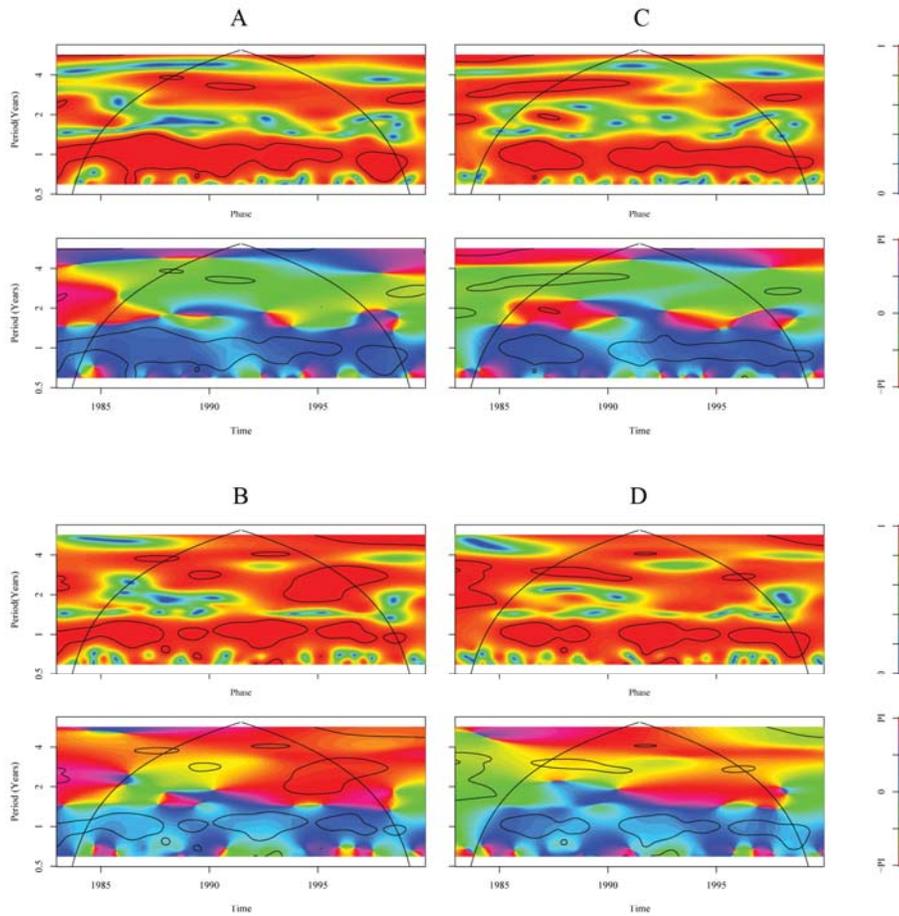


Figure 5.S1. Time Series for The El Niño Southern Oscillation: **A** SST 1+2, **B** SST 3, **C** SST 3.4, **D** SST 4

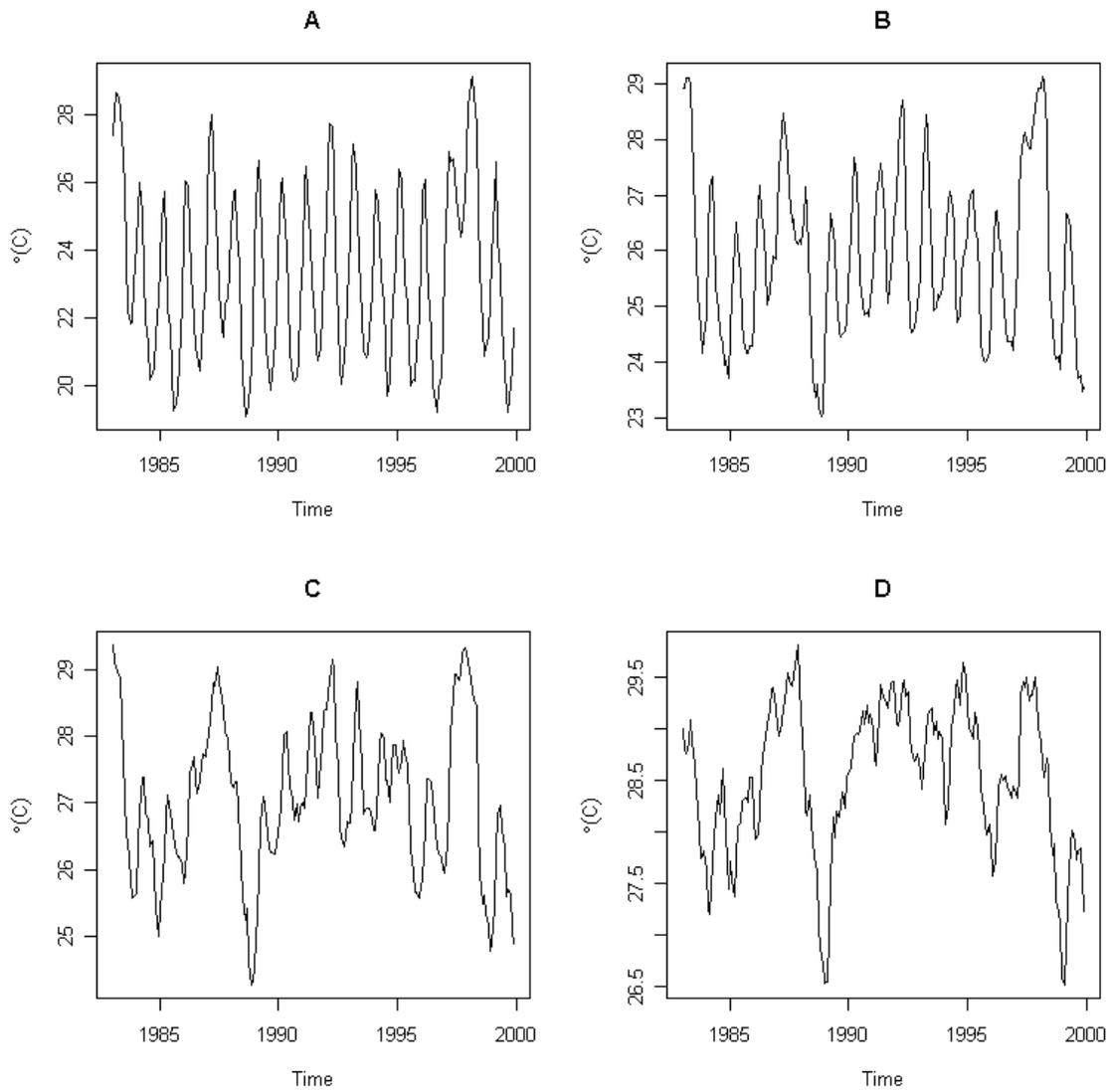


Figure 5.S2 Breakpoints for the rate of slide examination **A** F statistic for the falciparum malaria rate, the solid line is the 95% upper tail of the distribution for the F Statistic [24,25] **B** Empirical fluctuation period of the CUSUM test, the maximum value is in may 1992, the redline is the threshold value for breakpoint significance [23, 24] **C** Variance of the Kolmogorov-Zurbenko adaptive filter, the dashed line is the 95% upper tail of the exponential distribution for this statistic, the maximum value corresponds to may 1992.

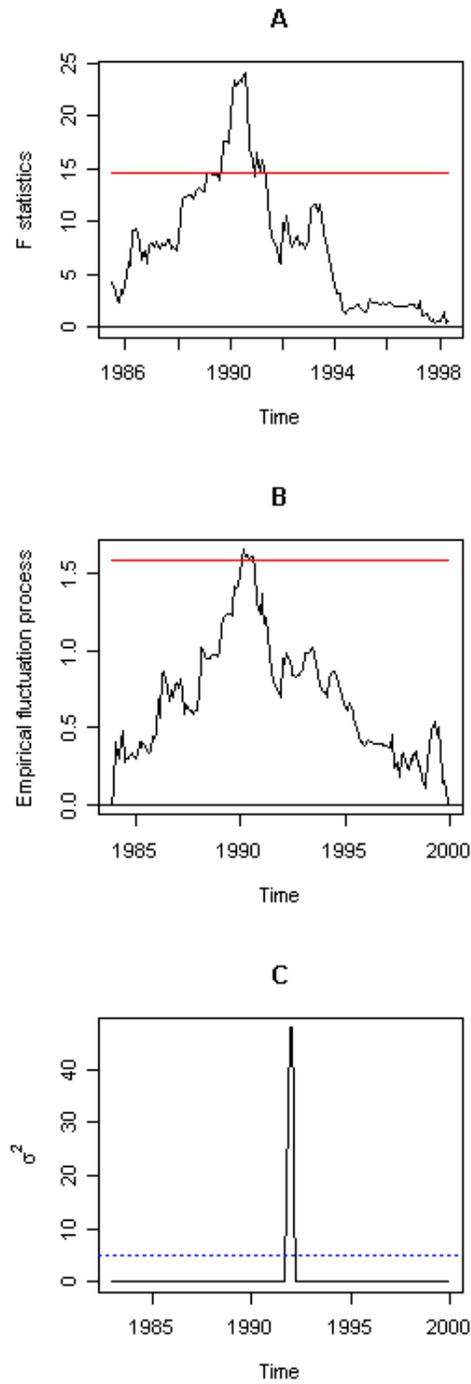


Figure 5.S3 Breakpoints for the *Plasmodium falciparum* and *P. vivax* rates **A** F statistic for the falciparum malaria rate, the solid line is the 95% upper tail of the distribution for the F Statistic [24,25] **B** Empirical fluctuation period of the CUSUM test, the maximum value is in may 1992 **C** Variance of the Kolmogorov-Zurbenko adaptive filter, the dashed line is the 95% upper tail of the exponential distribution for this statistic, the maximum value corresponds to may 1992 **D** F statistic for the vivax malaria rate, the solid line is the 95% upper tail of the distribution for the F Statistic [24,25] **E** Empirical fluctuation period of the CUSUM test for, the maximum value is in January 1992 **F** Variance of the Kolmogorov-Zurbenko adaptive filter. In **A**, **B** and **D**, **E** the redline is the threshold value for breakpoint significance [23, 24]. In **C** and **F** the blue dashed line is the 95% upper tail of the exponential distribution for this statistic, the maximum value corresponds to December 1992.

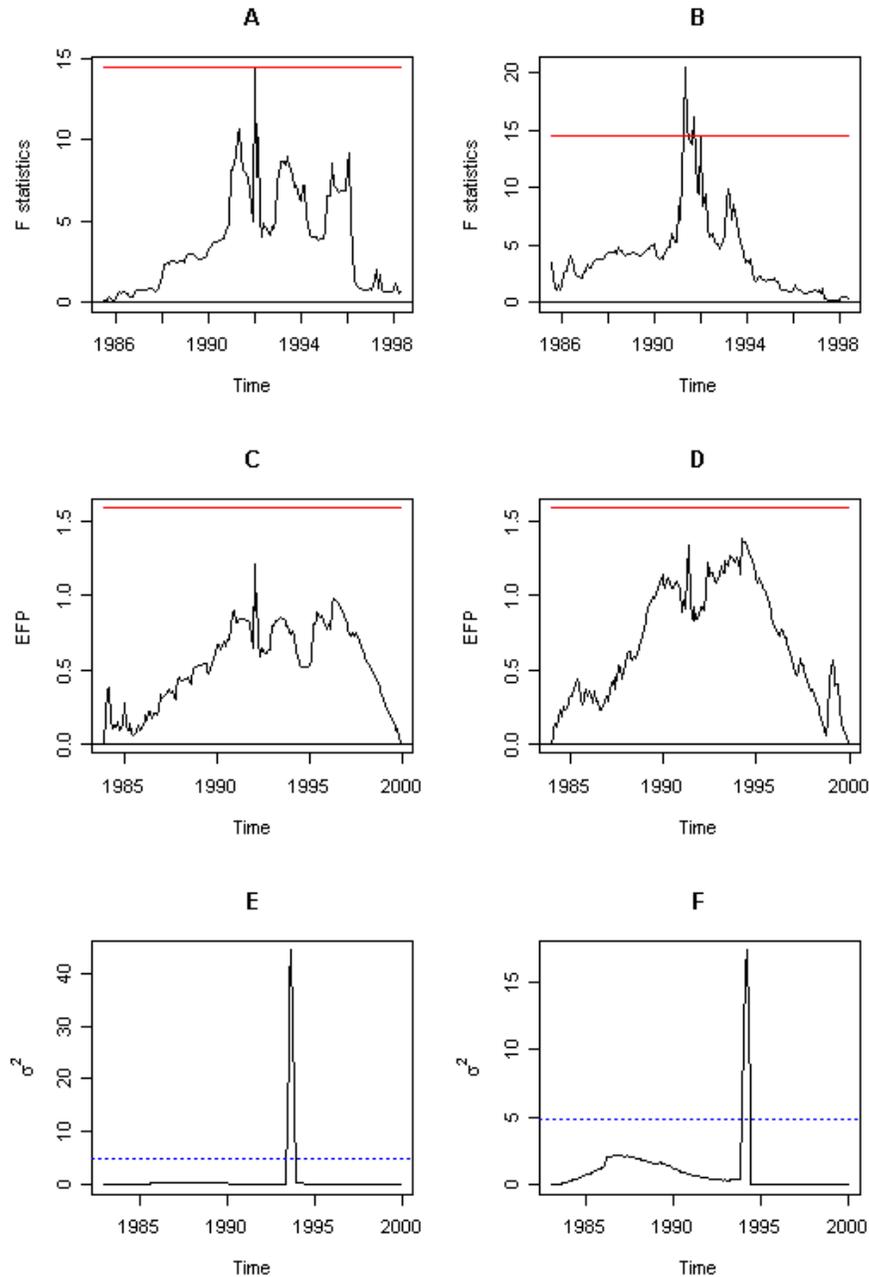


Figure 5.S4. Breakpoints for **A,B** Temperature and **C,D** Rainfall in Vanuatu using the F statistics(**A,C**) and the empirical fluctuation period of the CUSUM (**B,D**). For the F statistics the 30% percent of the data belonging to the extremes was left out.

