



**The Abdus Salam  
International Centre for Theoretical Physics**



2022-51

## **Workshop on Theoretical Ecology and Global Change**

*2 - 18 March 2009*

### **Shifting patterns: malaria dynamics and rainfall variability in an African highland**

PASCUAL SPADA Maria Mercedes\* and other authors

*University of Michigan  
Department of Ecology and Evolutionary Biology  
2019 Natural Science Bldg.  
830 North University Ave., Ann Arbor MI 48109-1048  
U.S.A.*

# Shifting patterns: malaria dynamics and rainfall variability in an African highland

M. Pascual<sup>1,2,\*</sup>, B. Cazelles<sup>3,4</sup>, M. J. Bouma<sup>5</sup>, L. F. Chaves<sup>1</sup> and K. Koelle<sup>1,†</sup>

<sup>1</sup>*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109-1048, USA*

<sup>2</sup>*Santa Fe Institute, Santa Fe, NM 87501, USA*

<sup>3</sup>*CNRS UMR 7625, Ecole Normale Supérieure, 46 rue d'Ulm, 75230 Paris, France*

<sup>4</sup>*IRD UR GEODES, 93143 Bondy, France*

<sup>5</sup>*Department of Infectious and Tropical Diseases, LSHTM, Keppel Street, London WC1E 7HT, UK*

The long-term patterns of malaria in the East African highlands typically involve not only a general upward trend in cases but also a dramatic increase in the size of epidemic outbreaks. The role of climate variability in driving epidemic cycles at interannual time scales remains controversial, in part because it has been seen as conflicting with the alternative explanation of purely endogenous cycles exclusively generated by the nonlinear dynamics of the disease. We analyse a long temporal record of monthly cases from 1970 to 2003 in a highland of western Kenya with both a time-series epidemiological model (time-series susceptible–infected–recovered) and a statistical approach specifically developed for non-stationary patterns. Results show that multiyear cycles of malaria outbreaks appear in the 1980s, concomitant with the timing of a regime shift in the dynamics of cases; the cycles become more pronounced in the 1990s, when the coupling between disease and rainfall is also stronger as the variance of rainfall increased at the frequencies of coupling. Disease dynamics and climate forcing play complementary and interacting roles at different temporal scales. Thus, these mechanisms should not be viewed as alternative and their interaction needs to be integrated in the development of future predictive models.

**Keywords:** malaria dynamics; climate variability; exogenous and endogenous cycles

## 1. INTRODUCTION

Patterns of malaria resurgence have been documented for several African highlands, raising pressing questions on the underlying causes of increases in disease incidence in these transition regions (Marimbu *et al.* 1993; Loevinsohn 1994; Kilian *et al.* 1999; Lindblade *et al.* 1999; Shanks *et al.* 2000; Hay *et al.* 2002a). Importantly, these changing patterns involve not only a general upward trend in cases but also a dramatic increase in the size of epidemic outbreaks over time. Thus, an understanding of malaria dynamics in these regions requires consideration of both epidemic cycles and longer-term trends, two aspects of change that are typically considered separately.

In particular, the year-to-year variation in seasonal epidemics, known as interannual variability, has been studied for different African highlands. It has been shown that cycles of periodicity longer than 1 year are present and it has been proposed that these cycles are generated by the dynamics of the disease itself (Hay *et al.* 2000). The argument rests on the observation that the dynamics of simple epidemiological models, known as SIR for susceptible, infected and recovered population numbers, with parameters plausible for malaria, have an endogenous period of *ca* 3 years consistent with that

observed for a time series of cases in Kenya (Hay *et al.* 2000). Although these intrinsic cycles decay to equilibrium in a purely deterministic model, they are well known to persist in the presence of noise (e.g. Alonso *et al.* 2007). Because they result from the population dynamics of transmission and immunity, they do not require an exogenous driver, such as climate fluctuations. In contrast, a more recent study of temporal patterns across multiple highlands concluded that temperature and rainfall play an important role in the interannual variability of malaria (Zhou *et al.* 2004). In this case, a time-series modelling approach was used that is fully non-mechanistic and, therefore, not directly interpretable in the light of the processes that underlie disease dynamics. As a result, these two opposite explanations for malaria's multiyear cycles are difficult to compare, and have been the subject of an unresolved debate on the role of climate interannual variability in the epidemic patterns of the disease (Hay *et al.* 2005). This debate has also questioned the evidence for a significant increase in rainfall variability itself in recent times, which underscores that disease cycles must increasingly be understood in the context of non-stationary patterns in both disease dynamics and environmental conditions.

In this paper, we re-examine the question of malaria's multiyear cycles with both epidemiological models and statistical approaches specifically developed to handle non-stationary patterns. We analyse a long temporal record of malaria cases in the past three decades in a highland of western Kenya. We show that both previously proposed hypotheses, endogenous dynamics and exogenous environmental factors, play a role in the temporal dynamics of the disease, but do so at different temporal scales. Specifically,

\* Author and address for correspondence: Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109-1048, USA (pascual@umich.edu).

† Present address: Department of Biology, Duke University, Durham, NC 27708, USA.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2007.1068> or via <http://www.journals.royalsoc.ac.uk>.

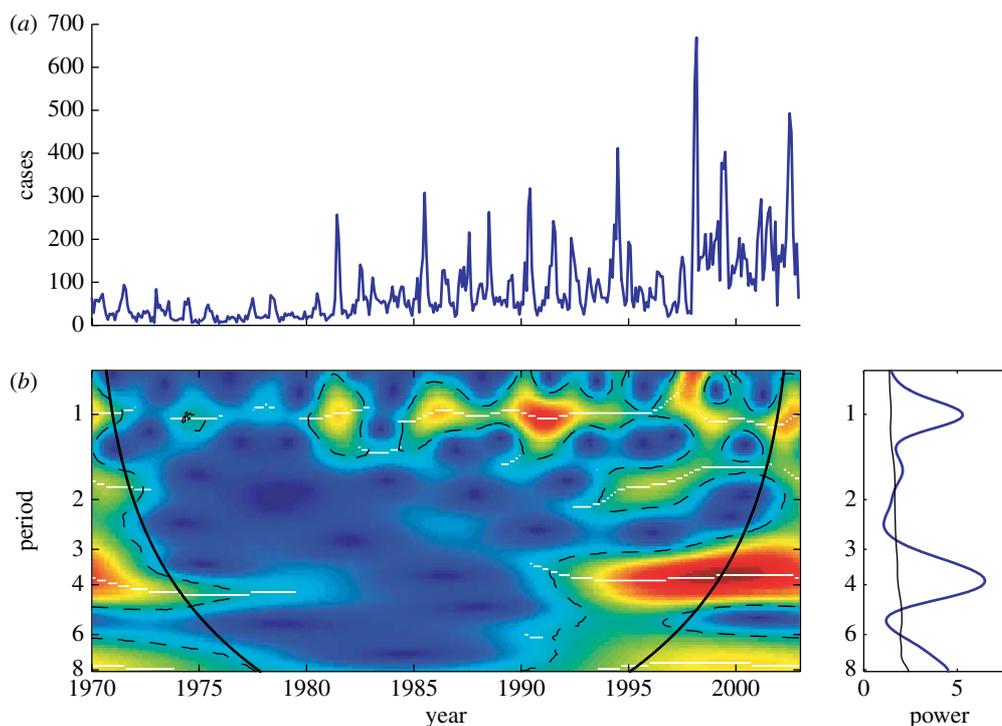


Figure 1. (a) Time series of cases at AHP. (b) The wavelet spectrum is shown with power increasing from blue to red colours. The continuous line corresponds to the 5% significance level (§3). The areas within this line indicate significant variability at the corresponding periods and times. A cycle of period four is apparent in the 1990s, accompanied by a shorter cycle of period between one and two. This frequency is present earlier and is already significant in the 1980s, as best seen by following the crest of the spectrum (white line) indicating the localization of local maxima in time and period. The bold continuous line is known as the cone of influence and delimits the effect of the treatment of the boundaries; only patterns within this line are therefore considered reliable. The malaria data have been square-root transformed to rescale the variance.

