A stochastic model for ecological systems with strong nonlinear response to environmental drivers: application to two water-borne diseases

Claudia Torres Codeço[1],*, Subhash Lele[2], Mercedes Pascual[3], Menno Bouma[4] and Albert I. Ko[5]

[1]*Oswaldo Cruz Foundation, Avenida Brasil, 4365, Residência Oficial, Rio de Janeiro, RJ 21045-900, Brazil*
[2]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*
[3]*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48104, USA*
[4]*Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK*
[5]*Gonçalo Muniz Research Center, Rua Waldemar Falcão 121, Brotas Salvador, BA, 40295-001, Brazil*

Ecological systems with threshold behaviour show drastic shifts in population abundance or species diversity in response to small variation in critical parameters. Examples of threshold behaviour arise in resource competition theory, epidemiological theory and environmentally driven population dynamics, to name a few. Although expected from theory, thresholds may be difficult to detect in real datasets due to stochasticity, finite population size and confounding effects that soften the observed shifts and introduce variability in the data. Here, we propose a modelling framework for threshold responses to environmental drivers that allows for a flexible treatment of the transition between regimes, including variation in the sharpness of the transition and the variance of the response. The model assumes two underlying stochastic processes whose mixture determines the system's response. For environmentally driven systems, the mixture is a function of an environmental covariate and the response may exhibit strong nonlinearity. When applied to two datasets for water-borne diseases, the model was able to capture the effect of rainfall on the mean number of cases as well as the variance. A quantitative description of this kind of threshold behaviour is of more general application to predict the response of ecosystems and human health to climate change.

Keywords: nonlinear; cholera; leptospirosis; seasonality; disease dynamics; extrinsic forcing

1. INTRODUCTION

Ecology is pervaded by models with threshold behaviour, that is, models showing drastic shifts in population abundance or species diversity in response to relatively small changes in the parameters that control their dynamics. Resource competition theory, for example, predicts that microbial community should change drastically when resource ratio crosses a critical point where growth limitation shifts from one resource to another (Grover 1997). Epidemic theory predicts that parasites should persist only if host population density increases above a critical value (Anderson & May 1992). Population dynamic models also describe drastic population responses to environmental factors. Ellis & Post (2004), for example, show that the effect of temperature and density on wolf population is dramatically different below and above a critical density value. Below this threshold, population density is positively and linearly associated with temperature due to its effect on prey availability. Above the threshold, the effect of temperature becomes weak when compared with density effects. Grenfell *et al.* (1998) and Stenseth *et al.* (2004) describe a similar threshold behaviour for soya sheep in islands. Jacobson *et al.* (2004) show that the rate of population increase of ungulates changes drastically as snow depth crosses a critical point. The existence of such thresholds has significant consequences for the impact of climate change on ecological systems.
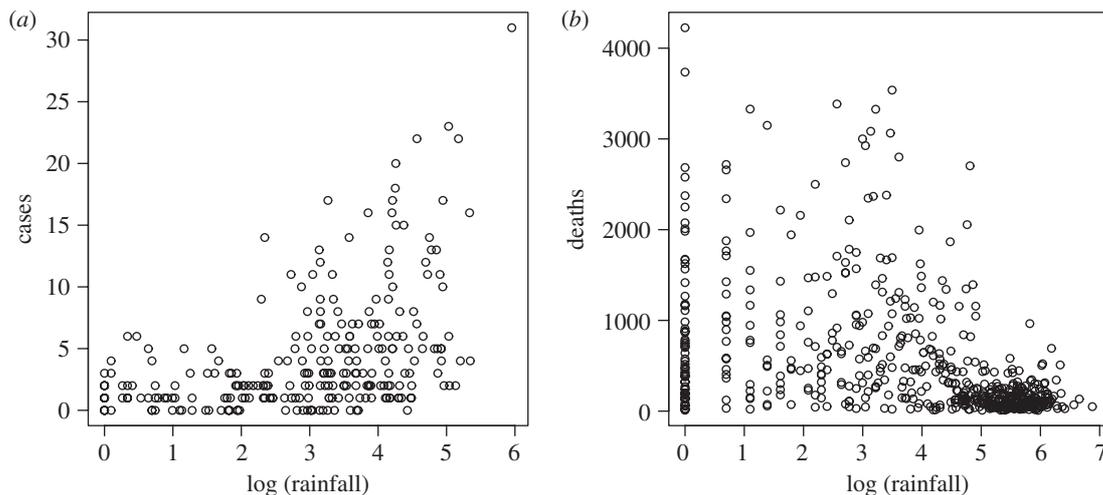
Figure 1. Two examples of systems with threshold response to the environmental driver. (*a*) Number of weekly severe cases of leptospirosis in a hospital in Salvador, Brazil, versus the amount of rain (mm) in the previous week (log-transformed), 1996–2001 (Flannery *et al.* 2001). (*b*) Number of monthly cholera deaths in Parganas, Bangladesh, versus the amount of rainfall in the same month (log-transformed), 1893–1940 (Pascual *et al.* 2002).