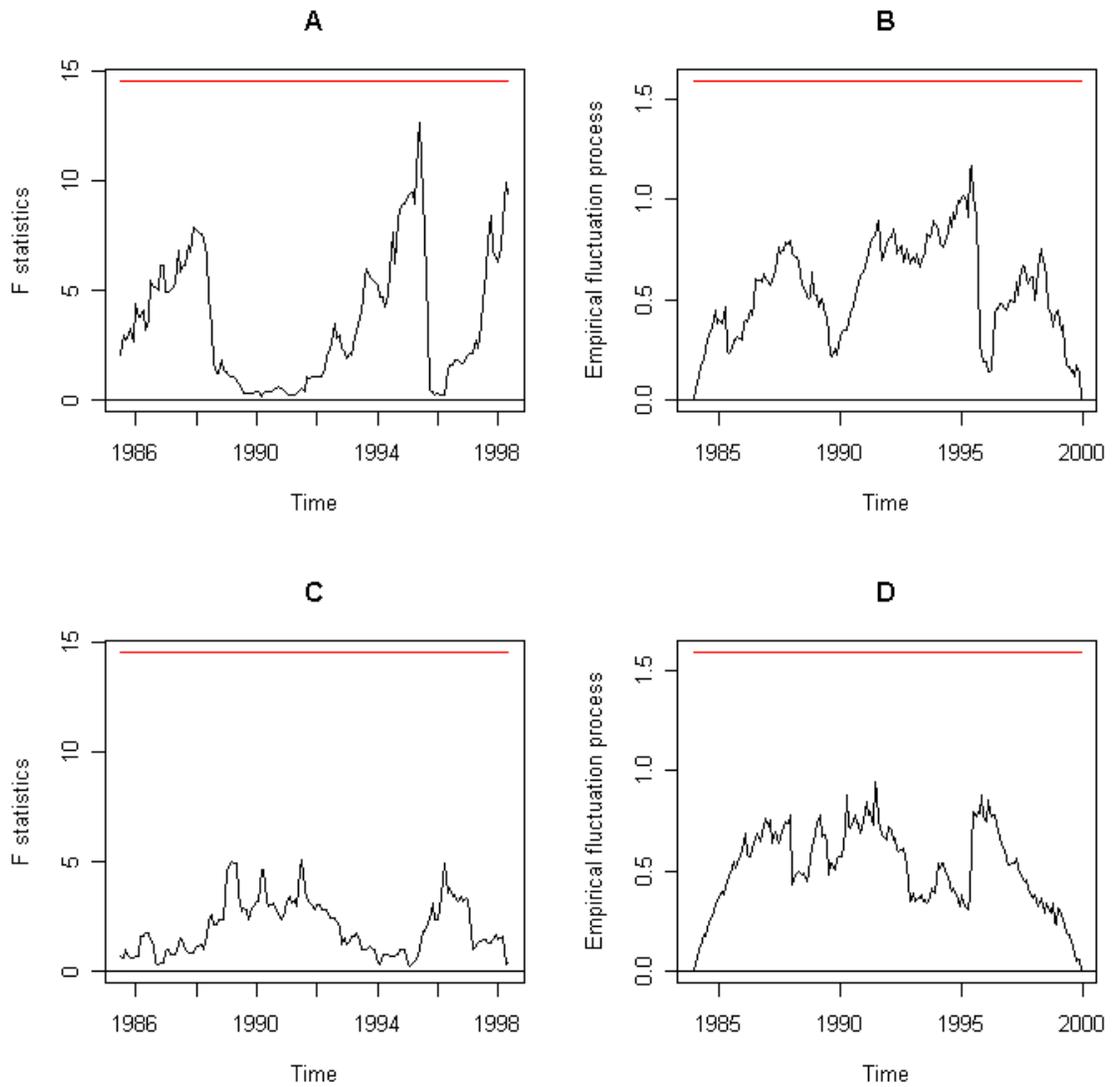
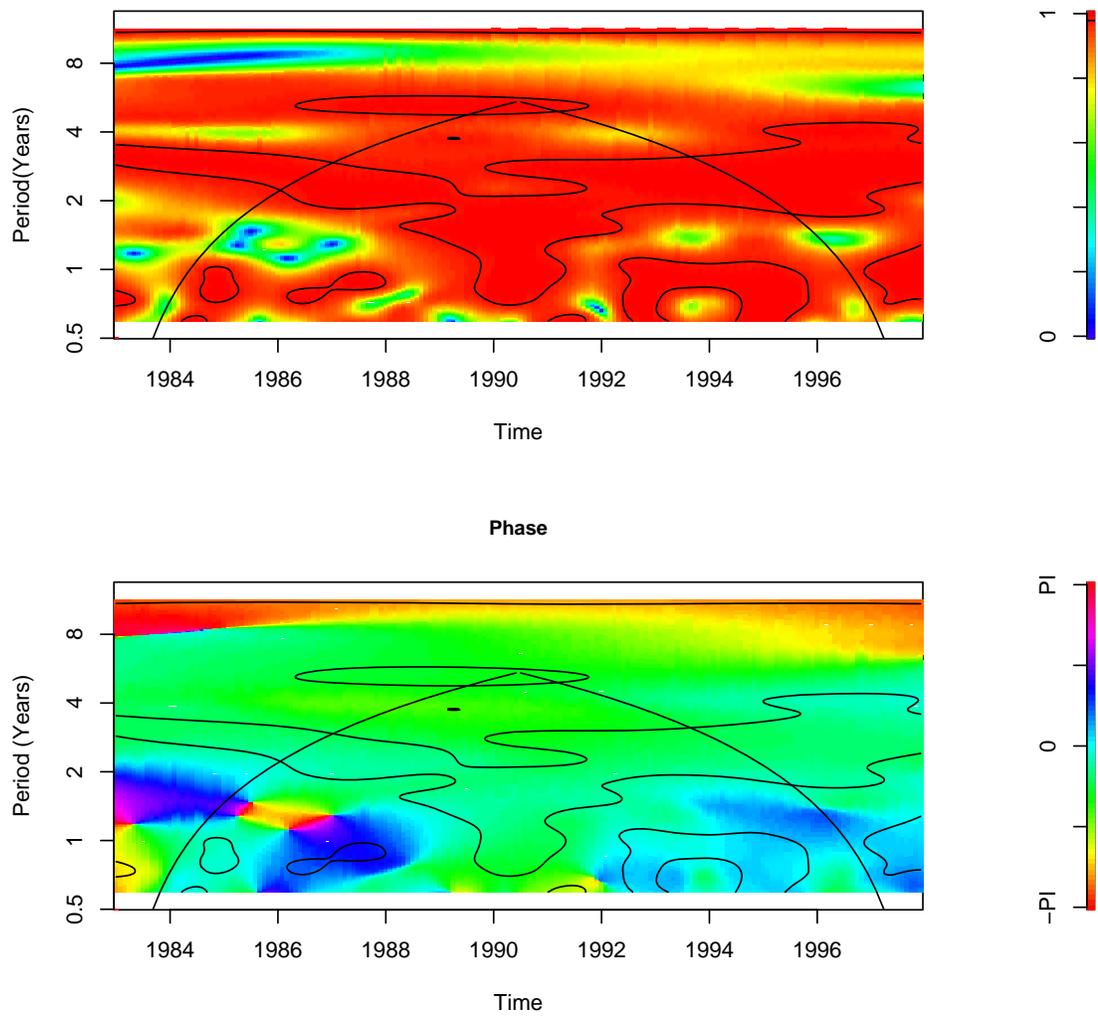


Figure 5.S5 Cross-Wavelet Coherency and Phase of *Plasmodium vivax* malaria rate with *P. falciparum* malaria rate. For technical details see legend of Figure 5.4.



References

1. May RM (1977) Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature*. 269:471-477.
2. Awerbuch T, Kiszewski AE, Levins R (2002) Surprise, nonlinearity and Complex behaviour. In: Martens P, McMichael AJ Editors. *Environmental Change, Climate and Health*. Cambridge University Press. pp 96-119.
3. Ricker WE (1963) Big effects from small causes: 2 examples from fish population dynamics. *J Fish Res Board Can* 20: 257-264.
4. Pascual M, Ahumada JA, Chaves LF, Rodo X, Bouma M (2006) Malaria resurgence in the East African highlands: temperature trends revisited. *Proc Natl Acad Sci USA* 103: 5829-5834.
5. Chaves LF, Pascual M (2007) Comparing early warning systems for neglected tropical diseases. *PLoS Negl. Trop. Dis.* 1: e33.
6. Zhou XN, Yang JG, Yang K, Wang XH, Hong QB, Sun LP, Malone JB, Kristensen TK, Berquist NR, Utzinger J (2008) Potential Impact of Climate Change on Schistosomiasis Transmission in China. *Am J Trop Med. Hyg.* 78: 188-194.
7. Holling CS (1973) Resilience and stability of ecological systems. *Ann Rev Ecol Sys* 4: 1-23
8. Levins R, Lopez C (1999) Toward an ecosocial view of health. *Int. J. Health Serv.* 29: 261-293.
9. Buxton PA, Hopkins GHE (1926) *Researches in Polynesia and Melanesia. An Account of investigations in Samoa, Tonga, The Ellice group and The New Hebrides, in 1924, 1925. Parts I-IV (relating principally to medical entomology)*. London: Memoir of the London School of Tropical Medicine and Hygiene.
10. Bastien P (1990) Public-health epidemiology in Vanuatu. *Med. J. Australia.* 152:13-17.
11. Bastien P, Saliou P, Gentilini M (1986) Le paludisme à Vanuatu: données épidémiologiques récentes. *Bull Soc Path Exot* 79: 476-489.
12. Bastien P (1987) Quinine resistant falciparum malaria in Vanuatu? A case report. *Southeast Asian J Trop Med Pub Health* 18:101-102.

13. Bastien P, Saliou P, Gentilini M (1988) Etude de la chloroquine-résistance de *Plasmodium falciparum* à Vanuatu (1980-1984): apparition, evolution, distribution. Bull Soc Path Exot 81: 226-237.
14. Bastien P (1987) Particularités épidémiologiques des accès perniciox à *Plasmodium falciparum* dans un contexte d' épidémie palustre, Vanuatu, 1975-1985. Med Trop. 47: 125-131.
15. Maitland K, Williams TN, Peto TEA, Day KP, Clegg JB, Weatherall DJ, Bowden DK (1997) Absence of malaria-specific mortality in children in an area of hyperendemic malaria. Trans R Soc Trop Med Hyg. 91: 562-566.
16. Maitland K, Williams TN, Bennett S, Newbold CI, Peto TE, Viji J, Timothy R, Clegg JB, Weatherall DJ, Bowden DK (1996) The interaction between *Plasmodium falciparum* and *P. vivax* in children on Espiritu Santo island, Vanuatu. Trans R Soc Trop Med Hyg. 90: 614-620.
17. Maitland K, Kyes S, Williams TN, Newbold CI (2000) Genetic restriction of *Plasmodium falciparum* in an area of stable transmission: an example of island evolution?. Parasitology. 120: 335-343.
18. Williams TN, Maitland K, Bennett S, Ganczakowski M, Peto TEA, Newbold CI, Bowden DK, Weatherall DJ, Clegg JB (1996) High incidence of malaria in alpha-thalassaemic children. Nature. 383: 522-525.
19. Kaneko A, Taleo G, Kalkoa M, Yaviong J, Reeve PA, Ganczakowski M, Shirakawa C, Palmer K, Kobayakawa T, Björkman A (1998) Malaria epidemiology, glucose 6-phosphate dehydrogenase deficiency and human settlement in the Vanuatu Archipelago. Acta Trop 70: 285-302.
20. Mouchet J (1997) Le Paludisme au Vanuatu, Rapport de Mission Sante No.5. Port Vila: ORSTOM.
21. Kaneko A, Taleo G, Kalkoa M, et al (2000) Malaria eradication on islands. Lancet 356: 1560-1564.
22. Mitchell TD, Hulme M, New M (2002) Climate data for political areas. Area 34:109–112
23. Zeileis A (2005) a unified approach to structural change tests based on ML scores, F statistics and OLS residuals. Econ Rev 24: 445-466.
24. Zeileis A, Kleiber Ch, Krämer W, Hornik K (2003) Testing and dating structural changes in practice. Comp Stat Dat Anal 44:109-123.

25. Zurbenko I, Porter PS, Rao ST, Ku JY, Gui R, Eskridge RE (1996) Detecting Discontinuities in Time Series of Upper-air Data: Development and Demonstration of an Adaptive Filter Technique. *J Clim* 9: 3548-3560.
26. Hansen BE (1992) Tests for parameter instability in regressions with I(1) processes. *J Bus Econ Stat* 10:321-335.
27. Hansen BE (1997) Approximate asymptotic p values for structural changes tests. *J Bus Econ Stat* 15: 60-67.
28. Bai J, Perron P (1998) Estimating and testing linear models with multiple linear models. *Econometrica* 66: 47-78.
29. Bai J, Perron P (2003) Computation and analysis of multiple structural change models. *J Appl Econometrics* 18: 1-22.
30. Shumway RH, Stoffer DS (2000) *Time series analysis and its applications* New York: Springer. 572 p.
31. Ploberger W, Krämer W (1992) The CUSUM test with OLS residuals. *Econometrica* 60: 271-285.
32. Chaves LF (2007) *Casas Muertas and Oficina No. 1*: internal migrations and malaria trends in Venezuela 1905-1945. *Parasitol Res.* 101: 19-23.
33. Maraun D, Kurths J (2004) Cross wavelet analysis: Significance testing and pitfalls. *Nonlinear Proc Geophys* 11: 505–514.
34. Chaves LF, Pascual M (2006) Climate Cycles and Forecasts of Cutaneous Leishmaniasis, a Nonstationary Vector-Borne Disease. *PLoS Med* 3: e295.
35. Scheffer M, Carpenter S, Foley JA, Folke C, Walker B (2001) Catastrophic shifts in ecosystems. *Nature* 413: 591-596.
36. Levins R, Tchuenche JM (2006) Stability in variance: a new criterion for stability. *Far East J Dyn Sys* 8: 27-36.
37. Vandermeer J, Yodzis P (1999) Basin boundary collision as a model of discontinuous change in ecosystems *Ecology* 80:1817-1827.
38. Hastings A (2004) Transients: the key to long-term ecological understanding?. *Trends Ecol Evol* 19: 39-45.
39. van Nes EH, Scheffer M (2007) Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift *Am Nat* 169: 738-747

40. Levin SA (1992) The problem of pattern and scale in ecology. *Ecology* 73: 1943-1967.
41. Cohen JE (1973) Heterologous immunity to malaria. *Q Rev Biol.* 48:467-489.
42. Molineaux L, Storey J, Cohen JE, et al (1980) A longitudinal study of human malaria in the west African Savanna in the absence of control measures, relationships between different *Plasmodium* species, in particular *P. falciparum* and *P. malariae*. *Am J Trop Med Hyg* 29: 725-737.
43. Boyd MF, Kitchen SF (1938) Vernal vivax activity in persons simultaneously inoculated with *Plasmodium vivax* and *Plasmodium falciparum* *Am J Trop Med* 18: 505-514
44. Bruce MC, Donnelly CA, Alpers MP, et al (2000) Cross-species interactions between malaria parasites in humans. *Science* 287:845-848.
45. O'Meara WP, Collins WE, McKenzie FE (2007) Parasite Prevalence: A Static Measure of Dynamic Infections *Am J Trop Med Hyg* 77: 246 - 249.
46. Carpenter SR, Brock WA (2006) Rising variance: a leading indicator of ecological transition. *Ecol Lett* 9: 308-315.
47. Killeen GF, Smith TA, Ferguson HM, et al (2007) Preventing childhood malaria in Africa by protecting adults from mosquitoes with insecticide-treated nets. *PLoS Med.* 4: e229.
48. Williams TN, Maitland K, Foley DH (1995) Malaria sporozoite rates for *Anopheles farauti* s.s. Laveran (Diptera, Culicidae) from Vanuatu *Ann Trop Med Parasitol* 89: 305-307.
49. Becker NG, Dietz K (1995) The Effect of household distribution on transmission and control of highly infectious diseases. *Math Biosci* 127: 207-219.
50. Woolhouse MEJ, Dye C, Etard JF, et al (1997) Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc Natl Acad Sci USA* 94:338-342.
51. Cardinal MV, Lauricella MA, Marcet PL, Orozco MM, Kitron U, Gurtler RE (2007) Impact of community-based vector control on house infestation and *Trypanosoma cruzi* infection in *Triatoma infestans*, dogs and cats in the Argentine Chaco. *Acta Trop* 103: 201-211.

52. Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull Entomol Soc Amer* 15: 237-240.
53. Le Menach A, Takala S, McKenzie FE, Perisse A, Harris A, Flahault A, Smith DL (2007) An elaborated feeding cycle model for reductions in vectorial capacity of night-biting mosquitoes by insecticide-treated nets. *Malar J* 6:10
54. Lengeler C, Sharpe B (2003) *Indoor Residual Spraying and Insecticide-treated Nets*. Washington: Global Health Council. 17-24 p.
55. Lengeler C (2004) Insecticide-treated bed nets and curtains for preventing malaria. *Cochrane Database Syst Rev*: CD000363.
56. Lengeler C, Armstrong-Schellenberg J, D'Alessandro U, Binka F, Cattani J (1998) Relative versus absolute risk of dying reduction after using insecticide-treated nets for malaria control in Africa. *Trop Med Int Health* 3: 286-290.
57. Abdulla S, Schellenberg JA, Nathan R, et al (2001) Impact on malaria morbidity of a programme supplying insecticide treated nets in children aged under 2 years in Tanzania: community cross sectional study. *BMJ* 322: 270–73.
58. Lindblade KA, Eisele TP, Gimnig JE, et al (2004) Sustainability of reductions in malaria transmission and infant mortality in western Kenya with use of insecticide treated bednets. *JAMA* 291: 2571-2580.
59. Lindsay SW, Gibson ME (1988) Bednets revisited- Old idea, new angle. *Parasitol. Today* 4: 270-272.
60. Mathanga DP, Campbell CH, Taylor TE, Barlow R, Wilson ML (2005) Reduction of childhood malaria by social marketing of insecticide-treated nets: a case-control study of effectiveness in Malawi. *Am J Trop Med Hyg* 73: 622-625.
61. Magris M, Rubio-Palis Y, Alexander N, Ruiz B, Galvan N, Frias D, Blanco M, Lines J (2007) Community-randomized trial of lambda-cyhalothrin-treated hammock nets for malaria control in Yanomami communities in the Amazon region of Venezuela. *Trop Med Int Health*. 12: 392-403.
62. Bhattarai A, Ali AS, Kachur SP, et al (2007) Impact of artemisin based combination therapy and insecticide treated nets on malaria burden in Zanzibar. *PLoS Med* 4:e309

63. Ilboudo-Sanogo E, Cuzin-Ouattara N, Diallo DA, et al (2001) Insecticide-treated materials, mosquito adaptation and mass effect: entomological observations after five years of vector control in Burkina Faso. *Trans R Trop Med Hyg* 95: 353-360.
64. Dabiré RK, Diabaté A, Baldet T, Paré-Toé L, Guiguemdé RT, Ouédraogo JB, Skovmand O (2006) Personal protection of long lasting insecticide-treated nets in areas of *Anopheles gambiae* s.s. resistance to pyrethroids. *Malar J* 5: 12
65. Killeen GF, Smith TA (2007) Exploring the contributions of bed nets, cattle, insecticides, and excito-repellency to malaria control: a deterministic model of mosquito host-seeking behaviour and mortality. *Trans Roy Soc Trop Med Hyg* 101: 867-880.
66. Lehane M (2005) The biology of blood sucking in insects. Cambridge University Press. 321 p.
67. Takken W (2002) Do insecticide-treated bednets have an effect on malaria vectors? *Trop Med Int Health* 7: 1022-1030.
68. Geissbuhler Y, Chaki P, Emidi B, et al (2007) Interdependence of domestic malaria prevention measures and mosquito-human interactions in urban Dar es Salaam, Tanzania. *Malar J* 6: e126
69. Walker K, Lynch M (2007) Contributions of *Anopheles* larval control to malaria suppression in tropical Africa: review of achievements and potential. *Med Vet Entomol* 21: 2-21.
70. Chaves LF, Cohen JM, Pascual M, Wilson ML (2008) Social exclusion modifies the effects of deforestation and climatic variability on a vector-borne disease. *PLoS Neg. Trop. Dis.* 2: e176.
71. Kumar R, Hwang JS (2006) Larvicidal efficiency of aquatic predators: A perspective for mosquito biocontrol. *Zool Stud* 45: 447-466.
72. Blaustein L, Chase JM (2007) Interactions between mosquito larvae and species that share the same trophic level. *Ann Rev Entomol* 52: 489-507
73. Knight TM, Chase JM, Goss CW, et al (2004) Effects of interspecific competition, predation, and their interaction on survival and development time of immature *Anopheles quadrimaculatus*. *J Vector Ecol* 29: 277-284.
74. Yasuoka J, Levins R (2007) Ecology of vector mosquitoes in Sri Lanka: suggestions for future mosquito control in rice ecosystems. *Southeast Asian J Trop Med Public Health* 38: 646-657.

75. Chaves LF, Hernandez MJ, Dobson AP, et al (2007) Sources and sinks: revisiting the criteria for identifying reservoirs for American cutaneous leishmaniasis. *Trends Parasitol* 23: 311-316.
76. Mathanga DP, Campbell CH, Taylor TE, Barlow R, Wilson ML (2006) Socially marketed insecticide-treated nets effectively reduce *Plasmodium* infection and anaemia among children in urban Malawi. *Trop Med Int Health* 11: 1367-1374.
77. Molez JF (1999) Les mythes représentant la transmission palustre chez les Indiens d'Amazonie et leurs rapports avec deux modes de transmission rencontrés en forêt. *Sante* 9: 159-162.
78. Yasuoka J, Mangione TW, Spielman A, Levins R (2006) Impact of education on knowledge, agricultural practices and community actions for mosquito control and mosquito-borne disease prevention in rice ecosystems in Sri Lanka. *Am J Trop Med Hyg.* 74: 1034-1042.
79. Skarbinski J, Massaga JJ, Rowe AK, Kachur SP (2006) Distribution of free untreated bednets bundled with insecticide via an integrated child health campaign in Lindi Region, Tanzania: lessons for future campaigns. *Am J Trop Med Hyg* 76: 1100-1106.
80. Mukabana WR, Kannady K, Kiama GM, Ijumba JN, et al (2006) Ecologists can enable communities to implement malaria vector control in Africa. *Malar J* 5:9
81. Morales H, Perfecto I (2000) Traditional knowledge and pest management in the Guatemalan highlands. *Agr Hum Val* 17: 49-63.
82. Stenseth NC, Ottersen G, Hurrell JW, Mysterud A, Lima M, et al. (2003) Studying climate effects on ecology through the use of climate indices: The North Atlantic Oscillation, El Niño Southern Oscillation and beyond. *Proc R Soc Biol Sci* 270: 2087–2096.
83. Patz JA, Campbell-Lendrum D, Holloway T, Foley JA (2005) Impact of regional climate change on human health. *Nature* 438: 310–317.
84. Hales S, Weinstein P, Soares Y, Woodward A (1999) El Niño and the Dynamics of Vectorborne Disease Transmission. *Environ Health Persp* 107: 99-102
85. Chase JM, Knight TM (2003) Drought-induced mosquito outbreaks in wetlands. *Ecol Lett* 6: 1017-1024.

86. Bouma MJ, Dye C (1997) Cycles of malaria associated with El Niño in Venezuela. *JAMA* 278:1772-1774.
87. Patz JA, Olson SH (2006) Malaria risk and temperature: Influences from global climate change and local land use practices. *Proc Ntnl Acad Sci USA* 103: 5635-5636.
88. Habluetzel A, Cuzin N, Diallo DA, et al (1999) Insecticide-treated curtains reduce the prevalence and intensity of malaria infection in Burkina Faso. *Trop Med Int Health* 4: 557-564.
89. Killeen GF, Kihonda J, Lyimo E, Oketch FR, et al (2006) Quantifying behavioural interactions between humans and mosquitoes: evaluating the protective efficacy of insecticidal nets against malaria transmission in rural Tanzania. *BMC Infect Dis.* 6:161.
90. Killeen GF (2003) Following in Soper's footsteps: northeast Brazil 63 years after eradication of *Anopheles gambiae*. *Lancet Infect Dis.* 3: 663-666.
91. Kitron U (1985) Malaria, agriculture, and development: lessons from past campaigns. *Int J Health Serv.* 17:295-326
92. Levins R (1995) Toward an integrated epidemiology. *Trends Ecol Evol* 10: 304.

CHAPTER VI

THE ECOLOGICAL DYNAMICS OF VECTOR-BORNE DISEASES UNDER THE EFFECTS OF A MULTIDIMENSIONAL CHANGING ENVIRONMENT: SPATIAL PATTERNS OF CUTANEOUS LEISHMANIASIS; THE ROLE OF SOCIAL MARGINALITY AND DEFORESTATION IN COSTA RICA

Introduction

American cutaneous leishmaniasis (ACL), a neglected infectious disease [1-4], is one of the main emerging and re-emerging vector-borne diseases in the Americas. It is a zoonotic vector-borne disease, caused by several species of *Leishmania* (Kinetoplastida: Trypanosomatidae) parasites and transmitted by sand flies (Diptera: Psychodidae). The (re)emergence of ACL has been associated with deforestation in the neotropics. For example, infection is highest among people living close to forest edges [5,6], and also elevated among workers that extract natural resources in forested areas [7,8]. This association with forest proximity/deforestation has led to the view that large-scale landscape transformation may reduce ACL emergence [6,8,9]. However, studies of ACL and forest cover thus far have ignored the multidimensionality of factors that shape patterns of infectious diseases [10]. Such multidimensionality is underscored by Schmalhausen's law, which states that biological systems at the boundary of their tolerance along any dimension of existence become more vulnerable to small changes along other such dimensions [11]. We suggest that this general principle is relevant to understanding environmental change and infectious

diseases, and more generally ecosystem functioning and diversity conservation, given the interactions of these phenomena with social and economic realms. Here we examined county-level ACL case data from 1996 through 2000 for Costa Rica, a country that proportionally has had the largest rate of landscape transformation in the New World [12-14], and report results contrary to the perspective that forest cover is the major risk factor for this disease. We began by qualitatively assessing the patterns of clustering of the disease incidence and risk factors, and the landscape level associations between ecosystems and the vectors. These analyses indicated that landscape alone does not explain the spatial distribution of ACL. Based on this information we proceeded with more quantitative analyses relating risk factors to the disease. Our more comprehensive analysis demonstrated that living close to the forest was negatively associated with infection incidence once social marginality was evaluated as a key variable in explaining disease pattern. The effects of these drivers are not monotonic, but rather display "breakpoints" or threshold values at which the shape and magnitude of the relationship change [15-17]. Forest cover certainly plays an important role in modulating the response of pathogen transmission to other environmental changes [18], specifically climate variability by the El Niño Southern Oscillation (ENSO). However, we have identified possible ecological mechanisms related to infection risk that may explain these macroscopic patterns, and suggest alternatives in planning development policies if the long term goals of biodiversity conservation, control of infectious diseases, and sustainable human well-being are to be pursued in concert.

Methods

Data

The monthly number of cases of American Cutaneous Leishmaniasis (ACL) from January 1996 through December 2000 was obtained from the epidemic surveillance service of Costa Rica “*Vigilancia de la Salud*” for the 81 counties that comprise the country. The total number of cases for this period was 3379. County-level data on the percent of people living <5 km from the forest (%close) and percent forest cover, as of 2000, were obtained from [19]. Social marginality is in general referred as the lack or limited access to resources that ensure a satisfactory quality of life [20]. Social Marginalization index values (MI), based on the 2000 Costa Rican national census, were obtained from [20]. This marginalization index is a robust measure of social outcast status since it is constructed using several variables associated with social exclusion, including income, literacy, level of education, average distance to health centers, health insurance coverage, etc. Monthly rainfall data were obtained from 14 weather stations across the country available at the Earth Observing Laboratory, National Center for Atmospheric Research [<http://data.eol.ucar.edu/>], and the yearly average was calculated for each station. Ordinary kriging was employed to interpolate average rainfall values across the country using the Geostatistical Analyst extension in ArcGIS 9.1, and averages of mean, minimum (MinRfill) and maximum yearly rainfall for each county were calculated. An elevation data layer in raster format with 30 arc-second ($\sim 1 \text{ km}^2$) resolution was obtained from the United States Geologic Survey (USGS)

[<http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html>], and minimum (ME), maximum, average and standard deviation of elevation for each county were calculated using Hawth's Analysis Tools for ArcGIS [21].

Data on species and locations of sand fly captures were obtained from systematic reviews on human biting species from Costa Rica [22,23].

Coordinates of sand fly captures were compared against the Central American Ecosystems Map [<http://mitchnts1.cr.usgs.gov/data/otheragency.html>] created by Costa Rica's Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) and described by [24]. The ecosystem map, derived from Landsat satellite imagery, was created with ArcGIS at a resolution of 1 km² grid cells and then used to define the ecological type in which each of the sand fly species was located.

Statistical Methods

Kuldorff's Scan Statistic. This method finds spatio-temporal clusters by detecting the excess of cases in a given region under the assumption that cases are generated by an inhomogeneous Poisson point process with an intensity, μ , proportional to the population at risk. The method is implemented by moving a circular window systematically through the study area, starting at the centroid of each location in the dataset [25]. The window expands to include the nearest region centroids, and its maximum size does not exceed 50% of the total population at risk size for the study period. The null hypothesis of a Poisson process is tested through a maximum likelihood ratio test that compares it to an

alternative model stating that this assumption is false, with the significance tested through multinomial Monte Carlo. The analysis was implemented with the Clusterseer software and significance of clusters was tested with 999 Monte Carlo randomizations. We assumed that the population at risk was that of the whole county, and used data from the 1983 and 2000 Costa Rican censuses [<http://ccp.ucr.ac.cr/>] with linear interpolation from January 1996 through December 2000.

Local Indicators of Spatial Autocorrelation (LISA). We used this technique to analyze the patterns of clustering in potential risk factors for the disease. LISA, a local adaptation of Moran's I, compares the value of the variable of interest in a given county with those in neighboring counties. The degree of similarity between neighboring counties was compared to that expected by chance to determine where clusters of high or low values occur [26]. To ensure the robustness of results, both queen contiguity and four-nearest neighbors were used as weights and the output compared for each variable using the GeoData Analysis software package.

Negative Binomial Generalized Linear Models (NB-GLM) with breakpoints.