rainfall exhibits a clear association with cycles of *ca* 2 and 3 years, while disease dynamics generates cycles of longer period. An interaction between these two factors, rainfall forcing and disease dynamics, is also described. Malaria's multiyear cycles become apparent in the 1980s and are most pronounced in the 1990s, when the coupling between disease and rainfall is also most evident and the variance of rainfall at the frequencies of coupling becomes stronger. These changes are concomitant with the timing of a regime shift, a break point, in the dynamics of malaria at the beginning of the 1980s, suggesting that the association between malaria and rainfall is in part responsible for this qualitative change in disease dynamics. However, a long-term trend in transmission identified by the simple SIR model indicates that other mechanisms are also at play. The timing and possible roles of drug resistance, climate change and land-use practices are discussed. A seasonal mechanism for the role of rainfall is identified in which the 'short' rain season plays a major role by influencing cases in the following year.

## 2. RAINFALL AND MALARIA DATA

The malaria data consist of a monthly time series described in Shanks *et al.* (2005b) from a tea estate in western Kenya known as the African Highland Produce (now the Findlay Farms). These records correspond to the confirmed cases from 1970 to 2003 from positive blood slides for symptomatic inpatients and outpatients, recorded prospectively on a weekly basis. The plantation itself consists of approximately 20 individual estates each with a group of 500–1000 workers, plus their dependents.

We have aggregated the cases monthly for the analyses of interannual variability (figure 1a). We refer to this time series as AHP, for African Highland Produce.

There have been changes in computing the weekly cases from the hospital and outpatient records over time, which may have introduced a source of measurement error. For this reason, a second malaria time series at an adjacent plantation was also considered (this has already been the subject of previous studies, e.g. Hay *et al.* 2000), as a basis for comparison and discussion (figure S5 in the electronic supplementary material). We refer to this time series as BBK, for Brooke Bond Farms. Although the medical care is from a separate system, it operates with very similar policies. The local populations of the two plantations are of similar size, but BBK consists of fewer cases because the records do not include outpatients. The tea estates provide health care for all employees and their dependents on the plantations; this is unlikely to result in accessibility biases which can be expected from ordinary health centre or hospital records.

The rainfall data consist of three monthly time series for local meteorological stations provided by the Kenyan Meteorological Department. (Figure S1 in the electronic supplementary material.) The three stations are: Hail Research Station (0°22' S, 35°16' E; altitude 6483 ft), Kericho Chagaik Estate 0°20' S, 35°20' E; 6000 ft) and Kaisugu Kericho (0°19' S, 35°22' E; 7000 ft). The first time series starts in 1980, while the other two are available from 1970. Therefore, we analyse an average monthly time series obtained from the three datasets (1980–2003) and the individual datasets (1970–2003). (See the legend of figure S1 in the electronic supplementary material for the treatment of missing data.)

### 3. MATERIAL AND METHODS

#### (a) Wavelet power spectrum

In contrast to Fourier analysis, wavelet analysis is well suited for the study of signals whose spectra change with time. This time–frequency analysis of the signal provides information on the different frequencies (i.e. the periodic components) as time progresses (Torrence & Compo 1998; Cazelles *et al.* 2007). See Cazelles *et al.* (2007) for a detailed description of the method and a summary of applications to disease and ecological data.

The wavelet power spectrum (WPS) estimates the distribution of variance between frequency  $f$  and different time locations  $\tau$ . To compare the WPS with classical spectral methods, the global wavelet spectrum is computed as time average of the WPS for each frequency component.

#### (b) Wavelet cross-coherence

As given in Fourier analysis, the univariate WPS can be extended to quantify statistical relationships between two time series  $x(t)$  and  $y(t)$  by computing the wavelet coherence

$$R_{x,y}(f, \tau) = \frac{|\langle W_{x,y}(f, \tau) \rangle|}{|\langle W_x(f, \tau) \rangle|^{1/2} \cdot |\langle W_y(f, \tau) \rangle|^{1/2}},$$

where  $\langle \rangle$  indicates smoothing in both time and frequency;  $W_x(f, \tau)$  is the wavelet transform of series  $x(t)$ ;  $W_y(f, \tau)$  is the wavelet transform of series  $y(t)$ ; and  $W_{x,y}(f, \tau) = W_x(f, \tau) \cdot W_y^*(f, \tau)$  is the cross-WPS. The wavelet coherence provides local information about where two non-stationary signals,  $x(t)$  and  $y(t)$ , are linearly correlated at a particular frequency (or period).  $R_{x,y}(f, \tau)$  is equal to 1 when there is a perfect linear relationship at a particular time and frequency between the two signals (Cazelles *et al.* 2007).

#### (c) Wavelet significance

As with other time-series methods, it is crucial to assess the statistical significance of the patterns exhibited by the wavelet approach. To this end, bootstrap methods have been used to quantify the statistical significance of the computed patterns. The idea is to construct, from observed time series, control datasets that share some properties with the original series, but which are constructed under the following null hypothesis: *the variability of the observed time series or the association between two time series is no different to that expected from a purely random process.* The construction of our control datasets was performed by classical bootstrap (Efron & Tibshirani 1994). By applying the wavelet analysis to this control dataset, we estimated the probability distribution of the wavelet quantities of interest under  $H_0$ . Then, the original values computed from the raw series can be compared with these distributions under the null hypothesis, to extract, for instance, the 99th or the 95th quantiles of these distributions.

#### (d) Regime shift

Methods to detect regime shifts seek to identify a point in time that separates the time series into two segments with different qualitative dynamics. Under the null hypothesis of no break point ( $H_0$ ), a single model is appropriate to capture the dynamics of the whole time series. The alternative hypothesis ( $H_1$ ) consists of two models that differ before and after a break point  $m$ . A recently proposed approach applied to fisheries relies on fitting a vector autoregressive model to a multidimensional system for which multiple time series for interacting variables are available (Solow & Beet 2005). Because a single time

series is available here and seasonality is clearly important for infectious diseases such as malaria, we used instead a seasonal autoregressive model (SAR; Brockwell & Davis 2002). Under  $H_0$ , we fit the model to the whole time series, selecting the order of the autoregressive and seasonal parts using the Akaike information criterion. Diagnostics and problems with overfitting are considered when selecting these orders. Under  $H_1$ , we considered break points occurring at all possible months. For each considered break point  $m$ , we fit two SAR models, one using the time-series data before the break point and the other after the break point. The log likelihood of the two model fit was computed for each break point  $m$  (figure S2 in the electronic supplementary material). To test  $H_0$  against  $H_1$  (for a specific break point  $m$ ), we used the likelihood ratio (LR) statistic, defined as minus twice the difference between the maximum log likelihood under  $H_0$  and that under  $H_1$ . A parametric bootstrap test was used to determine the significance of this statistic (Efron & Tibshirani 1994; Solow & Beet 2005), with the distribution of the LR statistic generated from repeated fits of the null and alternative models to simulated data produced with the model fitted under  $H_0$ . The significance level is given by the proportion of simulations for which the LR statistic is larger than that observed for the data (Solow & Beet 2005).