Classical approaches for modelling systems with threshold behaviour assume the existence of two (or more) non-interacting regimes: 'low-density' and 'high-density' (Ellis & Post 2004) or 'regular' and 'explosive' (Engel *et al.* 2001). Depending on the current regime, the system's response to covariates is governed by a different probability distribution or a different dynamic equation. The value of the threshold is often fixed and deterministic, and identified from data by visual inspection of scatter plots (Jacobson *et al.* 2004) or by investigating the threshold value that optimizes the fit of piecewise functions (Grenfell *et al.* 1998; Engel *et al.* 2001). This approach, however, often hides several difficulties that arise when either looking for theoretically predicted threshold phenomena in real datasets or representing observed thresholds using mathematical models. First, in real data, the existence of thresholds may be blurred by stochasticities and inadequate data resolution (Lloyd-Smith *et al.* 2005) which may soften the observed transition between regimes, making the identification of critical points a difficult task. Moreover, the variability in the system's response and not just its mean can vary drastically across a threshold. Figure 1 shows two examples of this pattern that will be considered in this paper. Similar patterns arise in the abundance of blue-green algae as a function of N : P ratios in lakes (Smith, 1983).

Here we propose a statistical modelling framework that allows for a flexible treatment of the transition between regimes, including different degrees of sharpness in the transition as well as drastic changes in system variability. The proposed model assumes two underlying stochastic processes, each of which dominates at one end of the environmental spectrum. Between these two extremes, the observations are a mixture of the two processes and the mixture proportion is a function of the driver. Thus, the transition between states can be more or less gradual and the actual shape of the transition is determined from the data as a function of the covariates. With this approach, we model situations where shifts between regimes do not follow a simple on and off response.

We illustrate the model with an application to time-series data for two water-borne diseases, leptospirosis and cholera, and their seasonal response to rainfall. The proposed threshold model captures not only the change in mean disease levels, but also their variances, as a function of rainfall, providing a functional form for the stochastic transition between regimes.

2. TWO EXAMPLES OF ENVIRONMENTAL THRESHOLDS

Figure 1 shows scatter plots of disease versus rainfall for leptospirosis in Salvador, Brazil (Flannery *et al.* 2001) and cholera in Parganas, Bangladesh (Pascual *et al.* 2002). Leptospirosis is a water-borne acute bacterial disease that affects both humans and animals. Human infection occurs by contact with rodent urine-contaminated environment, surface water, soil and objects (WHO 2005). In many tropical urban centres, leptospirosis is an endemic disease, presenting a baseline incidence throughout the year and sporadic outbreaks after heavy rains, possibly due to the increased contact of people with contaminated flood water (Sarkar *et al.* 2002; Karande *et al.* 2003).