We introduced breakpoints in predictors by transforming the predictor using a breakpoint basis function of the form:

$$B_L(x) = \begin{cases} x - c & x < c \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$B_R(x) = \begin{cases} x - c & x > c \\ 0 & \text{otherwise} \end{cases}$$

where c is the breakpoint where the functions $B_L(x)$ and $B_R(x)$ join each other, and are used to separate the relationship between the response and the predictors to the left and the right of the break point respectively. This technique is known as hockey stick regression [27]. Four models were fitted using maximum likelihood for NB-GLM with logarithmic link and fixed over-dispersion parameter [28]. Nonlinear forms observed in the Generalized Additive Models (GAM) presented in Protocol S1 were approximated by using the following models:

$$\text{Rate}_i = \mu_0 + \beta_1 \text{ME}_i + \beta_2 B_L(\text{MI}_i) + \beta_3 B_R(\text{MI}_i) + \beta_4 B_L(\% \text{close}_i) + \beta_5 B_R(\% \text{close}_i) + \beta_6 B_R(\log(\text{MinRfl}_i)) + \beta_7 B_L(\log(\text{MinRfl}_i)) + \beta_8 (B_L(\log(\text{MinRfl}_i)))^2 + \varepsilon \quad (2.1)$$

$$\text{Rate}_i = \mu_0 + \beta_1 \text{ME}_i + \beta_2 B_L(\text{MI}_i) + \beta_3 B_R(\text{MI}_i) + \beta_4 B_L(\% \text{Close}_i) + \beta_5 B_R(\% \text{close}_i) + \beta_6 (B_R(\% \text{close}_i))^2 + \beta_7 B_R(\log(\text{MinRfl}_i)) + \beta_8 B_L(\log(\text{MinRfl}_i)) + \beta_9 (B_L(\log(\text{MinRfl}_i)))^2 + \varepsilon \quad (2.2)$$

$$\text{Rate}_i = \mu_0 + \beta_1 \text{ME}_i + \beta_2 B_L(\text{MI}_i) + \beta_3 B_R(\text{MI}_i) + \beta_4 (B_R(\text{MI}_i))^2 + \beta_5 B_L(\% \text{close}_i) + \beta_6 B_R(\% \text{close}_i) + \beta_7 B_R(\log(\text{MinRfl}_i)) + \beta_8 B_L(\log(\text{MinRfl}_i)) + \beta_9 (B_L(\log(\text{MinRfl}_i)))^2 + \varepsilon \quad (2.3)$$

$$\begin{aligned}
\text{Rate}_i = & \mu_0 + \beta_1 \text{ME}_i + \beta_2 \text{B}_L(\text{MI}_i) + \beta_3 \text{B}_R(\text{MI}_i) + \beta_4 (\text{B}_R(\text{MI}_i))^2 + \\
& \beta_5 \text{B}_L(\% \text{close}_i) + \beta_6 \text{B}_R(\% \text{close}_i) + \beta_7 (\text{B}_R(\% \text{close}_i))^2 \\
& + \beta_8 \text{B}_R(\log(\text{MinRfl}_i)) + \beta_9 \text{B}_L(\log(\text{MinRfl}_i)) + \beta_{10} (\text{B}_L(\log(\text{MinRfl}_i)))^2 + \varepsilon
\end{aligned} \tag{2.4}$$

Models have the same predictors described for the GAM presented in Protocol S1. For the purpose of comparison a simpler null model without breakpoints was also fitted:

$$\text{Rate}_i = \mu_0 + \beta_1 \text{ME}_i + \beta_2 (\text{MI}_i) + \beta_3 (\% \text{close}_i) + \beta_4 (\log(\text{MinRfl}_i)) + \varepsilon \tag{3}$$

as well as a model assuming smooth non-linear relationships with MI and %close:

$$\begin{aligned}
\text{Rate}_i = & \mu_0 + \beta_1 \text{ME}_i + \beta_2 (\text{MI}_i) + \beta_3 (\text{MI}_i)^2 + \beta_4 (\% \text{close}_i) + \beta_5 (\% \text{close}_i)^2 \\
& + \beta_6 \text{B}_R(\log(\text{MinRfl}_i)) + \beta_7 \text{B}_L(\log(\text{MinRfl}_i)) + \beta_8 (\text{B}_L(\log(\text{MinRfl}_i)))^2 + \varepsilon
\end{aligned} \tag{4}$$

To make comparisons reliable, the variance over-dispersion parameter of the negative binomial response was fixed to 1, and not estimated independently for each model [26]. Models were fitted by off-setting the logarithm of population size on the right hand side of the equations as recommended for rate models [29]. Values for the break points “c” were estimated a priori by minimizing the value for Akaike Information Criterion (AIC) of a function fitting the model while considering

breakpoint values for predictors: MI, %close and MinRfill using an algorithm based on the Newton method [30]. The model selected as "best" was further subjected to a process of model selection by backward elimination as described in [1,27]. Goodness of fit for the final model was assessed using a Chi² test with degrees of freedom (df) defined as n-p-1, where n is the number of observations, p the number of parameters estimated in the model, and the additional df accounts for the dispersion parameter of the negative binomial. Diagnostics for spatial autocorrelation were carried out by regressing residuals on the centroids of each county. The error (ϵ) was assumed to be identically and independently normally distributed for the linear predictor of the NB-GLM [28,29].

Linear Models and Analysis of Covariance (ANCOVA). Parameters have a linear relationship with the response variable and were computed using ordinary least squares [27]. Models incorporated ENSO, county and their interaction as predictors. The definition for covariates and the response are similar to those used for the Linear Mixed Effects Models (LMEM), as are the assumptions about the error (ϵ , see Protocol S1). The linear model used for the ANCOVA is given by:

$$\log(\text{Cases}(t)_i + 1) = \mu + \beta_1 \text{ENSO}(t) + \beta_2 \text{County}_i + \beta_3 \text{ENSO}(t) * \text{County}_i + \epsilon(t) \quad (5)$$

In the process of model building, autoregressive components were tested but they were not significant. However, for the sake of comparison, the fitting of

model (5) only included the data from 1997 through 2000. Diagnostics for spatial autocorrelation were carried out by regressing residuals on the centroids of each county.

Results

We began by exploring whether the spatial distribution of disease incidence was heterogeneous across the country, a pattern that might be expected from the considerable heterogeneity of ecosystems in Costa Rica. Figure 6.1 shows that disease incidence and social marginalization (described in Materials and Methods) achieved their highest values in the same counties, a pattern not found for other ecological variables such as minimum rainfall, minimum elevation, landscape composition index, proportion of forest cover, or proportion of people living within 5 Km of the forest edge. This pattern was confirmed by spatial statistical analyses that detected overlapping geographical clusters for both disease and social marginalization, a pattern that was again absent for other variables and robust to the methodology applied to find clusters (Figure 6.1, Figure 6.S1 and Figure 6.S3). To further investigate counties where ACL was clustered, we analyzed the percentage of various landscape compositions using principal component analysis (PCA) for the most common landscape units known to harbor human biting sand fly species (Tables 6.S1, 6.S2, 6.S3). No clear effect of landscape composition was found, as counties where the disease was clustered were within the ranges of variability of all counties in the country. We further tested the robustness of this result using

multidimensional scaling, a method lacking the linearity constraints of PCA, with strikingly similar results (Tables 6.S1, 6.S2, 6.S3 and Figure 6.S2).

To examine further and more quantitatively the factors determining observed spatial patterns of ACL, we fitted GAMs to the five-year ACL incidence rate (total cases during 1996 through 2000 divided by the 2000 population) as a function of several variables (see statistical methods in Appendix S4 for a detailed description and Table 6.S4). A process of model selection by backward elimination (see Appendix S4) resulted in the following relevant variables: the marginalization index (MI), % of people living close to the forest (% close), log(minimum rainfall) and minimum elevation (ME). All variables except minimum elevation exhibited non-linear relationships with disease incidence, explaining 78% of the variance. Because GAMs are difficult to interpret and the fitted smoothed functions of GAMs showed clear qualitative changes (see Table 6.S4 and Figure 6.S3), we fitted somewhat simpler negative binomial generalized linear models (NB-GLM) that incorporated breakpoints (see Materials and Methods). The best model selected using this methodology accounted for 72% of variability ($1 - \text{residual deviance} / \text{null deviance}$). Furthermore, major qualitative differences in the association of rates with some relevant variables were more easily visualized (Figure 6.2). Interestingly, a simpler model, not incorporating breakpoints, explained only 60 % of the variability (i.e. model deviance), and failed to capture the significance of the relationship between disease rates and proportion of people living <5 km from the forest border within each county. This breakpoint relationship with covariates was further supported by smaller Akaike

Information Criterion (AIC) for breakpoint models, as compared to a model in which the relationships with covariates were described by smooth functions with the same number of parameters (second degree polynomial; Table 6.1).

To address effects of hierarchically nested geopolitical units (e.g., counties belonging to provinces) and of interannual climatic variability (El Niño Southern Oscillation (ENSO)), we fitted Linear Mixed Effects Models (LMEM). These models incorporated geopolitical subdivisions of the country as nested random factors, and ENSO as a continuous predictor (details in Appendix S4). Neither ENSO nor the geopolitical nesting of counties had significant effects based on bootstrap model comparisons, with the highest variability explained by unknown factors (Table 6.S5). These results could indicate that the effects of ENSO were very local (county scale), and different across counties. To test the hypothesis of localized ENSO effects, we fitted an Analysis of Covariance (ANCOVA) to the counties where disease was clustered. The results showed a statistically significant interaction between ENSO and the considered counties (goodness of fit $R^2=85\%$). The effects of ENSO are variable, with some counties showing an increase and others a decrease in incidence during a cycle of the oscillation (Figure 6.3). The only variable that showed a significant difference between these two groups was the percentage of forest cover, with a significantly larger fraction ($P<0.05$) in counties where incidence decreased (Figure 6.3).

Discussion

The finding that ACL tended to afflict socially marginal populations more heavily is common to other infectious diseases, and has been historically documented in public health studies particularly at small spatial scales [10,31,32]. We have shown here that social marginalization also can explain patterns of ACL at larger geographical scales. When this influence is taken into account, risk of infection is diminished among those living close to forests, an unexpected pattern in light of previous studies on the role of this habitat type. The pathway by which social marginalization promotes transmission of *Leishmania* in this context probably is linked to a major environmental problem affecting the tropics: destruction of forests and associated biodiversity. Forest clearing worldwide [33,34], and especially in Costa Rica, is concurrent with development of large scale commercial agriculture [12,13,14], including monocultures of several commercial crops where ACL is clustered, and with accelerated human population growth [14]. This shift towards market-based agricultural production and rapidly expanding population is associated with new inequities in land tenure [35], increased numbers of landless peasants, and hence further pressure to cut down forests for local subsistence agriculture [13] and extraction of other natural resources [36,37].

Risk of ACL infection in rural Costa Rica has been especially associated with the exposure to forests close to agricultural environments [38,39]. The latter could imply that populations living inside or close to fragmented forests intermixed with crops where the overall biodiversity of the landscape is reduced

could have a higher risk of infection when compared with those where the agricultural practices and crops allow the maintenance of biodiversity. Supporting this idea is the ecological knowledge about biodiversity in disturbed, fragmented, and isolated landscapes proceeds through a series of well-documented, ecological syndromes, starting with habitat destruction and associated biodiversity reduction [40], followed by loss of keystone species and resulting structural changes leading to reduced biodiversity [41]. Changes in biodiversity due to deforestation are probably of importance to ACL since the major reservoirs of *Leishmania* species are small mammals, including marsupials, rodents and sloths [5,6]. Forest fragmentation has been shown to increase densities of these species, because in small and isolated habitat fragments, large predators are lost first, leading to major changes in inter-specific interactions that decrease mammal biodiversity and lead to the dominance of rodents [42,43]. This scenario, extensively studied for Lyme disease which is another rodent-associated, vector-borne disease, involves increased diversity of hosts providing a “dilution” effect on transmission [43,44]. Similar mechanisms may be at play for ACL as suggested by mathematical models of transmission dynamics and by field studies that show only a small number of mammal species are infected with *Leishmania* spp. among those that are frequently bitten by sand fly vectors [3,45,46].

Changes in landscape quality are also likely to affect composition of the arthropod vector community [47]. Interestingly, sand fly species richness is greater in traditional, shaded coffee agroecosystems than in those that are

intensified and unshaded [48]. More generally, traditional coffee production supports similar biodiversity as undisturbed forests [49]. Reduced forest cover in our study modulated the effects of climate variability (ENSO), an interaction that may operate through multiple pathways. Increased temperatures in modified landscapes can directly affect transmission of vector-borne diseases [50]. The negative effects of climatic variability on crops, accompanied by associated increases in reliance on the exploitation of forest resources [1,51], may have large impacts on the ACL transmission system. In addition, disruption of trophic structures known to increase densities of certain small mammals, including possible *Leishmania* spp. reservoirs [52-54], can be amplified by ENSO anomalies that alter resources [1,52]. The influences of rainfall and elevation on the spatial distribution of ACL are probably mediated through the effects of humidity and temperature on the biology of both vectors and parasites [1].

Future work should examine the role of local climate variability encompassing multiple ENSO events over a longer time span, as was previously done at the coarse scale of the whole country [1]. A special emphasis should be put on elucidation of mechanisms acting at a local scale, since operational control strategies require further details about local characteristics increasing the risk of transmission, while always contextualizing these risk factors within the multidimensional nature of human disease. This can be achieved by considering aspects as diverse as the demographic structure of cases and the relationships between forest fragmentation and biodiversity on the endemic areas of the disease. Another effort could explore the relationship between ACL and different

systems of agricultural production that might affect the ecology of transmission, as well as the perception and measures of protection that people take under different socio-economic conditions [10,47,48]. In a more theoretical realm, further attention should be given to a corollary of Schmalhausen's law of fundamental relevance to the resilience of ecosystems and their response to environmental change, namely the increase in the variance of systems under stress [e.g., 55]. Finally, our work underscores the need to place the control of ACL, and more generally of neglected tropical diseases and malaria, within a framework that encompasses ecologically sound development and viable solutions to the trade-offs between agriculture and conservation, such as shaded coffee production [48,49,51]. The quality of the landscape matrix is not only relevant to biodiversity conservation, as already recognized in studies of agroecosystems [37], but also to preventing the emergence and exacerbation of infectious diseases.

Table 6.1. Breakpoint values for the natural logarithm of minimum rainfall, the marginalization index, and the percentage of people living close to the forest for the studied models.

Model	Log(Min Rainfall)	Margin Index	% Close	No. Parameters	AIC
I	7.78	4.13	49.40	8	567.7
II	7.78	4.13	49.98	9	569.5
III	7.77	4.13	49.99	9	569.1
IV	7.77	4.13	49.99	10	570.7
Smooth	7.79	Poly 2	Poly 2	8	574.2
Null	—	—	—	4	595.4

The number of parameters does not include the dispersion parameter for the negative binomial generalized linear models, which was set to 1 (see Protocol S1 for details).

Table 6.S1. Ecosystems and number of locations where human biting sand fly species have been caught in Costa Rica (S6, S7)

Locations	Ecosystem
83	Agriculture
6	Evergreen Tropical Forest
1	Deciduous Tropical Forest
1	Manglar Forest
1	Embalse

Table 6.S2. Principal Component Analysis (PCA) for the landscape units where human biting sand flies have been caught in Costa Rica (S6, S7).

Component	1 st	2 nd	3 rd	4 th
Proportion of Variance	0.64	0.26	0.09	0.01
Cumulative Proportion	0.64	0.90	0.99	1.00

Table 6.S3. Factor Loadings for Ecosystems in components 1 & 2

Component	1 st	2 nd
Agriculture	0.610	---
Evergreen Low-Lands	-0.247	-0.880
Evergreen Montane	-0.523	0.473
Evergreen Sub-Montane	-0.541	---

Table 6.S4. Parameters, smooth function degrees of freedom and significance for the GAM described in equation 1.

Parametric Coefficients				
Parameter	Estimate	S.E.	T	P
μ_0	-8.185	0.094	-86.63	<2e-16
Smooth Terms				
Variable	EDF	Rank	F	P
s(MI)	3.504	8	3.345	0.03
s(% Close)	6.698	9	2.403	0.05
s(ME)	1.332	3	3.571	0.02
s(log(MinRfill))	5.562	9	5.943	6E-05

Table 6.S5. Comparison of Linear Mixed Effects models

Models	Likelihood ratio test	Bootstrap P
6.1 and 6.2	0.3815	0.611
6.2 and 6.3	5.931e-08	0.437
6.3 and 6.4	9.920e-08	0.923

Table 6.S6. Parameters for the model presented in Figure 6.2

Parameter	Estimate	S. E.	T	P
μ_0	-7.61828	0.471846	-16.146	< 2e-16*
β_1 (ME)	-0.00157	0.000386	-4.083	0.000112*
β_2 (B_R (MI))	0.375278	0.098044	3.828	0.000271*
β_3 (B_R (% Close))	-3.93909	1.42758	-2.759	0.007318*
β_4 (B_L (% Close))	0.019768	0.009035	2.188	0.031871*
β_5 (B_L (\ln (MinRainfall)))	2.434161	1.26397	1.926	0.058023
β_6 (B_L (\ln (MinRainfall)) ²)	-5.80016	1.202431	-4.824	7.50E-06*
β_7 (B_R (\ln (MinRainfall)))	5.512966	1.726986	3.192	0.002083*

*Statistically significant (P<0.05)

Table 6.S7 Analysis of Covariance for the model in (5)

Factor	ANCOVA		
	DF	F	P
		311.54	<0.000
Intercept	1	1	1
County	8	3.229	0.0891
ENSO	1	2.36	0.0621
County*ENSO	8	3.611	0.0113
Error	18		

Table 6.S8 Parameters for the linear model in (5). Intercept and ENSO are respectively the intercept and slope for Talamanca County, the reference county. For all other counties, intercept and slopes are found by adding the values in the table to the values for the reference county.