#### (e) The time-series susceptible–infected–recovered model

Time-series susceptible–infected–recovered (TSIR) models for infectious diseases consist of two main components. The first is a procedure to reconstruct the time series of susceptibles and the second is a transmission equation (Finkenstadt & Grenfell 2000; Koelle & Pascual 2004). Our model here is a simplification of the TSIRS (Time Series Susceptible–Infectious–Recovered–Susceptible) model in Koelle & Pascual (2004), originally formulated for diseases with temporary immunity. Here, we consider that there is no loss of immunity and that the total population is constant in time with a constant turnover time  $T$  of individuals in the tea estate. Under the latter assumption, the reconstruction of susceptibles  $S_t$  is straightforward

$$S_t = S_{t-1} - C_t + B - D \frac{S_{t-1}}{N}, \quad (1)$$

where  $C_t$  is the number of cases at time  $t$ ; the constant  $D$  is the number of total deaths per time interval obtained as  $N/T$ ; and  $B$  is the number of births per time interval, equal to  $D$ , since the total population size  $N$  is constant. We assume that the initial fraction of susceptible individuals is 1 consistent with the observations of negligible levels of immunity to malaria in the highlands in 1970. The transmission equation for the dynamics of cases is given by

$$C_t = \beta_{t-1} \beta_{\text{seas}} \left( \sum_{k=1:9} C_{t-k} \frac{S_{t-1}}{N} \right) \varepsilon_{t-1}, \quad (2)$$

where  $\varepsilon_t$  is an error term; and the transmission rate  $\beta$  has two components, a seasonal one,  $\beta_{\text{seas}}$ , and a long-term  $\beta_t$  encompassing variability at longer time scales than seasonal. We assume that infected individuals are able to transmit the disease for a period of nine months (Hay *et al.* 2000). Because  $\beta_t$  is not specified but determined from the model fit itself, the model is semi-parametric. Thus, we fit the model with the semi-parametric approach described in Koelle & Pascual (2004) and Koelle *et al.* (2005), using log-transformed

malaria cases. The approach essentially consists of two steps: (i) the fit of the transmission equation (2) using weighted linear regression after a log transformation and (ii) the smoothing of the residuals to obtain  $\beta_t$  from the residuals of the linear regression. The smoothing bandwidth  $h$ , which is the same for the weighted regression and the smoothing of the residuals, is chosen using cross-validation by removing one time point at a time in the data and refitting the model. The value of  $h$  determines the shape of the trend in the transmission rate  $\beta_t$ . Besides seasonality itself, there are two places in (2) where evidence for extrinsic forcing is reflected:  $\beta_t$  and the error terms  $\varepsilon_t$ . We refer to these error terms as the residuals of the model in the text. The variability in these two terms,  $\beta_t$  and  $\varepsilon_t$ , reflects sources of interannual variability in the dynamics of cases that are not captured by either the fluctuations of susceptibles or changes in seasonal transmissibility.

#### (f) *TSIR stochastic simulations*

To address whether the 4-year cycle observed in the data can be generated by the dynamics of the fitted TSIR model in the absence of forcing at shorter periods other than the seasonal one, we simulated the model repeatedly by sampling the residuals  $\varepsilon_t$  with replacement. Thus, in these simulations, the noise is uncorrelated and the circa-biennial cycles are no longer present. For each simulated time series, we obtained the wavelet spectrum of (the square-root transformed) cases and determined whether significant power occurred within a given window in the frequency–time domain. The specific quantities used to assess the occurrence of variability at periods close to four are described in the legend of [table 1](#). To further assess the possible interplay of intrinsic SIR dynamics with extrinsic forcing at the circa-biennial scale, we generated surrogate time series that have the same power spectrum than the residuals but differ in their temporal order ([Kaplan & Glass 1995](#)). We then simulated the fitted model with these surrogate residuals.

## 4. RESULTS

We begin by characterizing the dominant temporal scales present in the interannual variability of malaria with a WPS (§3). By contrast to the Fourier power spectrum, which gives an average picture of the dominant periodicities in the time series, the wavelet spectrum is localized in time ([Cazelles \*et al.\* 2007](#)). Thus, this analysis is especially well suited for patterns of variability that change with time, such as transient cycles, allowing the identification of not only dominant periods but also their timing. [Figure 1b](#) shows that malaria cases in AHP exhibit interannual variability around two main temporal scales with respective periods of *ca* 2 and 4 years. Both these cycles are most pronounced in the 1990s, while the former is already apparent earlier on in the 1980s. An 8-year period is also found in the late 1990s, although it lies outside the ‘cone of influence’ and is therefore not considered reliable. A similar pattern is found for cases in BBK, except that the longer cycle has a shorter period of *ca* 3 years ([figure S5b](#) in the electronic supplementary material), as previously shown in [Hay \*et al.\* \(2000\)](#). As for AHP, the power of these cycles is most intense in the 1990s. A longer cycle of 6 years is also apparent in the BBK time series. The variability in both time series at the *ca* 2-year frequency encompasses variability at period 2 or slightly greater than 2 years (e.g. in 1995–2000, [figure 1b](#)), as well as a shorter period between 1 and 2

[Table 1](#). Results of bootstrap tests for the presence of the 4-year cycle in the dynamics of the SIR model with noise. (CN indicates results for the simulated time series with correlated noise, while white noise (WN) indicates results for the simulated time series with uncorrelated noise. Different windows in the WPS were considered: the interval of time is specified in column 1; that for the period is fixed from 3.5 to 4.4 years. The average period (AvP, second column) of the peak in the WPS was computed for the malaria data. All time slices within a given window in the WPS were inspected for the presence of a maximum. A maximum is defined as a peak only if its  $p$  value was smaller than a threshold value (0.05). Times with no peak or no significant peak were given a value of zero and averaged as such. Then, the probability of finding a significant peak in the considered window was determined for the simulated time series (columns 3 and 4). This probability is high for the time series as a whole (within the cone of influence, for years 1974–1999). It increases from the 1970s to the 1990s and is higher for correlated noise than for WN. The average  $p$  value of all points within a given window was also computed (column 7) and the probability that this value for the simulated time series is smaller than a given threshold ( $p=0.05$ ) was also computed (columns 5 and 6). This indicated whether on average the probability of finding a significant power in that window is high. The probabilities again increase from the 1970s to the 1990s, and when noise is correlated. Since these  $p$  values are averages, they vary with the size of the window and are higher when a narrower period band is considered *ca* 4 years.)

time	AvP	CN	WN	CN	WN	Pav
1974–1999	3.97	0.95	0.73	0.00	0.01	0.29
1975–1979	4.33	0.46	0.28	0.05	0.11	0.18
1977–1981	4.30	0.45	0.26	0.04	0.14	0.21
1979–1983	0	0.48	0.25	0.06	0.16	0.30
1981–1985	0	0.51	0.25	0.09	0.18	0.05
1983–1987	0	0.54	0.28	0.14	0.21	0.07
1985–1989	0	0.58	0.27	0.19	0.22	0.79
1987–1991	0	0.62	0.30	0.24	0.25	0.59
1989–1993	3.68	0.64	0.32	0.26	0.27	0.24
1991–1995	3.76	0.66	0.34	0.29	0.31	0.04
1993–1997	3.86	0.65	0.39	0.34	0.33	0.04
1995–1999	3.82	0.65	0.41	0.37	0.36	0.01

years (e.g. 1981–1986, [figure 1b](#)). For simplicity, we refer hereafter to these components of variability as the circa-biennial cycle.

To further characterize dynamic patterns in disease incidence, we ask whether a significant change has occurred in the type of qualitative dynamics at some particular time in the last three decades. This type of change is known in the literature as a regime shift or break point (e.g. [Ricker 1963](#); [Solow & Beet 2005](#)). Because infectious diseases exhibit threshold behaviour, the question of the existence of break points is particularly relevant when conditions underlying transmission are themselves likely to be non-stationary. Here, we adapt a method previously applied to fisheries ([Solow & Beet 2005](#)) to take into consideration seasonality (§3). The analysis shows that malaria dynamics at AHP exhibit a shift at the beginning of the 1980s, with different dynamic regimes before and after the statistically chosen and significant June 1981 break point ( $p<0.001$ ; [figure S2](#) in the electronic supplementary material).