Historically, large cholera outbreaks in Bangladesh have been associated with deficient monsoon rainfalls (see Pascual *et al.* 2002, for a review of environmental drivers). This pattern has been classically attributed to the short supply of potable water during droughts. Similarly, the typical seasonal pattern in this endemic region involves two peaks per year with a drastic reduction in cases during the monsoons. A number of other hypotheses have been proposed linking rainfall to seasonal cholera cases through environmental factors that affect the concentration of the pathogen in aquatic environments (brackish and freshwater). Larger inocula of *Vibrio cholerae* in the environment would result from warming up of the water and associated

plankton blooms outside the rainy season (Lobitz *et al.* 2000). Increased rainfall would dilute the concentration of the pathogen through an increase in volume of the environmental reservoir, lower salinity (below 15 0/00; Miller *et al.* 1982), lower pH (below 8.5; Huq *et al.* 1984) and reduced numbers of lysogenic vibriophages in the water (Faruque *et al.* 2005). Since a heavy bacterial load must be consumed to initiate a life-threatening case of cholera, the number of total deaths by cholera goes down in response to this dilution effect. A negative effect of rainfall has also been described for other regions of the world where periodic floods occur, as in the region of Madras, India (Pascual *et al.* 2002) and the Brazilian Amazon (Codeço 2001). However, the relationship between rainfall and cholera is complex with historical patterns of positive seasonal association in dry regions such as the Punjab and the northern districts of Madras where intermittent outbreaks occur (Pascual *et al.* 2002), as well as in more endemic, less dry regions at longer lags in southern Madras (Ruiz-Moreno *et al.* 2007). Evidence for positive interannual associations have also been recently described between floods and cholera transmission following El Niño events in recent decades in a rural area south of Dhaka (Koelle *et al.* 2005).

Despite the anecdotal evidence and the many known mechanisms by which rainfall can drive the seasonal outbreaks of water-borne diseases, a quantitative analysis of the nonlinear response of incidence to this environmental driver is still missing. Here, we illustrate a positive response of leptospirosis, with increased incidence after heavy rainfall in Salvador, Brazil, and a negative response of cholera, with a decrease in cases during the monsoon season in Bangladesh.

Both diseases show a different response at low and high values of the environmental driver and a sharp transition between these behaviours. At one extreme, the number of cases are consistently low (relatively to the average), while at the other extreme, a mixture of high and low cases is found. With the threshold model, the objective is to describe the effect of the environmental driver not only on the expected number of cases but also on the system's variability.

2.1. Threshold models: description and inference

To identify and describe the threshold behaviour suggested by the above disease patterns, we introduce a threshold model that incorporates two underlying processes and considers that the observed response is a mixture of these two processes. At extreme values of the environmental driver, one of the processes predominates. Between these two extremes, the observations are a mixture of the two processes and the mixture proportion is a function of the values of the environmental covariate.

McLachlan & Peel (2000) provide a good introduction to the general theory of probabilistic mixtures of distributions and likelihood-based inference. We use here the theory of estimating functions (Godambe 1991), instead of applying the full probabilistic modelling approach and likelihood-based

Table 1. Parameter estimates for the threshold model.

parameter	leptospirosis		cholera	
	point estimate	95% CI	point estimate	95% CI
μ_1	1.77	[1.44,2.08]	2214	[2132,2348]
μ_2	13.8	[12,17]	157	[156,159]
a	-6.25	[$-10,-5$]	1.2	[0.62,1.86]
b	1.29	[0.8,2.1]	-1.06	[$-1.39,-0.81$]

inference, and consider only the mean and the variance functions of the process to estimate the parameters of the model.

Let μ_1, μ_2 denote the mean responses and σ_1^2, σ_2^2 denote the variances of the two processes, respectively. Let $\pi(z)$ denote the mixture proportion that is a function of the observed covariates Z. Then, the mean and the variance of the observed mixed process are given by

$$E(Y|z) = \pi(z)\mu_1 + (1-\pi(z))\mu_2, \qquad (2.1)$$

$$\mathrm{Var}(Y|z) = \pi(z)\sigma_1^2 + (1-\pi(z))\sigma_2^2 \\ + \pi(z)(1-\pi(z))(\mu_1 - \mu_2)^2. \qquad (2.2)$$