Parameter	Estimate	Std. Error	<i>t</i>	P
Intercept (Talamanca)	-4.5411	0.5007	-9.07	3.92E-08
ENSO (Talamanca)	-3.1518	0.4478	-7.038	1.44E-06
Aguirre	-3.6201	0.7081	-5.113	7.28E-05
Buenos Aires	-2.9878	0.7081	-4.22	0.000515
Corredores	-3.3055	0.7081	-4.668	0.000191
Coto Brus	-3.4213	0.7081	-4.832	0.000134
Golfito	-1.7606	0.7081	-2.486	0.022943
Limon	-2.6633	0.7081	-3.761	0.00143
Osa	-2.6661	0.7081	-3.765	0.001417
Perez Zeledón	-3.8459	0.7081	-5.431	3.69E-05
ENSO*Aguirre	3.681	0.6333	5.812	1.66E-05
ENSO*Buenos Aires	1.7485	0.6333	2.761	0.01287
ENSO*Corredores	3.3862	0.6333	5.347	4.41E-05
ENSO*Coto Brus	2.737	0.6333	4.322	0.000411
ENSO*Golfito	3.0889	0.6333	4.877	0.000121
ENSO*Limon	3.0571	0.6333	4.827	0.000135
ENSO*Osa	3.5617	0.6333	5.624	2.46E-05
ENSO*Perez Zeledón	3.6914	0.6333	5.829	1.60E-05

Figure 6.1. Patterns of clustering and Schmalhausen's law. (A) Quinquennial (1996-2000) cutaneous leishmaniasis case rates (cases/population) in Costa Rica at the county level. Colors indicate clustering in monthly rates per 10,000 inhabitants obtained using the Scan method: blue corresponds to the most likely cluster, comprised of the Talamanca county, with a monthly rate 308 per 10,000 from January 1999 to December 2000 (loglikelihood ratio = 3020.06, $P < 0.001$); green depicts the second most likely cluster, comprised of the counties of Osa, Buenos Aires, Aguirre, Perez Zeledon, Golfito, Coto Brus, Aguirre y Corredores, with a rate of 7 per 10,000 from June 1996 to November 1999 (loglikelihood ratio = 515, $P < 0.001$); and red corresponds to the third most likely cluster, comprised of the county of Limon with a rate of 12 per 10,000 from April 1997 to May 2000 (loglikelihood ratio = 265, $P < 0.001$). (B) The county marginalization index (See Protocol S1 for details). Red and blue indicate clusters with high and low marginality, respectively, found using the LISA method with weights based on the 4 nearest neighbors (overall $I = 0.7096$, $P < 0.05$). (C) County rate as a function of the marginalization index. Black dots represent counties with less than 2 cases in the five years. This pattern, which we call Schmalhausen's pattern, shows a significant positive correlation between marginality and the rate of the disease ($r = 0.39$, $t = 3.8221$, $df = 79$, $P < 0.0002$), where a qualitative change in the relationship is apparent after and before a value of 4 in the marginalization index. Specifically, the variance increase for larger values of social marginalization, consistent with the prediction that new or anomalous conditions modify the system's sensitivity to other drivers.

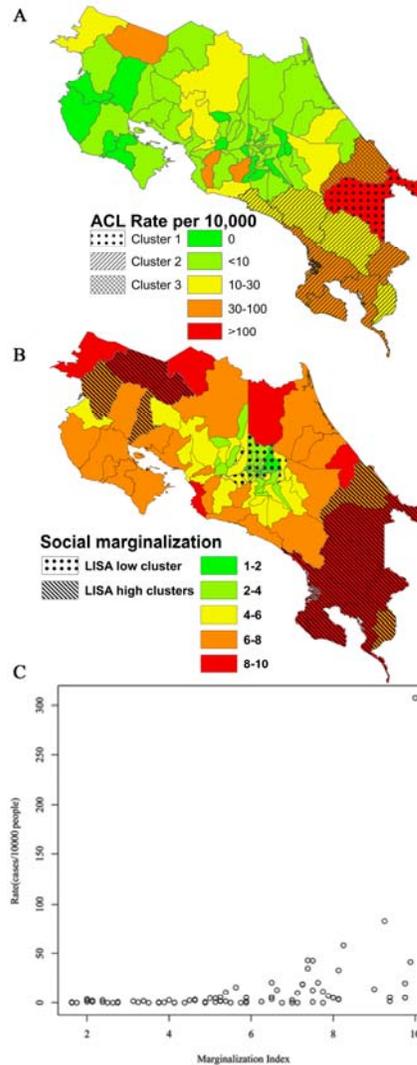


Figure 6.2. Breakpoints and discontinuous patterns of association. A schematic representation of the breakpoint in marginalization (MI) and people living close to the forest (%close), when minimum elevation (ME) is set to 500 m and rainfall (log(MinRfll)) is set at its breakpoint. The surface illustrates major qualitative differences in disease risk as a function of the covariates. Specifically, risk increases exponentially as the proportion of people living close to the forest decreases above the breakpoint. The change has the opposite sign and decreases in magnitude for smaller values below the breakpoint. Marginality exacerbates this difference above its own breakpoint. Parameters are those of the model selected as best. This model has 7 parameters (AIC=5768.7) and fits the data satisfactorily (Residual deviance = 79.718, df=72, $P>0.24$), explains 71.34% of the deviance (null deviance = 278.108) and is not different from the more complex models presented in Table 6.1, values for the coefficients are presented in Table 6.S6.

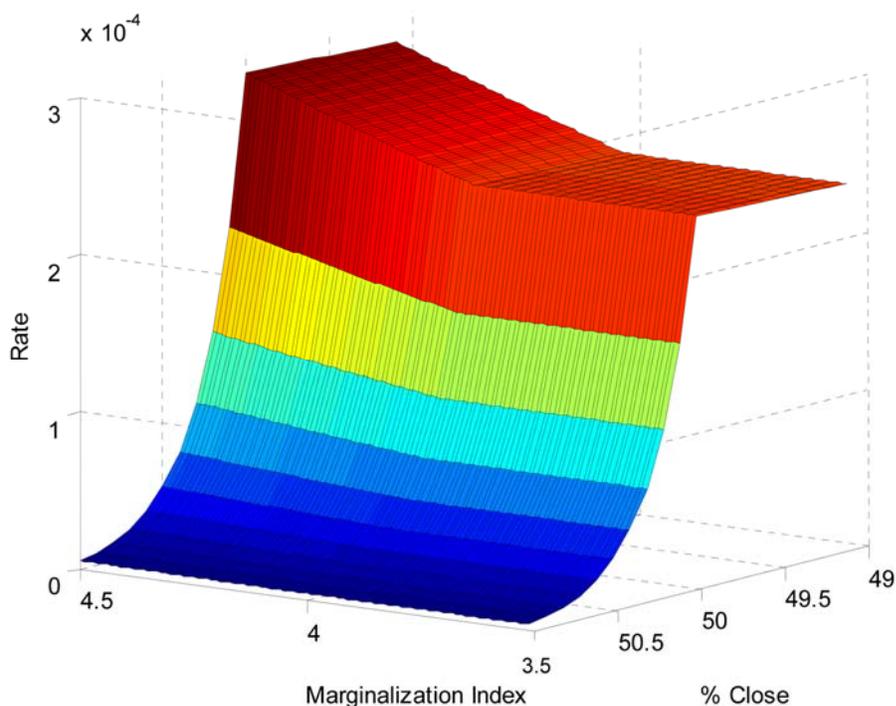


Figure 6.3. Cutaneous leishmaniasis in Costa Rica: Deforestation and El Niño Southern Oscillation (ENSO). (A) Local Effects of ENSO. Linear model results for a model testing for localized effects of ENSO in the counties where the disease was clustered. Color indicates clusters found with the spatio-temporal scan analysis of Figure 6.1, characters are used for the data in each individual county (For parameter values, see table in the appendix). For representation purposes, a small amount of noise was added in the x (ENSO) axis. The ANCOVA for this model showed the interaction of ENSO*County to be statistically significant ($P < 0.0113$, for more details see Tables 6.S7, 6.S8). The model has a high goodness of fit ($R^2 = 0.85$) that outperforms a similar model with the same number of parameters but that uses a first order autoregressive structure ($R^2 = 0.26$) instead of ENSO. (B) Differences in forest cover for counties where the incidence diminishes or increases with ENSO. In the boxplot, 1 stands for the counties where the annual rate decreases with ENSO (Talamanca, Limón, Golfito, Buenos Aires & Coto Brus) and 2 for those where the incidence increases with ENSO (Aguirre, Corredores, Osa & Pérez-Zeledón). The difference is statistically significant as shown by a one tail Welch's t-test (a test robust to differences in variance) in which the alternative hypothesis is that the difference in forest cover between 1 and 2 is larger than 0 (Welch's $t = 2.14$, d.f.=5.9, $p < 0.038$).

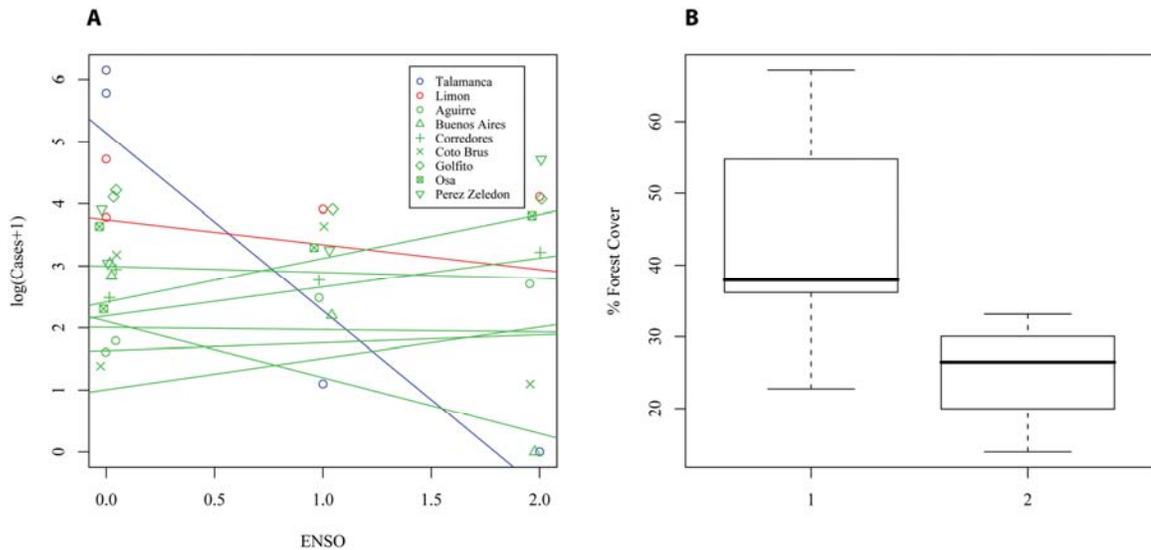


Figure 6.S1. Supplementary maps (A) Weather stations and interpolated values. Clusters of deforestation: (B) Queen contiguity. (C) 4 nearest neighbors. (D) Ecosystems of Costa Rica and number of sand fly species for each locality (see references [22],[23] in the main article).

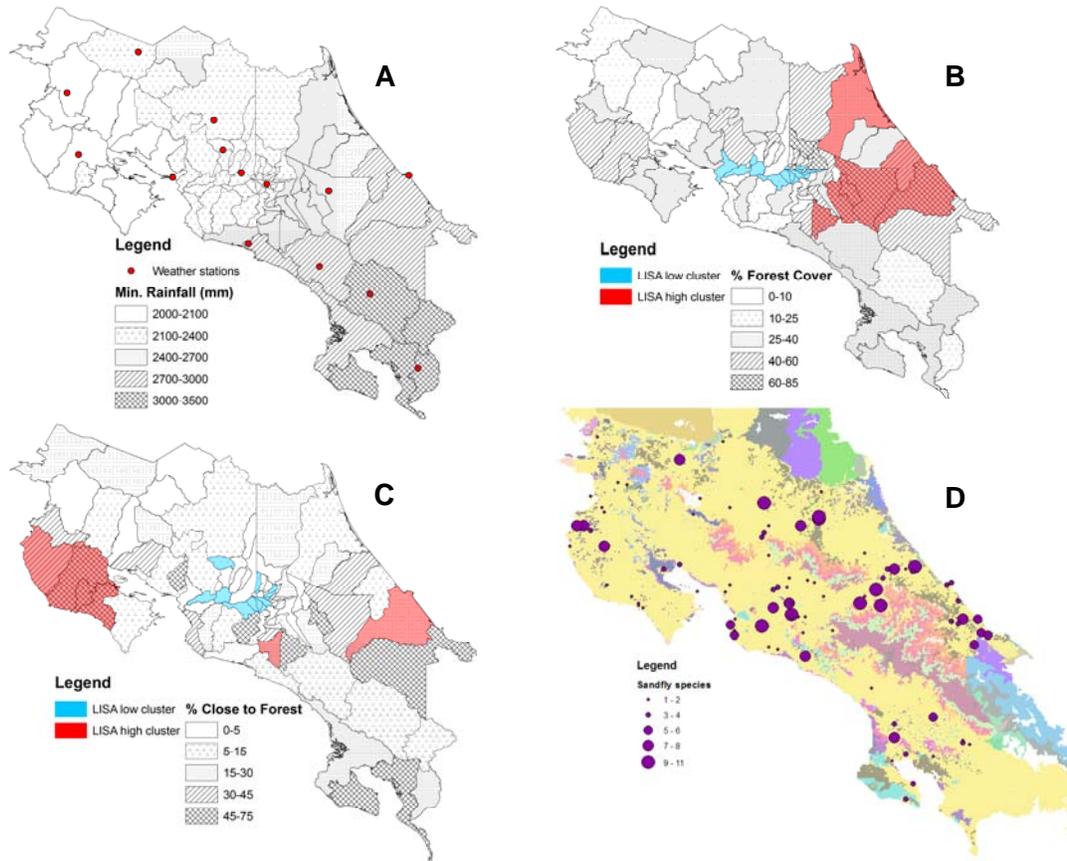


Figure 6.S2. Generalized Additive Model smooth functions. (A) Marginalization index. (B) % of People living within 5 km to the border of the forest. (C) Minimum elevation. (D) Log(minimum rainfall).