The timing of the break point corresponds with the apparent changes in both the size of epidemics ([figure 1a](#))

and the patterns of seasonal and interannual variability, with the appearance of the circa-biennial cycle in the wavelet spectrum (figure 1*b*). Although the spectrum suggests that both the circa-biennial and 4-year period may have been already present in the early 1970s despite the lower number of cases, this signal is too short and too close to the boundaries (technically, outside the cone of influence, figure 1*b*) to be reliable. What is clear is a change in the pattern of variability from the end of the 1970s to the beginning of the 1980s, accompanying the emergence of local epidemics.

Given these disease patterns, we can now examine a mechanistic basis for both the cycles and the long-term changes. To this end, we first address the endogenous versus exogenous origin of the cycles with a simple TSIR model that allows for long-term changes in transmission (§3). Although a TSIR model is too simple, a representation of the more complex processes underlying the dynamics of a vector-borne disease such as malaria (e.g. Aron & May 1982; Dietz 1988; Koella 1991), we used it here to follow up on the previous hypothesis that multiyear cycles could be explained by the dominant cycles of SIR dynamics (Hay *et al.* 2000). Evidence for this previous hypothesis relied on the period of an SIR model with parameters consistent with those in the malaria literature. Here, we tested this hypothesis further by fitting the TSIR model itself to the time-series data and asking whether it could indeed capture the cycles observed in the data. In the model, the transmission rate parameter ( $\beta_t$ ) is allowed to vary in time in a non-specified fashion, with this variation determined by fitting the model to the data. This semi-mechanistic (semi-parametric) approach allows us to identify evidence for extrinsic forcing by considering both the variability in  $\beta_t$  and in the residuals of the model, that is, in the error terms reflecting the variability in the transmission rate that is unaccounted for by the model itself (Koelle & Pascual 2004; Koelle *et al.* 2005). Two key questions are addressed with this approach. (i) Is there evidence for a temporal trend in the transmission rate? (ii) Are the observed dominant frequencies in malaria cases captured by SIR dynamics? Figure 2*a* shows that a long-term trend in transmission is identified by the model. Figure 2*b* shows the wavelet spectrum of the residuals of the model (equation (1); compare to figure 1*b*). While the circa-biennial cycles are still evident in the residuals, the longer, 4 years, cycle exhibits much lower power and reduced significance (figure 2*b*).

These results suggest that the 4-year cycle involves the endogenous dynamics of the disease while the shorter circa-biennial one does not. To examine the dynamics of the fitted TSIR model, we can consider three types of simulations: (i) deterministic runs of the model (that is, without noise), (ii) stochastic runs with uncorrelated, white noise (WN), and (iii) stochastic runs with correlated noise (CN). In the WN simulations, noise is included by sampling with replacement from the residuals of the model. In the CN simulations, surrogate time series of error terms are obtained that conserve the power spectrum (and therefore, the correlation structure) of the residuals but not their temporal order (Kaplan & Glass 1995). Variability in the simulated time series were then analysed with wavelets to determine the presence of a significant signal in specific windows of time and frequency (§3). These analyses tell us whether the model is indeed capable

of generating a period close to 4 years, and whether exogenous forcing, in the form of noise, interacts with the endogenous dynamics of the disease in generating this longer cycle. A purely deterministic model is unable to generate a period of four cycle (results not shown), while both types of stochastic simulations generated this cycle in a significant number of runs (table 1). Interestingly, these cycles are more frequent in the CN simulations. Furthermore, they increase in likelihood from the 1970s to the 1990s (table 1). These results indicate that noise plays a role in generating the cycles, and that both forcing at the circa-biennial scale and the higher transmission rate of the 1990s facilitate this pattern.

The presence of the circa-biennial pattern in the residuals of the TSIR model fit suggests that the shorter cycles are driven exogenously (figure 2*b*). If this is the case, then the temporal patterns observed for the cases should reflect the variability of an environmental factor relevant to the dynamics of malaria transmission. We investigated one such factor, rainfall, at three local stations in the region (figure 3*a* and electronic supplementary material, figure S1). Figure 3*b* shows the patterns of interannual variability for rainfall at the Chagaik station. The wavelet spectrum exhibits a similarity to that of malaria cases for the short periods between 1 and 3 years. In particular, dominant periods are identified *ca* 2 years. There is an intensification of the power at that scale during the 1990s, reflecting a higher fraction of the variance in those modes. A period six is also present in the rainfall spectrum. Similar patterns are seen for a time series of average monthly rainfall across the three stations (figure S3 in the electronic supplementary material) and for the other rainfall stations individually (results not shown). To quantitatively assess the correspondence of the wavelet spectra for malaria and rainfall, we obtained their cross-coherence spectrum (§3). Figure 4*a* shows that there is strong and significant cross-coherence between malaria cases at AHP and average rainfall across the three stations, for the circa-biennial scale. (A similar pattern is obtained for the residuals of the model and rainfall, figure S4 in the electronic supplementary material). This pattern is also present for the Chagaik rainfall time series alone (figure 4*b*). In addition, the coupling between these malaria cases and rainfall changes over time, with no significant cross-coherence in the 1970s (figure 4*b*). The phase difference for the circa-biennial cycle shows that rainfall anticipates malaria in the intervals of coupling with a time lag of two to three months for 1995–2000 and a shorter lag closer to zero in 1985–1990.

The analysis so far concerns only the interannual variability of the disease. A role of rainfall in multiyear cycles must find an explanation at the seasonal scale since this environmental driver is, together with temperature, one of the main potential influences on vector dynamics. The average seasonal cycle at AHP exhibits two main peaks per year: a small peak in February–March and a larger peak typically in June–July (figure 5*a*). Rainfall in the region also displays two distinct seasons, the October–November–December short rains and the March–April–May ‘long’ rains (figure 5*b*). Interestingly, accumulated rainfall from November to January is significantly correlated to total cases in the first peak, with a correlation coefficient of 0.72 ( $p=0.0002$ ; figure 5*c*). We considered only the later part of the short rains because longer lags