In the water-borne diseases we consider here, we model the individual processes such that $\sigma_1^2 = \mu_1$ and $\sigma_2^2 = \mu_2$. This is similar to modelling the component processes as having a Poisson distribution. One may also consider other mean and variance functions. Furthermore, note that if one takes $\mu_1 = \mu_2$, this approach allows one to model observations where the variance of the observed process is a mixture of the variance of the component processes depending on the covariates, but the mean is unaffected by these values. We model the proportion in the mixture using the logistic function, namely,

$$\pi(z) = \exp(a + bz)/(1 + \exp(a + bz)). \qquad (2.3)$$

This flexible function takes values in the (0,1) range and allows for sharp or slow changes in the mixture proportion. Of course, any other function that takes values between 0 and 1 may also be used. In our examples, the change in the mixture proportion as a function of the covariate is fairly sharp and hence we call these models 'threshold models'.

Let us denote the set of parameters that we want to estimate as θ. To estimate θ, we minimize the objective function as follows (Lele & Taper 2002):

$$\sum_{i=1}^{n} \frac{(y_i - E(Y|z;\theta))^2}{\mathrm{Var}(Y|z;\theta)} + \log(\mathrm{Var}(Y|z;\theta)), \qquad (2.4)$$

with respect to the parameters. The theory of estimating functions can be used to prove that the estimators obtained by this procedure are consistent and asymptotically normal. The standard errors and confidence intervals (CI) for the parameters are obtained by the parametric bootstrap approach (Efron & Tibshirani 1993).
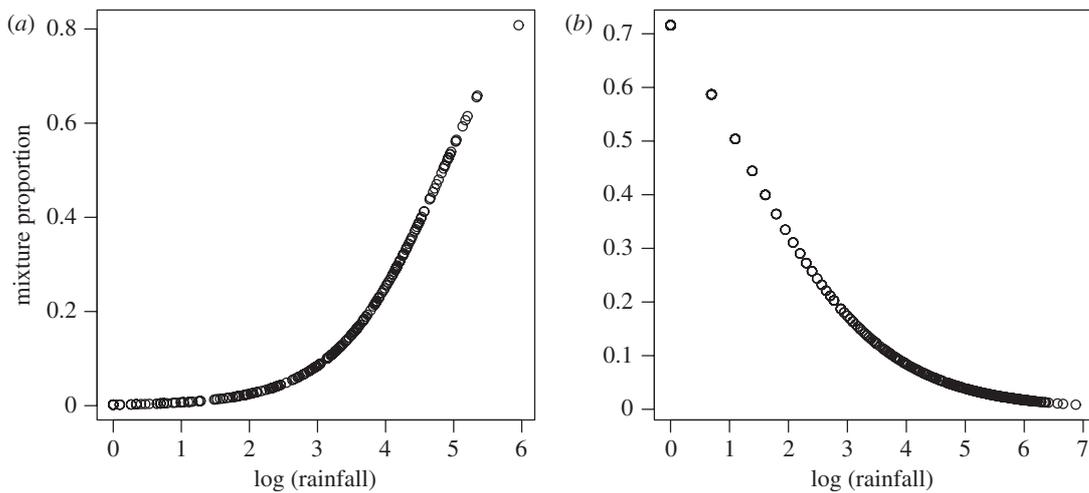
Figure 2. Estimated mixing proportion ($\pi(z)$) for (*a*) the leptospirosis and (*b*) cholera datasets.
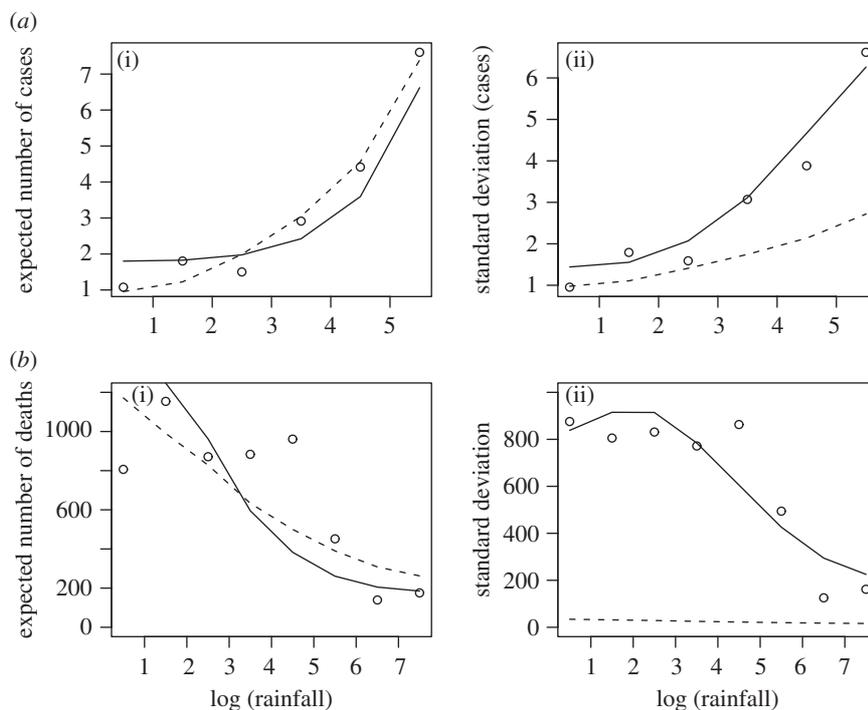