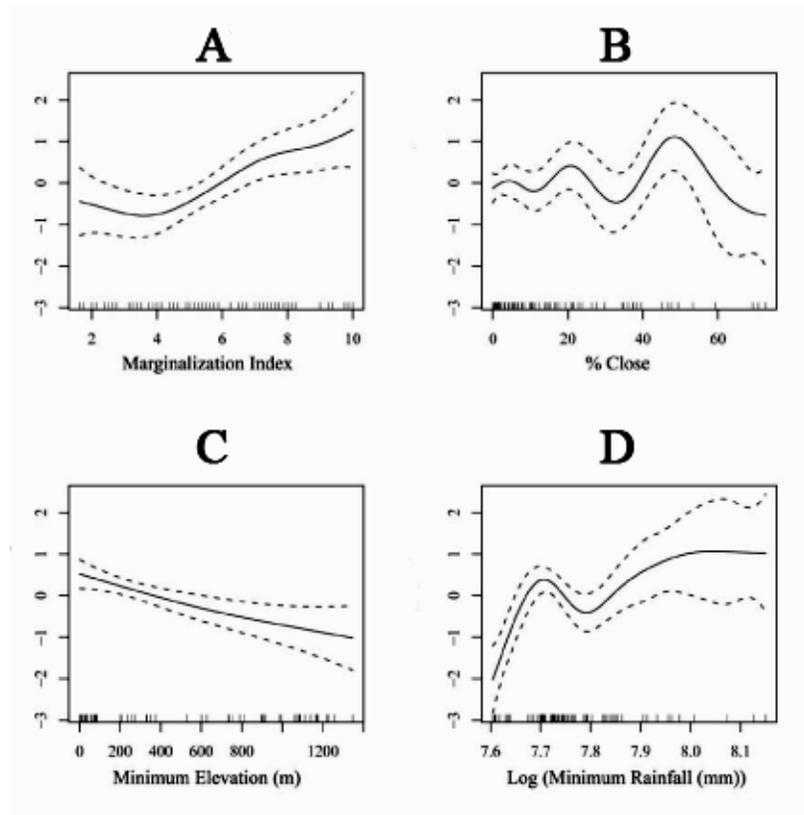
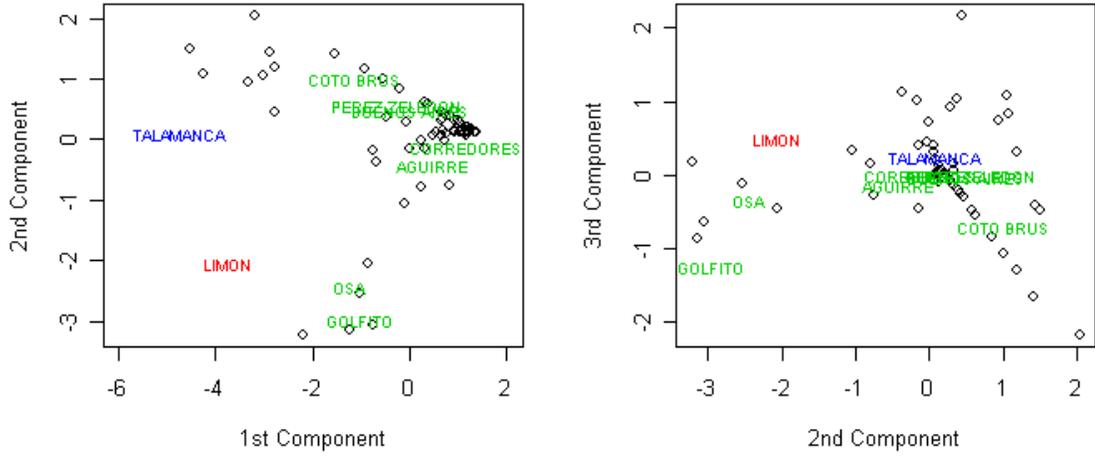
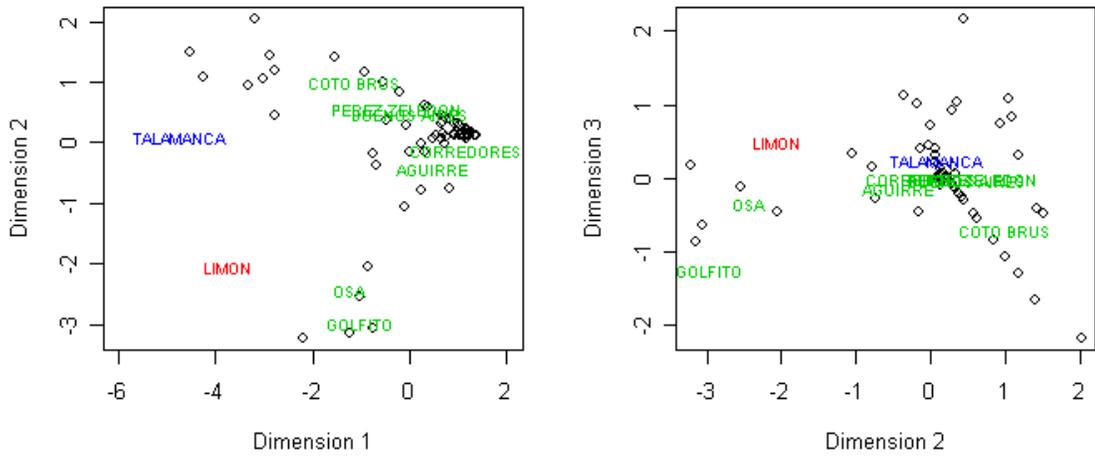


Figure 6.S3. Landscape dimension reduction. Top panels include the first three components from the PCA analysis presented in Tables S2 and S3. Bottom panels include three dimensions using a MDS analysis with 2.87 % for stress, a goodness of fit that is good.

Principal Component Analysis



Multidimensional Scaling



References

1. Chaves LF, Pascual M (2006) Climate cycles and forecasts of Cutaneous Leishmaniasis, a non-stationary vector borne disease. *PLoS Med* 3: e295.
2. Chaves LF, Pascual M (2007) Comparing models for early warning systems of neglected tropical diseases. *PLoS NTDs* 1: e33.
3. Chaves LF, Hernandez MJ, Dobson AP, Pascual M (2007) Sources and Sinks: revisiting the criteria for identifying reservoirs for American Cutaneous Leishmaniasis *Trends Parasitol* 23: 311-316.
4. Hotez PJ, Molyneux DH, Fenwick A, Ottesen E, Erlich-Sachs S, Sachs JD(2006) Incorporating a rapid impact package for neglected tropical diseases with programs for HIV/AIDS, tuberculosis and malaria. *PLoS Med* 3: e102.
5. Badaro R (1988) Current situation in regard to leishmaniasis in Brazil. In: *Research on Control Strategies for the Leishmaniasis; Proceedings of an International Workshop held in Ottawa, Canada 14 June 1987*. B.C. Walton et al. (eds). IDRC Manuscript Report 184e. International Development Research Centre, Ottawa, Canada, pp.91-100.
6. Yadon ZE, Rodrigues LC, Davies CR, Quigley MA (2003) Indoor and peridomestic transmission of American cutaneous leishmaniasis in northwestern Argentina: A retrospective case-control study. *Am. J. Trop. Med. & Hyg.* 68: 519-526.
7. Davies CR, Reithinger R, Campbell-Lendrum D, Feliciangeli D, Borges R, Rodriguez N. (2000) The epidemiology and control of leishmaniasis in Andean countries. *Cad. Saude Publica.* 16: 925-950.
8. King RJ, Campbell-Lendrum DH, Davies CR (2004) Predicting geographic variation in cutaneous leishmaniasis, Colombia. *Emer. Inf. Dis.* 10: 598-607.
9. Wijeyaratne PM, Jones Arsenault LK, Murphy CJ (1994) Endemic disease and development: the leishmaniasis. *Acta Trop.* 56: 349-364.
10. Levins R, Lopez C (1999) Toward an ecosocial view of health. *Int. J. Health Serv.* 29: 261-293.
11. Awerbuch T, Kiszewski AE, Levins R (2002) Surprise, nonlinearity and Complex behaviour In: Martens P, McMichael AJ Editors. *Environmental Change, Climate and Health*. Cambridge University Press. pp 96-119.

12. Harrison S (1991) Population growth, land-use and deforestation in Costa Rica, 1950-1984. *Interciencia*. 16: 83-93.
13. Rosero-Bixby L, Palloni A (1998) Population and deforestation in Costa Rica. *Pop & Environ*. 20: 149-185.
14. Rosero-Bixby L, Maldonado-Ulloa T, Bonilla-Carrion R (2002) Forests and population on the Osa Peninsula, Costa Rica. *Rev. Biol. Trop.* 50: 585-598.
15. Ricker WE (1963) Big effects from small causes: 2 examples from fish population dynamics. *J Fish Res Board Can* 20: 257-264.
16. May RM (1977) Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature* 269: 471-477.
17. Scheffer M, Carpenter S, Foley J, Folke C, Walker B (2001) Catastrophic shifts in ecosystems. *Nature* 413: 591-596.
18. Holling CS (1973) Resilience and stability of ecological systems. *Ann Rev Ecol Sys* 4: 1-23.
19. Bonilla-Carrion R, Rosero-Bixby L (2004) Presión demográfica sobre los bosques y áreas protegidas, Costa Rica 2000. In: Rosero-Bixby L Editor. *Costa Rica a la luz del censo del 2000*. San Jose: Centro Centroamericano de Poblacion, pp. 575-594.
20. González Quesada ME (2004) Índice de rezago social. In: Rosero-Bixby L Editor. *Costa Rica a la luz del censo del 2000*. San Jose: Centro Centroamericano de Poblacion, pp. 3-27.
21. Beyer HL (2004) Hawth's Analysis Tools for ArcGIS. Available at <http://www.spatial ecology.com/htools>.
22. Zeledón R, Murillo J, Gutiérrez H (1985). Anthropophilic sandflies and cutaneous leishmaniasis in Costa Rica. *Bol of. Sanit. Panam.* 99: 163-172.
23. Murillo J, Zeledón R (1985) Flebótomos de Costa Rica. Brenesia, Monografía del Museo Nacional de Costa Rica
24. Obando Acuña V (2002) Biodiversidad en Costa Rica. San Jose: INBIO.
25. Kulldorff M (1999) In: J. Glaz and N. Balakrishnan editors. *Scan Statistics and Applications*. Boston: Birkhauser pp. 303-322.

26. Anselin L (1995) Local Indicators of Spatial Association-LISA Geog. Anal. 27: 93-115.
27. Faraway JJ (2005) Linear Models with R. Boca Raton: Chapman Hall/ CRC.
28. Venables WN, Ripley BR (2002) Modern Applied statistics with S. New York: Springer-Verlag.
29. Faraway JJ (2005) Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. Boca Raton: Chapman Hall/ CRC.
30. Schnabel RB, Koontz JE, Weiss BE (1985) A modular system of algorithms for unconstrained minimization. ACM Trans. Math. Software. 11: 419-440.
31. Wise PH, Kotelchuck M, Wilson ML, Mills M (1985) Racial and socioeconomic disparities in childhood mortality in Boston. New Eng J Med 313: 360-366.
32. Celli A (1933) The history of Malaria in the Roman Campagna. London: John Bale, Sons & Danielsson, LTD.
33. Bawa KS, Dayanandan S (1997) Socioeconomic factors and tropical deforestation. Nature 386: 562-563.
34. Jha S, Bawa KS (2006) Population growth, human development, and deforestation in biodiversity hotspots. Conserv Biol 20: 906-912.
35. Fearnside PM (1993) Deforestation in Brazilian Amazonia: the effect of population and land tenure. AMBIO. 22: 537-545.
36. de Castro MC, Monte-Mor RL, Sawyer DO, Singer BH (2006) Malaria risk on the Amazon frontier. Proc Nat Acad Sci 103: 2452-2457.
37. Vandermeer J, Perfecto I (2007) The agricultural matrix and a future paradigm for conservation. Conserv Biol 21: 274-277.
38. Hidalgo HH, Jaramillo A (1977) Contribución a la epidemiología de la Leishmaniasis en Costa Rica. Acta Med Costarricense 20: 83-101.
39. Marramo NN, Mata LJ, Durack DT (1989). Cutaneous Leishmaniasis in rural Costa Rica. Trans Roy Soc Trop Med Hyg 83: 340.

40. Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press.
41. Beisner BE, Haydon DT, Cuddington K (2003) Alternative stable states in Ecology. *Front Ecol Environ* 1: 376-382.
42. Terborgh J, Lopez L, Nunez P, Rao M, Shahabuddin G, Orihuela G, Riveros M, Ascanio R, Adler GH, Lambert TD, Balbas L (2001) Ecological meltdown in predator-free forest fragments. *Science* 294: 1923-1926.
43. Ostfeld RS, Keesing F (2000) Biodiversity and disease risk: The case of lyme disease. *Conserv Biol* 14: 722-728.
44. Allan BF, Keesing F, Ostfeld RS (2003) Effect of forest fragmentation on Lyme disease risk. *Conserv Biol* 17: 267-272.
45. Aguilar CM, Fernández E, Fernández R, Deane LM (1984) Study of an outbreak of cutaneous Leishmaniasis in Venezuela. The role of domestic animals. *Mem Inst Oswaldo Cruz*. 79: 181–195.
46. Alexander B, Lozano C, Barker DC, McCann SHE, Adler GH (1998) Detection of *Leishmania (Viannia) braziliensis* complex in wild mammals from Colombian coffee plantations by PCR and DNA hybridization *Acta Trop* 69: 41-50.
47. Travi BL, Adler GH, Lozano M, Cadena H, Montoya-Lerma J (2002) Impact of habitat degradation on Phlebotominae (Diptera : Psychodidae) of tropical dry forests in northern Colombia. *J Med Entomol* 39: 451-456.
48. Alexander B, Agudelo LA, Navarro F, Ruiz F, Molina J, Aguilera G, Quiñones ML (2001) Phlebotomine sandflies and leishmaniasis risks in Colombian coffee plantations under two systems of cultivation. *Med Vet Entomol*. 15: 364-373.
49. Perfecto I, Rice RA, Greenberg R, Van der Voort ME (1996) Shade coffee: a disappearing refuge for biodiversity. *Bioscience*. 46: 598-608.
50. Lindblade KA, Walker ED, Onapa AW, Katungu J, Wilson ML (2000) Land use change alters malaria transmission parameters by modifying temperature in a highland area of Uganda. *Trop Med Int Health* 5: 263-274.
51. Scorza JV, Rojas E. (1988) Caficultura y leishmaniasis tegumentaria en Venezuela. *Bol Dir Malar San Amb* 28: 114-127.
52. Adler GH (1998) Impacts of resource abundance on populations of a tropical forest rodent. *Ecology*. 79: 242-254.

53. Lima M, Keymer JE, Jaksic FM (1999) El Niño southern oscillation driven rainfall variability and delayed density dependence cause rodent outbreaks in western South America: Linking demography and population dynamics. *Am Nat* 153: 476-491.
54. Davis S, Calvet E, Leirs H (2005) Fluctuating rodent populations and risk to humans from rodent-borne zoonoses. *Vector Borne Zoon Dis* 5: 305-314.
55. Carpenter SR, Brock WA (2006) Rising variance: a leading indicator of ecological transition. *Ecol Lett* 9: 308-315.

CHAPTER VII

CONCLUSIONS

Summary of Major Findings

- Early warning systems for vector-borne diseases are plausible tools. Vector-borne diseases dynamics were shown as predictable within short windows of time, i.e., one year, and predictions were improved by considering exogenous climatic drivers [1,2]. In general, linear models were the best for forecasting future dynamics. The success of forecasts may be linked to the fact that the relationship of Cutaneous Leishmaniasis with climatic drivers was robust during the studied period [2].
- The regulation of the interactions between two co-occurring parasites, *Plasmodium vivax* and *P. falciparum*, causing malaria seems to be under a bottom-up system of regulation. *P. falciparum* is dynamically sufficient while *P. vivax* is positively correlated with *P. falciparum*, a result expected only if the interaction between hosts and their immunity is one where hosts feed their immunity. Additionally, the development of immunity beyond parasite clearance seems to be very short, which justifies the focus of control measures on transmission interruption instead of vaccine development [3]. This result is re-inforced by the success that control

measures targeting transmission have had when compared with the limited success of vaccine trials.

- Relationships between disease dynamics and exogenous drivers can be transient and are sensitive to the context of transmission. Effects of climate change on disease dynamics will be dependent on adaptive measures adopted by human populations. The introduction of insecticide treated bednets for malaria control showed that after 20% of the population of Vanuatu were covered with nets, the mean rate of infection, the impact of exogenous and the unexplained variability was diminished by more than 50 % for both *P. falciparum* and *P. vivax* malaria in Vanuatu [4]. The spatio-temporal patterns of Cutaneous Leishmaniasis in Costa Rica seem to be ultimately defined by the degree of social exclusion of people for a given county. In the marginalized populations, where the burden of this disease is largest, qualitative differences in transmission are influenced by forest cover. In largely deforested areas incidence increases with El Niño Southern Oscillation events, whereas the opposite effect is seen in more forested counties [5].