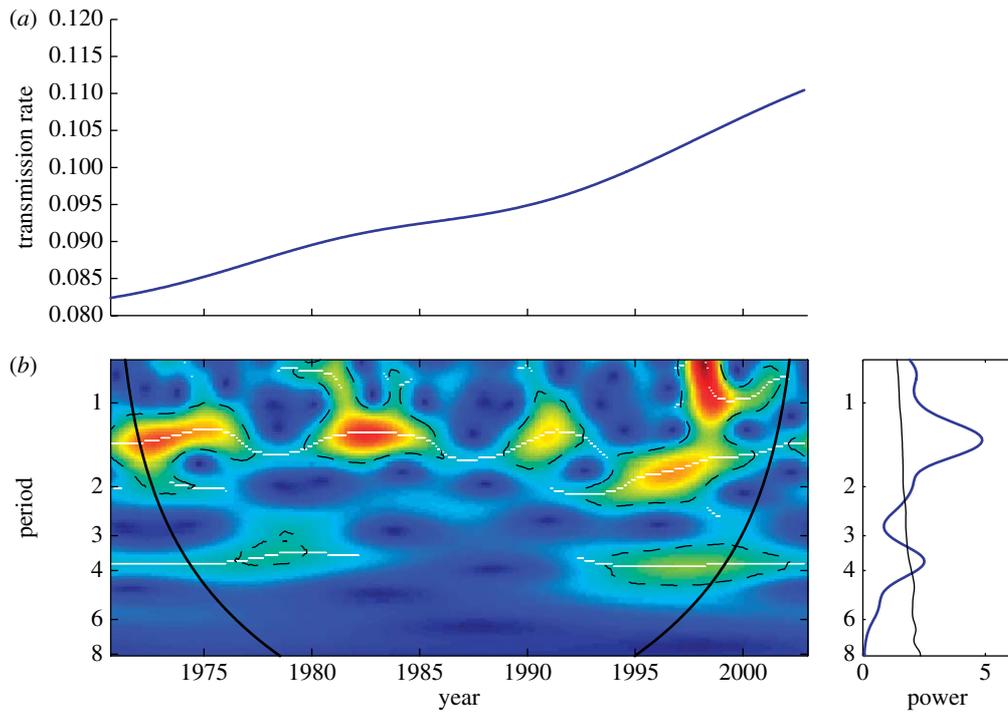


Figure 2. TSIR model results for the AHP malaria time series. (a) Long-term changes in the transmission rate,  $\log(\beta_t)$ . (b) Wavelet spectrum of the residuals of the TSIR model ( $\epsilon_t$  in equation (2)), showing significant variability at the circa-biennial scale, which is also present in the wavelet spectrum of the AHP case data. The 4-year mode appears only weakly, with reduced intensity and significance. (Because the model is fitted using log-transformed cases, the residuals were back-transformed by exponentiation to return to the original scale of the cases; and were further square-root transformed before the wavelet analysis for purposes of comparison with the original data in figure 1. Similar patterns are apparent in the non-transformed residuals. ( $N=50\,000$ ;  $T=10$  years; similar results are obtained for other values of these constants:  $N=10\,000$  and  $T$  between 5 and 20 years.)

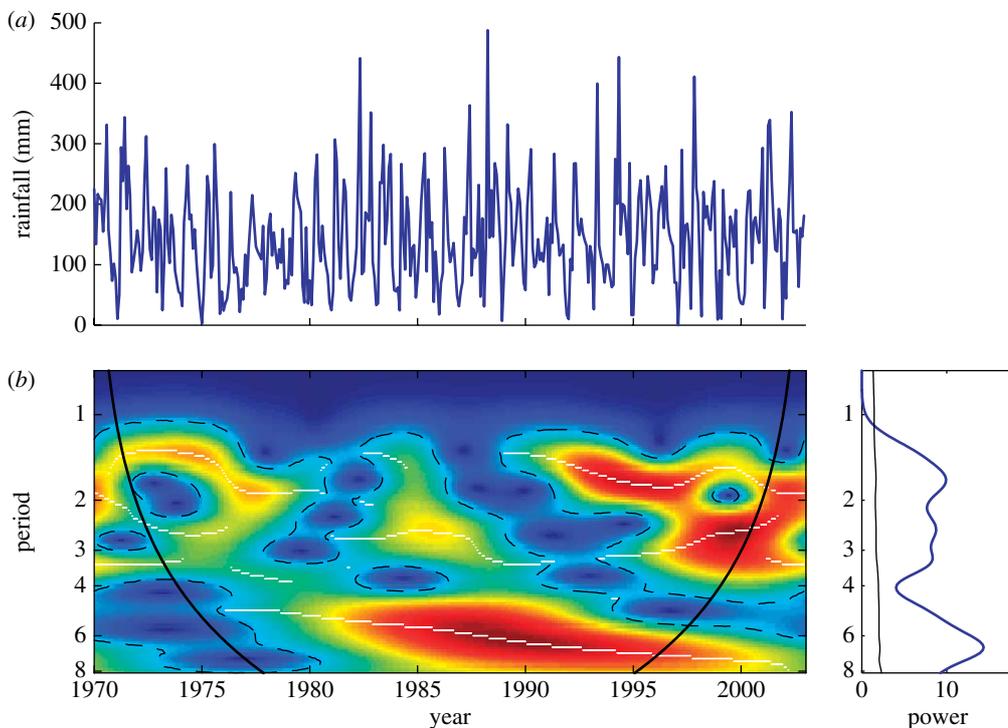


Figure 3. Patterns of rainfall interannual variability for one of the local stations (Chagaik). (a) The time series and (b) its wavelet spectrum. To remove seasonality and focus on the interannual variability, this environmental variable was filtered with a low-pass filter (Cazelles *et al.* 2007).

relative to cases would not be possible given the life history of the vector. In turn, the second and main peak in cases is significantly associated with the first peak in cases, with a correlation coefficient of 0.59 ( $p=0.0032$ ; figure 5d).

Similar patterns are found for the BBK time series (figure S6 in the electronic supplementary material).

Finally, the analyses of interannual variability were repeated for BBK; this supported the role of extrinsic

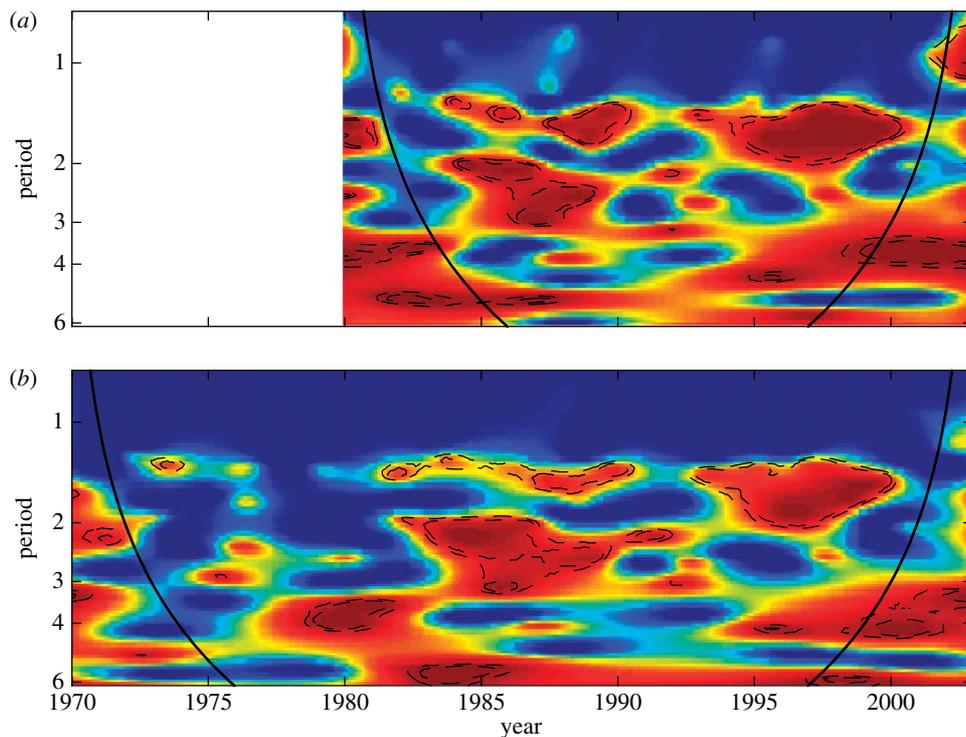


Figure 4. Wavelet cross-coherence between malaria at AHP and rainfall at interannual time scales. (a) The cross-coherence for the average of monthly rainfall at the three local stations. (Results start in 1980 because one of the time series begins at this later time and so does the average.) Cross-coherence varies between 0 and 1, shown here in a colour scale from blue to red. The lines correspond to the 5 and 10% significance levels. Only regions within these lines exhibit significant cross-coherence at those levels. (b) The same analysis but now for a local station stating in the 1970s (Chagaik). There is no significant cross-coherence in the 1970s at interannual time scales. After 1980s, rainfall and malaria exhibit coherent cycles at the circa-biennial scales. (The rainfall time series has been filtered with a low-pass filter to remove seasonality and periods below 1 year; similar patterns are obtained without the filtering with the additional expected coherence at the 1-year scale, results not shown).

forcing by rainfall. The residuals of the selected TSIR model show variability at the circa-biennial cycle and at the period three (figure S5 in the electronic supplementary material). The localization of both these cycles in time and frequency is similar to that of rainfall (compare with figure 3b and electronic supplementary material, figure S3). The cycle six loses its significance in the residuals indicating a possible intrinsic origin. Because this period is clearly present in the rainfall spectrum (figure 3b), and shows significant cross-coherence between cases and rainfall (not shown), it is difficult to assign a conclusive origin to this longer cycle. However, the model does not provide evidence for a long-term change in the transmission rate since  $\beta_t$  does not exhibit a clear trend in the selected model. We also found no evidence in these data for a significant qualitative shift in the dynamics. We discuss below these differences with AHP.