Suggestions for Future Research

- The models considered here for forecasting incorporate drivers in a linear fashion; however, relationships with climatic drivers can be far from linear and monotonic [e.g., 5]. This pattern calls for the development of statistical techniques for which a measurement similar to cross-correlation would be necessary to identify non-linear relationships with drivers that can be

incorporated through simple non-linear functions or step-wise linear functions as presented by Chaves et al [5]. Insights from past cycles of climate change, especially those associated with glaciations during the Pleistocene, can be useful to move to longer time horizons of prediction, since they are fundamental to understand the dynamics of regime shifts. This step will also require the integration of the evolutionary dynamics of parasites and hosts, given the time scale of such a study.

- The robustness of the results on the population level regulation of the interaction between *P. falciparum* and *P. vivax* can be tested by using models where the abstraction is based on the process of transmission through the force of infection [e.g., 6]. The direction or ability of the interaction to be conditional (i.e., sign changing), can be studied by extending the models and incorporating evolutionary dynamics in the models. The robustness of results also can be assessed by using continuous time modeling tools.
- The dynamics of variability deserve to become an object of study by itself in ecology. Why does variability in state variables increase or decrease with their average value has not been fully addressed in population dynamics. This is a fundamental question since the uncertainty in population dynamics can be much more important than its mean behavior. It is ultimately relevant to understand issues of sustainability, because variability influences the likelihood that a system goes extinct or becomes persistent. This question has special relevance for applied areas of

ecology such as biodiversity conservation or disease transmission management.

- Effects of biodiversity changes on emerging diseases mediated through deforestation need further study, especially for Cutaneous Leishmaniasis. Although Chaves et al [5] show evidence for different effects of ENSO in counties with different degrees of deforestation, the linkage to changes in biodiversity was not addressed [see 7]. This link could be investigated through differences in fragmentation, a robust proxy for changes in biodiversity.
- Ecology of human diseases needs to become fully aware of the socialized nature of its object of study. Although, Chaves et al [5] clearly show that social exclusion, a problem at the root of deforestation (the conventional risk factor for cutaneous leishmaniasis emergence) is a main determinant for the spatial pattern of cutaneous leishmaniasis, this dimension is largely ignored in the study of infectious diseases. Future work on the ecology of human diseases should emphasize the social-cultural component of transmission, since it is at the basis of coarse grained patterns (spatial and temporal), and it has been shown to be fundamental to the understanding of traditional human practices that can guide sound solutions to manage risks for disease transmission (e.g., bednets).

References

1. Chaves LF, Pascual M (2006) Climate Cycles and Forecasts of Cutaneous Leishmaniasis, a Nonstationary Vector-Borne Disease PLoS Medicine. 3: e295

2. Chaves LF, Pascual M (2007) Comparing Models for Early Warning Systems of Neglected Tropical Diseases. PLoS NTDs 1, e33
3. Chaves LF, Kaneko A, Björkman A, Pascual M (submitteda) Random, top-down or bottom-up co-existence of parasites: explaining the dynamics of malaria in multi-parasitic settings
4. Chaves LF, Kaneko A, Taleo G, Pascual M, Wilson ML (submittedb) Malaria transmission pattern resilience to climatic variability is mediated by insecticide treated nets.
5. Chaves LF, Cohen JM, Pascual M, Wilson ML (2008) Social Exclusion modifies climate and deforestation impacts on a Vector-borne disease PLoS NTDs 2, e176
6. Pascual M, Cazelles B, Bouma MJ, Chaves LF, Koelle K (2008) Shifting patterns: malaria dynamics and rainfall variability in an African highland. Proc Biol Sci 275:123-132.
7. Chaves LF, Hernandez M-J, Dobson AP, Pascual M (2007) Sources and sinks: revisiting the criteria for identifying reservoirs for American cutaneous leishmaniasis. Trends. Parasitol. 23, 311-316

APPENDICES

Appendix S1:

Most of the statistical procedures of this paper were implemented in R [s1]. For some specific procedures, additional software is indicated below.

The periodogram and de-trending of the time series:

The periodogram is defined as the squared spectral density of a time series [s2]:

$$I(v_k) = \left| n^{-1/2} \sum_{i=1}^n y_i e^{-2\pi i v_k t} \right|^2 \quad (\text{a1})$$

The periodogram assumes that the series under study is stationary. One way to achieve stationarity is to detrend the time series. Here, we used the method of Discrete wavelet shrinkage. This method is based on the discrete wavelet transform (DWT) which in turn is a data transformation whose algorithm computes the wavelet coefficients of a series using an orthonormal and compactly supported function [s3]. As a result, the method can capture gross features of a series while focusing on finer details when necessary [s4, s5]. The advantage of DWT over other methods such as non-parametric splines or local polynomials, is this emphasis on the localized, as opposed to the global, behavior of the series [s4, s5]. The coefficients of the discrete wavelet transform are set (“shrunk”) to zero if they are smaller than a critical value, determined by the time series length and the variability of the wavelet coefficients for a given scale [s4, s5]. In this study, the Daubechies wavelet basis was used, symmetric

and periodic edges were tested, the series was padded with its mean to ensure a dyadic (power of two) length, and the wavelet filter number was chosen using the basis that produces the median MSE to guarantee that the data is neither overfitted nor underfitted. Wavelets were fitted using the package wavethresh for R.

Maximum entropy spectral density and non-Parametric de-noising of non-stationary time series

We obtained the dominant frequency of the cycles with a second method, maximum entropy spectral density, to examine the robustness of our findings. The maximum entropy spectral density $Y(v_k)$ is applied to identify cycles in time series whose non-stationarity can be approximated by an autoregressive process [s6]. It is computed as follows:

$$Y(v_k) = \sigma_w^2 / \left| 1 + \sum_{j=1}^p \phi_j e^{-2\pi i j v_k} \right|^2 \quad (\text{a2})$$

Again, peaks in this density indicate dominant frequencies. To successfully apply this technique, it is necessary to first separate signal from noise in the data, which we achieved with the two following methods:

1) Smoothing splines. The principle of this method is the minimization of a function that accounts for the trade-off between the fit, measured through the MSE, and λ , the degree of smoothness [s7]:

$$y_t = \mu(x_t) + \varepsilon_t$$

$$\sum \rho_t (y_t - \hat{\mu}(x_t))^2 + \lambda \int [\mu''(x_t)]^2 \quad (\text{a3})$$

where ρ are weights used for robustness to outliers. The parameter λ was selected using generalized cross-validation [s7].

2) Singular spectrum analysis. This non-parametric technique separates trends and oscillatory components from noise in a time series. The method consists in the computation of the eigenvalues and eigenvectors from a covariance matrix $\{M\}$ whose element m_{ij} is the covariance between lags i and j . The projection of the time series on the eigenvectors (the principal components of the matrix) reconstructs the pattern of variability associated with the selected eigenvalue, resulting in a de-noised time series [s6]. The eigenvalues themselves indicate how much variance is accounted for by the different components. For the SSA the toolkit described in [s6] was used.

Wavelet Power Spectrum

The wavelet power spectrum (WPS_y) is obtained from the so-called wavelet transform ($WT_y(s)$). The best way to understand this transform is to consider that it is the product of the data and a function known as a wavelet (ψ) which is different from zero only for a range of values centered around the origin. By systematically translating the wavelet in time and computing this product, we obtain the transform

$$WT_y(s) = \sum_{j=1}^n y_j \psi^*[(j - i\delta t)/s] \tag{a4}$$

as a function of temporal scale s [s8, s9]. By contrast, the periodogram relies on the Fourier transform which uses sinusoidal functions that repeat continuously in

time. This difference underlies the localization in time of the wavelet power spectrum given by the squared wavelet transform

$$WPS_y = |WT_y|^2 \quad (a5)$$

Wavelet Cross-Spectrum and patterns of association in non-stationary time series

Wavelet coherency is computed using the wavelet power spectrum of the two series under study as

$$WCO_{yx} = \frac{|WT_y WT_x^*|}{(WPS_y WPS_x)^{1/2}} \quad (a6)$$

and varies between 0 and 1, with a value of 1 indicating maximum coherency.

The time lag separating the two time series under study can also be determined by computing the phase of the cross wavelet spectrum, defined as the angle separating the real and imaginary parts of the wavelet cross spectrum [s8, s9]:

$$\theta_{yx} = \tan^{-1} \frac{\Im(WT_y WT_x^*)}{\Re(WT_y WT_x^*)} \quad (a7)$$

Linear models and forecasts

One strategy for fitting time series models is to use their state space representation, that is, to find the underlying (not observed) process that produces the observed patterns in the time series [s2, s10].

State Space Representations

Time series can be seen as realizations of an unobserved stochastic process. Any unobserved process $\{y_t\}$ that can be expressed using observation

and state equations has a state space representation [s2]. For the model process $\{y_t\}$ presented in (1) the state space representation can be obtained by introducing the following state vector:

$$x_t = \begin{bmatrix} y_{t-13} \\ y_{t-12} \\ \vdots \\ y_{t-1} \\ y_t \end{bmatrix} . \quad (\text{a8})$$

The observation equation is:

$$y_t = [0 \ 0 \ \dots \ 0 \ 1]x_t \quad (\text{a9})$$

while the state equation is given by:

$$x_{t+1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \mu \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \ddots & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ \phi_1 \phi_{12} & \phi_{12} & 0 & \dots & 0 & \phi_1 \end{bmatrix} x_t + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \alpha_1 \end{bmatrix} T_{t-4} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \gamma_a \end{bmatrix} MEI_{t-13} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \varepsilon_t \quad (\text{a10})$$

where the error is identically and normally distributed, i.e., $\varepsilon_t \sim N(0, \sigma_w^2)$. For this unobserved process to have the same characteristics than those of the observed time series, it is necessary to address: (i) the correlation of the data over time (the smoothing problem), (ii) the fitting of every observation (the filtering problem) and (iii) the forecasting of future events (the prediction problem). These problems can be solved using Kalman Recursions [s2, s10]. The exact likelihood is computed via a state-space representation of the SAR process, and the innovations (i.e., residuals) and their variance found by a Kalman filter [s10].

References

- s1 R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- s2 Brockwell PJ, Davis RA (2002) Introduction to Time Series and Forecasting. 2nd ed. New York: Springer.
- s3 Gençay R, Selçuk F, Whitcher B (2002) An introduction to wavelets and other filtering methods in finance and economics. San Diego: Academic Press.
- s4 Shumway RH, Stoffer DS (2000) Time series analysis and its applications. New York: Springer. 572 p.
- s5 Faraway JJ (2006) Extending the linear model with R: Generalized Linear, Mixed Effects and Non-parametric regression models. Boca Raton: Chapman and Hall CRC press.
- s6 Ghil M, Allen RM, Dettinger DM, Ide K, Kondrashov D, Mann ME, Robertson A, Saunders A, Tian Y, Varadi F, Yiou P. (2002) Advanced spectral methods for climatic time series. Rev Geophys 40: 1-41
- s7 Venables WN, Ripley BD (2002) Modern applied statistics with S. New York: Springer. 495 p.
- s8 Maraun D Kurths J (2004) Cross wavelet analysis: significance testing and pitfalls. Nonlinear Proc Geoph. 11: 505-514
- s9 Torrence C, Compo G (1998) A practical guide to wavelet analysis. Bull Am Meteor Soc 79: 61-78.
- s10 Durbin J, Koopman SJ (2001) Time Series Analysis by State Space Methods. Oxford: Oxford University Press.

Appendix S2:

Linear Models

Seasonal Autoregressive (SAR): these models were fitted to the data using Kalman recursions for their state space representation [see s1, for technical details]. These models incorporate seasonality by considering autoregressive components with the period of the time series under study. For the purpose of comparison, the following three models published in Chaves & Pascual [s1] were considered:

$$\begin{aligned}y_t &= \mu + \phi_1 (y_{t-1} - \mu') + \phi_{12} (y_{t-12} - \mu') - \phi_1 \phi_{12} (y_{t-13} - \mu') + \alpha T_{t-4} + \gamma MEI_{t-13} + \varepsilon_t \\y_t &= \mu + \phi_1 (y_{t-1} - \mu') + \phi_{12} (y_{t-12} - \mu') - \phi_1 \phi_{12} (y_{t-13} - \mu') + \gamma MEI_{t-13} + \varepsilon_t \\y_t &= \mu + \phi_1 (y_{t-1} - \mu) + \phi_{12} (y_{t-12} - \mu) - \phi_1 \phi_{12} (y_{t-13} - \mu) + \varepsilon_t\end{aligned}\tag{1}$$

Where the error terms are assumed to be identical, independent and normally distributed (i.e., i.i.d. normal, $\varepsilon \sim N(0, \sigma^2)$) and $\mu' = \mu +$ the effects of covariates ($\gamma MEI, \alpha T$) when considered. These models were fitted using R [s2].

Basic Structural Model (BSM): The BSM is one of the simplest linear gaussian state space models. It decomposes an observed time series (y_t) into a local level (μ_t), which basically is a changing mean value or the equivalent to an intercept in

least squares regression, a trend (v_t), and seasonal $\left(\gamma_t, \sum_{k=0}^{11} \gamma_{t-k} = 0 \right)$ components:

$$\begin{aligned}
y_t &= \mu_t + \gamma_t + \varepsilon_t; \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \\
\mu_{t+1} &= \mu_t + \nu_t + \eta_t; \eta_t \sim N(0, \sigma_\eta^2) \\
\nu_{t+1} &= \nu_t + \zeta_t; \zeta_t \sim N(0, \sigma_\zeta^2) \\
\gamma_{t+1} &= -\sum_{k=0}^{10} \gamma_k + \omega_t; \omega_t \sim N(0, \sigma_\omega^2)
\end{aligned} \tag{2}$$

The equations are solved and smoothed using Kalman recursions. The parameters are the variances of the errors in the four equations, which are assumed to be i.i.d. normal [for details, s3]. The model was also fitted using R [s2].

Non-Linear Models

Generalized Additive Models (GAM): Additive models are a combination of parametric and non parametric models [s4,s5]. Unlike linear models, where linear parametric forms define the shape of the relationship between responses and predictors, in GAM smooth functions (f) that can be far from linear are used for such relationships. In this work, the smooth non-parametric functions were computed using a penalized smoothing spline approach in which the parameters for smoothing were obtained by generalized cross-validation of a function that weights the trade-off between the smoothing and the likelihood of the fitting [for technical details see s5]. An intensive process of model selection, based on likelihood tests and the Akaike Information Criterion [e.g., s1], led to the selection of the following two best models:

$$y_t = \mu + \varphi_1 y_{t-1} + f\left(y_{t-12}\right) + f\left(y_{t-13}\right) + \alpha T_{t-4} + \gamma MEI_{t-13} + \varepsilon_t \quad (3)$$

$$y_t = \mu + \varphi_1 y_{t-1} + f\left(y_{t-12}\right) + f\left(y_{t-13}\right) + \gamma MEI_{t-13} + \varepsilon_t$$

For the models in (3) the assumptions about the error in (1) hold. These models were fitted using the library mgcv of R [s2]

Feed-forward Neural Networks (FNN): Feed-forward neural networks are computer-based models that try to emulate the human brain in performing complex tasks [s4, s5]. In the application of these models to time series analysis, there is no mechanistic interpretation of the layers of neurons. These simply provide a phenomenological and flexible treatment of functional relationships between predictors and responses. In models with one hidden layer of “neurons”, an arbitrary functional form is decomposed into a sum of sigmoids [s6]:

$$f(y_1, y_2, \dots, y_d) = \beta_0 + \sum_{i=1}^k \beta_i G(\mu_i + \sum_{j=1}^d \gamma_{ij} x_j) \quad (4)$$

Where G is a univariate sigmoid function like the logistic expression $e^u/(1+e^u)$. Models with the same predictors as in (1) were fitted and tested using up to 3 neurons. These models do not have explicit assumptions about the errors. Parameters (weights in the FNN’s jargon) are obtained with a search that minimizes residuals over 100000 experimental fittings [s5]. These models were fitted using the library nnet of R [s2]. The goodness of fit was highest for models with the same predictors as the ones described in (1).

Non-linear Forecasting (NLF): Though originally used to assess the degree of determinism in the dynamics of populations [s7], NLF is ultimately a forecasting tool. It is a technique based on the multidimensional embedding of a time series, where the E-dimensional set of points $x_t = (y_t, y_{t-\tau}, \dots, y_{t-(E-1)\tau})$ is constructed and called the phase space, and τ represents the time delay in the observations [e.g. s7, s8]. NLF is implemented by first identifying neighboring points of x_t in delayed embedding space, and then obtaining a prediction of the response variable at time $t+k$ as simply the average over the future state of the neighbors [s7]:

$$\hat{x}_{t+k} = \frac{1}{u_t} \sum_{x_t \in u_t} x_{t+k} \quad (5)$$

A key step in using NLF is to choose the appropriate dimension E. One approach is to use several dimensions in order to find the value of E that optimizes predictions [s7]. Here we used as a guide for an initial value of E, results from the false nearest neighbors method [s9]. This method finds the minimum embedding dimension (E), by computing the minimum number of multidimensional points that are erroneously mapped on the neighborhoods of other points when the value of E is diminished from an initial value. For the time delay τ , the most common practices are to assign time delays (τ) of 1 [s7,s10]. Here, a criterion based on time delayed average mutual information (AMI) was used [s11]. This function accounts for linear and non-linear correlations in a time series and is computed as follows:

$$AMI = -\sum_{ij} p_{ij}(\tau) \ln \frac{p_{ij}(\tau)}{p_i p_j} \quad (6)$$

Where p_i is the probability to find a time series value in the i -th interval of some partition of the data, $p_{ij}(\tau)$ is the joint probability that an observation falls into the i -th interval and τ times units later, in the j -th interval. The AMI and the false nearest neighbors method were implemented using the library `tseriesChaos` for R [s2] and the NLF's were computed using the package `predict` of TISEAN s8.

Choosing the Embedding dimensions and lags for the Non-Linear Forecasting

The false nearest neighbors approach suggests embedding dimensions of two, three or four, since larger values do not lead to a considerable decrease in the % of false neighbors, indicating that state-space is successfully unfolded for lower values of E . The AMI method suggests that good lags for the delay τ in the NLF are 1, 5 and 10. Given the limitation of time series length, we chose to set $\tau=1$ and to consider values of two, three and four for the embedding dimension.

References

- s1. Chaves LF, Pascual M (2006) Climate cycles and forecasts of cutaneous leishmaniasis, a nonstationary vector-borne disease. *PLoS Med* 3: e295.
- s2. R Development Core Team. (2006) R: A language and environment for statistical computing. R Foundation for Statistical Computing: Vienna.
- s3. Durbin J, Koopman SJ (2001) Time series analysis by state space methods. Oxford: Oxford University Press.
- s4. Venables W, Ripley B (2002) Modern applied statistics with S 4th ed. New York: Springer.

- s5. Faraway JJ (2005) Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. Boca Raton: Chapman & Hall. CRC.
- s6. Ellner S, Gallant AR, Theiler J (1995) Detecting nonlinearity and chaos in epidemic data. Epidemic models, their structure and relation to data (D. Mollison Ed), pp, 229-247. Cambridge: Cambridge University Press.
- s7. Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature. 344: 734-741.
- s8. Hegger R, Kantz H, Schreiber T (1999) Practical implementation of nonlinear time series methods: the TISEAN package. Chaos 9: 413-435.
- s9. Kennel MB, Brown R, Abarbanel HDI (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A 45: 3403-3411.
- s10. Grenfell BT, Kleczkowski A, Ellner SP, Bolker BM (1994) Measles as a case study in nonlinear forecasting and chaos. Phil Trans Roy Soc London A. 348: 515-530.
- s11. Fraser AM, Swinney HL (1986) Independent coordinates for strange attractors from mutual information. Phys Rev A 33: 1134-1140.

Appendix S3. Model Selection and Parameter Values for models of *Plasmodium falciparum* and *Plasmodium vivax* rates before and after the breakpoint found using the *F* statistics

Parasite/ Breakpoint	$\hat{\phi}_1$	$\hat{\phi}_{12}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\mu}$	$\hat{\sigma}_\varepsilon$	P_{χ^2}	d.f.	Akaike Information Criterion	
<i>Plasmodium falciparum</i>	Before Breakpoint	0.72 ± 0.06	0.36 ± 0.10	1.38 ± 0.42	1.44 ± 0.46	0.10 ± 0.51	0.83 ± 0.57	11.51 ± 1.09	2.38	—	—	603.32
		0.70 ± 0.07	0.34 ± 0.10	1.41 ± 0.42	1.35 ± 0.44	0.13 ± 0.52	—	11.56 ± 1.02	2.40	0.14	1	603.42
		0.72 ± 0.06	0.36 ± 0.10	1.39 ± 0.42	1.45 ± 0.45	—	0.84 ± 0.57	11.51 ± 1.08	2.37	0.83	1	601.37
		0.74 ± 0.06	0.48 ± 0.08	1.28 ± 0.46	—	0.39 ± 0.61	0.26 ± 0.51	11.34 ± 1.48	2.44	<0.001	1	610.09
		0.74 ± 0.06	0.43 ± 0.09	—	1.28 ± 0.49	0.28 ± 0.52	0.87 ± 0.57	11.34 ± 1.37	2.46	<0.005	1	611.18
		0.70 ± 0.07	0.33 ± 0.10	1.43 ± 0.42	1.36 ± 0.44	—	—	11.56 ± 1.02	2.40	0.33	2	601.48
		0.74 ± 0.06	0.48 ± 0.08	1.30 ± 0.46	—	—	0.68 ± 0.57	11.34 ± 1.47	2.45	<0.005	2	608.35
		0.74 ± 0.06	0.43 ± 0.09	—	1.30 ± 0.48	—	0.88 ± 0.57	11.34 ± 1.37	2.46	<0.01	2	609.48
	After Breakpoint	0.84 ± 0.06	-0.01 ± 0.12	0.65 ± 0.33	1.11 ± 0.31	0.70 ± 0.55	0.89 ± 0.42	4.07 ± 1.26	1.89	—	—	329.76
		0.82 ± 0.06	-0.01 ± 0.12	0.63 ± 0.35	0.97 ± 0.32	0.63 ± 0.57	—	3.98 ± 1.15	1.94	<0.05	1	332.00
		0.83 ± 0.06	0.04 ± 0.12	0.75 ± 0.33	1.07 ± 0.33	—	0.86 ± 0.43	3.98 ± 1.25	1.91	0.75	1	329.35
		0.82 ± 0.02	0.15 ± 0.12	0.52 ± 0.39	—	0.39 ± 0.59	0.68 ± 0.46	3.91 ± 1.39	2.02	<0.005	1	337.76
		0.84 ± 0.06	0.05 ± 0.12	—	1.05 ± 0.33	0.90 ± 0.56	0.86 ± 0.43	4.01 ± 1.37	1.93	0.06	1	331.27
		0.81 ± 0.06	0.04 ± 0.12	0.72 ± 0.35	0.94 ± 0.33	—	—	3.90 ± 1.15	1.96	0.07	2	331.19
0.82 ± 0.06		0.17 ± 0.11	0.57 ± 0.39	—	—	0.68 ± 0.46	3.87 ± 1.41	2.02	<0.05	2	366.24	
0.83 ± 0.06		0.14 ± 0.11	—	0.96 ± 0.35	—	0.81 ± 0.45	3.87 ± 1.43	1.96	0.11	2	366.91	

Parasite/ Breakpoint	$\hat{\phi}_1$	$\hat{\phi}_{12}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\mu}$	$\hat{\sigma}_\varepsilon$	P_{χ^2}	d.f.	Akaike Information Criterion	
<i>Plasmodium</i> . <i>vivax</i>	Before Break- point	0.81 ± 0.05	0.26 ± 0.09	0.47 ± 0.18	—	-0.38 ± 0.22	—	4.88 ± 0.61	1.07	—	—	411.31
		0.80 ± 0.05	0.27 ± 0.08	0.55 ± 0.18	—	—	—	4.83 ± 0.62	1.08	0.09	1	412.18
		0.81 ± 0.05	0.31 ± 0.09	—	—	-0.51 ± 0.22	—	4.83 ± 0.66	1.09	<0.01	1	415.94
	After Break- point	0.71 ± 0.09	0.36 ± 0.15	0.11 ± 0.17	—	-0.13 ± 0.18	—	2.33 ± 0.42	0.73	—	—	166.98
		0.72 ± 0.08	0.36 ± 0.13	0.11 ± 0.16	—	—	—	2.33 ± 0.43	0.73	<0.02	1	166.75
		0.71 ± 0.09	0.40 ± 0.13	—	—	-0.14 ± 0.18	—	2.32 ± 0.45	0.73	<0.02	1	165.35

The parameters for the full model are in the first data row, and for the null model are in the last data row. Parameters are described in the text and are given as value ± standard error. P_{χ^2} is the significance of the chi-squared likelihood ratio test between each model and the full model, and d.f. its degrees of freedom.

Appendix S4

Principal Component Analysis (PCA) and Multidimensional Scaling (MDS).

These techniques were used to study the sand fly associated landscape composition of counties. Both PCA and MDS are dimension reduction techniques to identify main axes in the variability of multidimensional data sets. PCA is computed by finding the eigenvalues and respective eigenvectors of the variance-covariance matrix of a multidimensional dataset, producing scores (components) for each individual object in the dataset based on linear combinations of the variables [s1,s2]. In contrast, MDS is a different approach where distances among objects are computed using any of several different measures, returning coordinates for the points on the number of chosen dimensions for the analysis [s1]. For the MDS, Euclidean distances were used. Goodness of fit was measured using as a loss function the least squares on distances, or STRESS [s3]. For both techniques data on the proportion of county landscape cover associated to sand fly presence were normalized.

Generalized Additive Models (GAM). Additive models are a combination of parametric and non-parametric models [s2,s4]. Unlike linear models, where only linear parametric forms define the shape of the relationship between responses and predictors, GAM smooth functions, $s()$, that may be far from linear, can be used for such relationships. In this study, the smooth, non-parametric functions were computed using a penalized smoothing spline approach in which the parameters for smoothing were obtained by generalized cross-validation, using a

function that weights the trade-off between the smoothing and the likelihood of the fitting [s4]. An intensive process of model selection, based on the Akaike Information Criterion (AIC) and backward elimination [s5], led to the selection of the following best model:

$$\text{Rate}_i = \mu_0 + s(\text{ME}_i) + s(\text{MI}_i) + s(\% \text{Close}_i) + s(\log(\text{MinRfl}_i)) + \varepsilon \quad (\text{a.1})$$

where Rate_i denotes the natural log of the five-year disease incidence rate ($\ln(\text{cases}/\text{population size})$), ME_i , the minimum elevation, MI_i , the marginalization index, $\% \text{close}_i$, the percent of people living in a radius of 5 km to the forest edge, and MinRfall_i , the natural log of average annual minimum rainfall. The index i indicates the county and ε is the error which is assumed to be identical, independent and normal (i.e., $\varepsilon \sim N(0, \sigma^2)$). To handle the problem of logarithms for values of 0, we added 1 to all Rate values in (a.1). For this and subsequent models, unless otherwise indicated, all 81 counties were considered. Diagnostics for spatial autocorrelation were carried out by regressing residuals on the centroids of each county.

Linear Mixed Effects Models (LMEM). Mixed effects models can consider covariates as fixed effects or random effects. Fixed effects are unknown constants, while random effects are random variables [s4]. As a result, a parameter for a fixed effect measures the mean effect in a response by unit or category change in a covariate, while for random effects they measure the

variability due to a given covariate. We used these models to test for the spatial scale of highest variability in the political subdivisions of the country that included seven provinces subdivided in a total of 81 counties. For this analysis, we only considered the 59 counties where disease was present with >2 cases during the 5 year study period. Although this procedure may bias the results of the analysis, we considered it to be the most robust option to gain insights about the geopolitical scale of spatio-temporal variability in the data. This decision to only include a subset of the counties was taken because of the superior reliability of LMEM over their corresponding generalized versions, for which no maximum likelihood estimators have been derived [s4], and the potential artifacts in the scales of variability due to the abundance of 0 values under the assumptions of linear models. The response in the model was the annual incidence (*Cases* in the models to follow), defined as the yearly total number of cases for a county, with the regression weighted by the total population in the county. These models were also used to test for an effect of the El Niño Southern Oscillation (ENSO) for the whole country by introducing a continuous predictor varying from 0 to 2, indicating the different phases of ENSO: 0 for years non-El Niño years (1997, 2000), 1 for the El Niño year (1998) and 2 for the year after this event (1999). Data from 1996 were excluded because they were lost as an autoregressive component in the response. This strategy was implemented to economize degrees of freedom (1 as opposed to the 3 needed by using a categorical predictor). Models were fitted by using restricted maximum likelihood estimators

(REML) and compared through a parametric bootstrap [4]. Four models were considered:

$$\log(\text{Cases}(t)_{kj} + 1) = \mu + \log(\text{population}) + \phi \log(\text{Cases}(t-1)_{kj} + 1) + \beta \text{ENSO}(t) + \sigma(t)_j + \delta(t)_{kj} + \varepsilon(t) \quad (\text{a.2})$$

$$\log(\text{Cases}(t)_{kj} + 1) = \mu + \log(\text{population}) + \phi \log(\text{Cases}(t-1)_{kj} + 1) + \sigma(t)_j + \delta(t)_{kj} + \varepsilon(t) \quad (\text{a.3})$$

$$\log(\text{Cases}(t)_k + 1) = \mu + \log(\text{population}) + \phi \log(\text{Cases}(t-1)_k + 1) + \delta(t)_k + \varepsilon(t) \quad (\text{a.4})$$

$$\log(\text{Cases}(t)_k + 1) = \mu + \log(\text{population}) + \phi \log(\text{Cases}(t-1)_k + 1) + \varepsilon(t) \quad (\text{a.5})$$

The comparison between models a.2 and a.3 allows a test of the null hypothesis that ENSO has no effect at the country scale, and the comparison between models a.3 and a.4 for the null hypothesis that effects due to the geopolitical hierarchy (counties belonging to provinces) is irrelevant. The comparison between models a.4 and a.5 allows a test of the existence of a source of common variability across the counties or localized variability at the county level. Models are autoregressive (ϕ), with σ and δ representing province and county level variability, t the time, k the county, and kj the county k belonging to province j . σ , δ , ε are assumed to be i.i.d. normal. Diagnostics for spatial autocorrelation were carried out by regressing residuals on the centroids of each county.

References

- s1. Borg I, Groenen P (1997) Modern multidimensional scaling: theory and applications. New York: Springer-Verlag.
- s2. Venables WN, Ripley BR (2002) Modern Applied statistics with S. New York: Springer-Verlag.
- s3. Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika. 29: 1-27.

s4. Faraway JJ (2005) Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. Boca Raton: Chapman Hall/ CRC.

s5. Faraway JJ (2005) Linear Models with R. Boca Raton: Chapman Hall/ CRC.