## 5. DISCUSSION

Our results bring together two different perspectives that have long been seen as alternative explanations for the interannual variability of malaria in African highlands. Both the endogenous and exogenous origins of multiyear cycles are supported by our analyses of the AHP data but for different temporal scales, respectively. The circa-biennial cycle appears to be driven exogenously, while the longer cycle of 4 years is consistent with endogenous SIR dynamics, involving the processes of transmission and immunity. One climatic driver, rainfall, was identified whose interannual variability is coherent with that of

malaria cases (and the model residuals) in both the time and the frequency domains for the circa-biennial scale. Interestingly, there is also evidence for the interplay of exogenous and endogenous factors: simulations of the fitted TSIR model suggest that forcing at the circa-biennial scale can resonate with the dynamics of the disease to facilitate the generation of 4-year cycles. The stochastic dynamics of the disease should be explored further with more realistic mathematical models specifically formulated for malaria transmission.

The role of rainfall is further supported by the seasonal findings for AHP. The short rains appear associated with the small peak of malaria at the beginning of the following year. This is important because the number of cases in this first outbreak show themselves a significant correlation with the number of cases later in the year. Thus, anomalous rainfall in the short rains would display a ripple effect and affect the total number of cases in the following year. Interestingly, although the highest precipitation occurs during the long rains, the short rains in eastern Africa have been shown to exhibit a larger degree of interannual variability relative to climatology (Hastenrath *et al.* 1993; Clark *et al.* 2003). One striking example is given by the short rains in 1997 which are reported to have been 5–10 times the normal in many East African locations (Clark *et al.* 2003; World Meteorological Organization). Climate studies have also described the correlation of the short rains to sea surface temperatures (SSTs) in the Indian Ocean, with an association that varies over interdecadal scales for East African coastal precipitation (Clark *et al.* 2003). While the geographical focus of

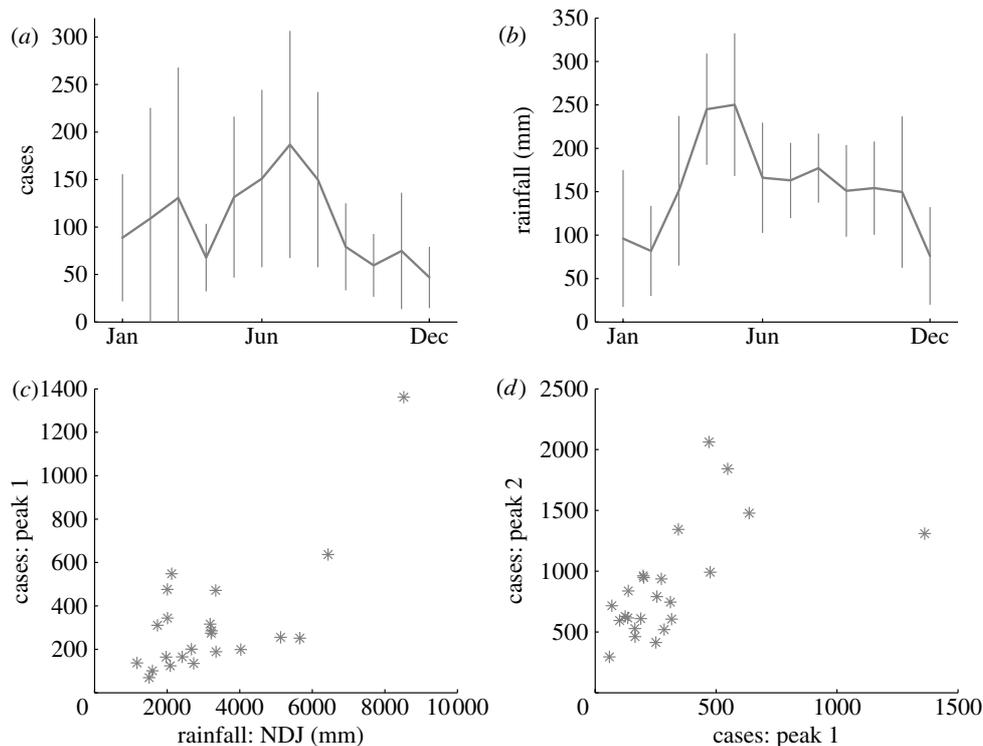


Figure 5. Seasonal patterns of malaria and rainfall. (a) Monthly average cases (continuous line) for AHP. The vertical bars correspond to 1 s.d. from the mean. (b) Monthly average precipitation from the mean of the three local stations (continuous line). (c) The total cases in the first peak for February–March–April are shown as a function of total precipitation for November–December–January. (d) The number of cases from May to December is plotted as a function of the number of cases in February–March–April. Correlations values are reported in the text. The significant association is not the result of the long-term trend in the mean, since a similar result is obtained when cases are normalized by one particular month (December), which effectively detrends the data.

those studies differs from ours, results are reported to hold in a larger region of eastern Africa. The described role of rainfall described here should be investigated further in the context of larger scale climate variability, in particular SST in the Indian Ocean.

This first peak between January and March in the ‘dry season’ is observed in some time series of malaria in the region (e.g. Hay *et al.* 2002b), and may not be associated with *Anopheles gambiae*, the vector usually mentioned in the context of Kericho (Malakooti *et al.* 1998; Shanks *et al.* 2005a,b). Africa’s main malaria vectors species, *Anopheles funestus*, and several subspecies of *A. gambiae* employ different habitats and consequently have differences in seasonal prevalence. Where *A. funestus* uses more (semi-) permanent, often swampy, water bodies and benefits from surrounding vegetation such as grassy ponds, *A. gambiae* is strongly associated with shallow temporary breeding sites (Zahar 1985). Trends and changes between years in the relative dominance of species have been documented: in Kenya, for example, the introduction of artificial fish ponds has been associated with *A. funestus* becoming more dominant over a number of years (Lockhart *et al.* 1969), and in Kericho itself, dramatic changes between *A. gambiae* and *A. funestus* were observed in the 1940s (Heisch & Harper 1949), with *A. funestus* being incriminated as the cause of an epidemic which started in March 1948. The early year maximum (February) of this vector also transpires from other locations in Kenya (Chandler *et al.* 1975). A better understanding of the first peak, and therefore of the seasonal cycle, will require an examination of this conjecture, a challenging problem giving the extremely low densities

of adult mosquito vectors in these regions (Koenraadt *et al.* 2006).

There are similarities and differences between the two time series: the results for the association between rainfall and cases at the seasonal and circa-biennial scales are consistent for both estates. The endogenous 4-year cycle present at AHP contrasts with the 3-year rainfall driven cycle at BBK. There is also evidence for a break point at AHP when epidemics start in the 1980s; this is complemented by a long-term upward trend in the transmission rate. However, only hospitalized (inpatient) cases are included in BBK, and thus considerably fewer cases are reported for this estate (approx. half of those at AHP). Furthermore, more effective antimalarials were introduced at BBK after 1999; this seemed to keep the number of cases consistently lower than that at AHP in the following years (Shanks *et al.* 2005b). The lower incidence at BBK could mask a regime shift in the 1980s and when coupled to the overall decrease of cases in recent years may explain the lack of detection of long-term change in transmission. Because the fitting of the model requires a log transformation (Koelle & Pascual 2004) and the number of malaria cases at BBK is capped by admission capacity of the hospital, the apparently smaller epidemics at BBK may not be sufficient to reveal a change in  $\beta_t$ . It is possible that our findings could be re-evaluated when revised and extended data for both tea estates become available. Given that the patterns we have described are consistent in their relationship with rainfall for both estates, it is particularly urgent to gather more data from the estate with both inpatients and outpatients and to

compare the observed incidence in the tea plantations with those in the highland region of Kenya.

The circa-biennial scale in our analyses comprises variability between 1 and 2 years but also at a period slightly longer than 2 years. This variation in the exact localization of the cycle can result from a biennial cycle combined with a bimodal seasonal pattern with two peaks per year. Variation in the seasonal timing of the anomaly, from one peak to the next, can introduce cycles both shorter and longer than 2 years. We note that the latter period is consistent with the so-called quasi-biennial oscillation described for rainfall in the literature (e.g. Kabanda & Jury 1999).

Epidemics and the multiyear cycles that modulate their size are evident in the 1980s and become particularly pronounced in the 1990s at AHP, a pattern consistent with our finding of a regime shift at the beginning of the 1980s. This shift can be interpreted as a significant qualitative change in the dynamics, corresponding to the emergence of larger epidemics. Two possible contributing mechanisms are suggested by our analyses. The first one is a change in the coupling between malaria and rainfall in the last two decades. The second is a long-term trend in the transmission rate of malaria identified by the TSIR model. Given the simplicity of this model, we emphasize that the change in this parameter could result from different causes capable of effectively increasing local transmission. The model itself cannot specify any given mechanism *per se*. One possibility is the known increase in drug resistance (Malakooti *et al.* 1998; Shanks *et al.* 2005b). Another possibility is an increase in the transmission rate due to environmental change, including higher temperatures (Pascual *et al.* 2006) and change in mosquito life history mediated by land-use and agricultural practices (Wilson *et al.* 1990; Lyimo & Koella 1992; Lindblade *et al.* 2000; Koenraadt *et al.* 2004; Mutero *et al.* 2004; Kebede *et al.* 2005). Another possible driver of the observed trend is increased human mobility between highlands and endemic regions (Prothero 1965; Shanks *et al.* 2005a). In particular, this may be the case for Kericho since the production of the tea states relies on migrant workforce.

It is probable that a combination of these factors, with important synergies (e.g. Lindblade *et al.* 2000), are at play behind the described trend. Regardless of origin, the trend in transmission appears to interact with the interannual dynamics, increasing the likelihood of the longer cycles towards the end of time series. Future work should consider how to fit more realistic models for vector-borne diseases to the data, not only to revisit our findings but also to specifically address different hypotheses about the underlying trend in both cases and transmission. The development of recent methods to fit continuous-time mathematical models of infectious diseases with dynamical and measurement noise provides a promising avenue (Ionides *et al.* 2006). Important aspects that are not accounted for in a TSIR formulation relate to the dynamics of the vector, the delay in the development of the pathogen and the complexities of the duration of the infectious and immune states in malaria. Moreover, simplifying assumptions in our model included the treatment of population size as constant with a fixed turnover rate. Consideration of different turnover rates led to similar results. Although changing patterns of immigration have been described for these highlands that

are clearly important to transmission (Lindblade *et al.* 2000), time-series data on population sizes and immigration patterns in the tea plantations are not available for the whole time span of interest. Further modelling studies are warranted; our results emphasize the importance of addressing outbreak patterns and long-term change together, from a dynamical perspective, if we are to understand the shifting patterns of malaria epidemics in transition regions. Climate change scenarios should consider not just temperature but patterns of change in rainfall variability.

This project was initiated as a part of the Working Group on Global Change and Infectious Disease, supported by the National Center for Ecological Analysis and Synthesis (NCEAS), a centre funded by NSF and UC Santa Barbara. We especially thank Simon Hay for providing the malaria data and Dennis Shanks for generously sharing information on the data and the tea plantations; also, Bond Kenya, Ltd. (now Unilever Tea Kenya, Ltd.) and African Highlands Produce (now Findlay Farms) for the data from their health facilities. We also thank the Kenyan Meteorological Department for providing the rainfall data. M.P. is grateful to the Université Pierre et Marie Curie for a short-term position in Paris. Support was also provided by a Centennial Fellowship by the James S. McDonnell Foundation in Global and Complex Systems, and by joint funding from the National Science Foundation–National Institutes of Health (Ecology of Infectious Diseases Grant EF 0430 120) and the National Oceanic and Atmospheric Administration (Oceans and Health Grant NA 040 AR 460019), to M. P.

## REFERENCES

- Alonso, D., McKane, A. J. & Pascual, M. 2007 Stochastic amplification in epidemics. *J. R. Soc. Interface*, **4**, 575–582. (doi:10.1098/rsif.2006.0192)
- Aron, J. L. & May, R. M. 1982 The population dynamics of malaria. In *The population dynamics of infectious diseases* (ed. R. M. Anderson), pp. 139–179. London, UK: Chapman and Hall.
- Brockwell, P. J. & Davis, R. A. 2002 Introduction to time series and forecasting, p. 434. 2nd edn. New York, NY: Springer.
- Cazelles, B., Chavez, M., Constantin de Magny, G., Geguan, J. F. & Hales, S. 2007 Time-dependent spectral analysis of epidemiological time series with wavelets. *J. R. Soc. Interface* **4**, 625–636. (doi:10.1098/rsif.2007.0212)
- Chandler, J. A., Highton, R. B. & Hill, M. N. 1975 Mosquitoes of the Kano plain, Kenya. Part I. Results of indoor collections in irrigated and non-irrigated areas, using human bait and light traps. *J. Med. Entomol.* **12**, 202–207.
- Clark, C. C., Webster, P. J. & Cole, J. E. 2003 Interdecadal variability of the relationship between the Indian Ocean Zonal Mode and East African Coastal rainfall anomalies. *J. Clim.* **16**, 548–554. (doi:10.1175/1520-0442(2003)016<0548:IVOTRB>2.0.CO;2)
- Dietz, K. 1988 Mathematical models for transmission and control of malaria. In *Malaria* (eds W. H. wendsdorfer & I. McGregor), pp. 1091–1133. Edinburg, UK: Churchill Livingstone.
- Efron, B. & Tibshirani, R. J. 1994 *An introduction to the bootstrap*. Boca Raton, FL: CRC Press.
- Finkenstadt, B. F. & Grenfell, B. T. 2000 Time series modelling of childhood diseases: a dynamical systems approach. *Appl. Stat.* **49**, 187–205.

- Hastenrath, S., Nicklis, A. & Greischar, L. 1993 Atmospheric-hydrospheric mechanisms of climate anomalies in the western equatorial Indian Ocean. *J. Geophys. Res.* **98**, 20 219–20 235.
- Hay, S. H., Myers, M. F., Burke, D. S., Vaughn, D. W., Endy, T., Ananda, N., Shanks, G. D., Snow, R. W. & Rogers, D. J. 2000 Etiology of interepidemic periods of mosquito-borne disease. *Proc. Natl Acad. Sci. USA* **97**, 9335–9339. (doi:10.1073/pnas.97.16.9335)
- Hay, S. I., Cox, J., Rogers, D. J., Randolph, S. E., Stern, D. L., Shanks, G. D., Myers, M. F. & Snow, R. W. 2002a Climate change and the resurgence of malaria in the east African highlands. *Nature* **415**, 905–909. (doi:10.1038/415905a)
- Hay, S. I., Noor, A. M., Simba, M., Busolo, M., Guyatt, H. L., Ochola, S. A. & Snow, R. W. 2002b Clinical epidemiology of malaria in the highlands of western Kenya. *Emerg. Infect. Dis.* **8**, 543–548.
- Hay, S. I., Shanks, G. D., Stern, D. I., Snow, R. W., Randolph, S. E. & Rogers, D. J. 2005 Climate variability and malaria epidemics in the highlands of East Africa. *Trends Parasitol.* **21**, 54–56. (doi:10.1016/j.pt.2004.11.007)
- Heisch, R. B. & Harper, J. O. 1949 An epidemic of malaria in Kenya highlands transmitted by *A. funestus*. *J. Trop. Med. Hyg.* **52**, 187–190.
- Ionides, E., Breto, C. & King, A. A. 2006 Inference for nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **103**, 18 438–18 443. (doi:10.1073/pnas.0603181103)
- Kabanda, T. A. & Jury, M. R. 1999 Interannual variability of short rains over northern Tanzania. *Clim. Res.* **13**, 231–241. (doi:10.3354/cr013231)
- Kaplan, D. & Glass, L. 1995 Understanding nonlinear dynamics. *Texts in applied mathematics*, pp. 342–344. New York, NY: Springer.
- Kebede, A., McCann, J. C., Kiszewski, A. E. & Ye-Ebiyo, Y. 2005 New evidence of the effects of agro-ecologic change on malaria transmission. *Am. J. Trop. Med. Hyg.* **73**, 676–680.
- Kilian, A. H. D., Langi, P., Talisuna, A. & Kabagambe, G. 1999 Rainfall pattern, El Niño and malaria in Uganda. *Trans. R. Soc. Trop. Med. Hyg.* **93**, 22–23. (doi:10.1016/S0035-9203(99)90165-7)
- Koella, J. C. 1991 On the use of mathematical models of malaria transmission. *Acta Trop.* **49**, 1–25. (doi:10.1016/0001-706X(91)90026-G)
- Koelle, K. & Pascual, M. 2004 Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera. *Am. Nat.* **163**, 901–913. (doi:10.1086/420798)
- Koelle, K., Rodo, X., Pascual, M., Yunus, Md. & Mostafa, G. 2005 Refractory periods to climate forcing in cholera dynamics. *Nature* **436**, 696–700. (doi:10.1038/nature03820)
- Koenraadt, C. J. M., Majambere, S., Hemerik, L. & Takken, W. 2004 The effects of food and space on the occurrence of cannibalism and predation among larvae of *Anopheles gambiae* s.l. *Entomol. Exp. Appl.* **112**, 125–134. (doi:10.1111/j.0013-8703.2004.00186.x)
- Koenraadt, C. J. M., Paaijmans, K. P., Schneider, P., Githeko, A. K. & Takken, W. 2006 Low larval vector survival explains unstable malaria in the western Kenya highlands. *Trop. Med. Int. Health* **11**, 1195–1205. (doi:10.1111/j.1365-3156.2006.01675.x)
- Lindblade, K. A., Walker, E. D., Onapa, A. W., Katungu, J. & Wilson, M. L. 1999 Highland malaria in Uganda: prospective analysis of an epidemic associated with El Niño. *Trans. R. Soc. Trop. Med. Hyg.* **93**, 480–487. (doi:10.1016/S0035-9203(99)90344-9)
- Lindblade, K. A., Walker, E. D., Onapa, A. W., Katungu, J. & Wilson, M. L. 2000 Land use change alters malaria transmission parameters by modifying temperature in a highland area of Uganda. *Trop. Med. Int. Health* **5**, 263–274. (doi:10.1046/j.1365-3156.2000.00551.x)
- Lockhart, J. D. F., Highton, R. B. & McMahan, J. P. 1969 Public health problems arising out of man-made fish ponds in the western province of Kenya. *East Afr. Med. J.* **46**, 471–480.
- Loevinsohn, M. E. 1994 Climatic warming and increased malaria incidence in Rwanda. *Lancet* **343**, 714–718. (doi:10.1016/S0140-6736(94)91586-5)
- Lyimo, E. O. & Koella, J. C. 1992 Relationship between body size of adult *Anopheles gambiae* s.l. and infection with the malaria parasite *Plasmodium falciparum*. *Parasitology* **104**, 233–237.
- Malakooti, M. A., Biomndo, K. & Shanks, G. D. 1998 Re-emergence of epidemic malaria in the highlands of western Kenya. *Emerg. Infect. Dis.* **4**, 671–676.
- Marimbu, J., Ndayiragije, A., Le Bras, M. & Chaperon, J. 1993 Environment and malaria in Burundi. Apropos of a malaria epidemic in a non-endemic mountainous region. *Bull. Soc. Path. Exotique* **86**, 399–401.
- Mutero, C. M., Ng'ang'a, P. N., Wekoyela, P., Githure, J. & Konradsen, F. 2004 Ammonium sulphate fertiliser increases larval populations of *Anopheles arabiensis* and culicine mosquitoes in rice fields. *Acta Trop.* **89**, 187–192. (doi:10.1016/j.actatropica.2003.08.006)
- Pascual, M., Ahumada, J., Chaves, L. F., Rodo, X. & Bouma, M. 2006 Malaria resurgence in East African highlands: temperature trends revisited. *Proc. Natl Acad. Sci. USA* **103**, 5829–5834. (doi:10.1073/pnas.0508929103)
- Prothero, R. M. 1965 *Migrants and malaria*. London, UK: Longmans.
- Ricker, W. E. 1963 Big effects from small causes: two examples from fish population dynamics. *J. Fish. Board Can.* **20**, 257–264.
- Shanks, G. D., Biomndo, K., Hay, S. I. & Snow, R. W. 2000 Changing patterns of clinical malaria since 1965 among a tea state population located in the Kenyan highlands. *Trans. R. Soc. Trop. Med. Hyg.* **94**, 253–255. (doi:10.1016/S0035-9203(00)90310-9)
- Shanks, G. D., Biomndo, K., Guyatt, H. L. & Snow, R. W. 2005a Travel as a risk factor for uncomplicated *Plasmodium falciparum* malaria in the highlands of western Kenya. *Trans. R. Soc. Trop. Med. Hyg.* **99**, 71–74. (doi:10.1016/j.trstmh.2004.04.001)
- Shanks, G. D., Hay, S. I., Omumbo, J. A. & Snow, R. W. 2005b Malaria in Kenya's western highlands. *Emerg. Infect. Dis.* **11**, 1425–1432.
- Solow, A. R. & Beet, A. R. 2005 A test for a regime shift. *Fish. Oceanogr.* **14**, 236–240. (doi:10.1111/j.1365-2419.2005.00329.x)
- Torrence, C. & Compo, G. P. 1998 A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**, 61–78. (doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2)
- Wilson, M. L., Agudelo-Silva, F. & Spielman, A. 1990 Increased abundance, size, and longevity of food-deprived mosquito populations exposed to a fungal larvicide. *Am. J. Trop. Med. Hyg.* **43**, 551–556.
- Zahar, A. R. 1985 *Vector bionomics in the epidemiology and control of malaria part I. The WHO African Region*. WHO publication VBC/85. Geneva, Switzerland: WHO.
- Zhou, G., Minakawa, N., Githeko, A. K. & Yan, G. 2004 Association between climate variability and malaria epidemics in the East African highlands. *Proc. Natl Acad. Sci. USA* **101**, 2375–2380. (10.1073/pnas.0308714100)