Figure 3. Observed and predicted means and standard deviations of (*a*) leptospirosis cases (i,ii) and (*b*) cholera deaths (i,ii) as a function of rainfall (open circles, observed; solid line, threshold model; dashed line, standard Poisson regression model).

3. RESULTS

Table 1 shows the point estimates and their bootstrap 95% CI for the two datasets. All calculations were performed using the R v. 1.5 statistical package (scripts are available upon request).

For leptospirosis, we found an expected incidence of 1.77 cases per week in the drought scenario, and 13.8 cases per week after extreme floods. The other parameters, a and b, describe the effect of the environmental driver on the mixture of the two processes. Since rainfall is associated with increasing reporting of leptospirosis cases, parameter b is positive. Figure 2a shows the estimated mixture function for leptospirosis. The abrupt change in the mixture proportion when log (rainfall) exceeds 2 supports the hypothesis of a threshold response to rainfall.

For cholera, the expected number of deaths in the drought scenario is 2214 (95% CI: 2132, 2348) per month. During extreme floods, deaths go down to an average of 157 (95% CI: 156,159). The mixture function shows a negative association between rainfall and deaths (figure 2b), and the estimated response is less steep than that for leptospirosis.

To evaluate goodness of fit, we compared empirically measured and model-derived estimates of mean number of cases per level of rainfall intensity, as well as its standard deviation (figure 3). Empirical estimates were obtained by calculating mean and standard deviations for the observations within the rainfall intervals: [0,1], [1,2], …, [5,6]. For the threshold model, mean and average were calculated using fitted values within the same intervals. We further included, for the sake of comparison, estimates from a standard Poisson

regression model in which expected cases are considered as a function of rainfall. From the results, it is clear that the threshold model captures the effect of rainfall reasonably well on the expected number of cases and the variance, for both diseases. By contrast, the simple Poisson regression model captures well only the expected value but not the variability.

4. DISCUSSION

The probabilistic model introduced here allows the estimation of nonlinear functional responses of ecological systems to environmental variables. Contrary to previous approaches, this model is capable of capturing sharp transitions between system states without resorting to *a priori* definition of cut points. The presence of thresholds is left to the user as it is not imposed by the model. For example, in the leptospirosis case, the steep mixing function suggests the presence of a threshold response to rainfall; hypothetically, when rainfall exceeds the structure of the city's run-off system, flooding occurs. In the absence of flooding, leptospira contamination is restricted to a few sites; in its presence, contamination spreads to large areas and the number of exposed individuals increases drastically. From the estimated mixing function, a threshold point for hospital alerts could be determined and used in combination with rainfall forecasts.

On the other hand, a less sharp mixing function is estimated for cholera and may not be compatible with a mechanistic threshold (on–off) response. The relationship between rainfall and cholera is known to be more elusive, involving both positive and negative responses depending on geographical region, temporal lag and overall water availability in the environment (Pascual *et al.* 2002). The negative nonlinear response at zero lag described here for seasonal cholera to rainfall in Parganas is also present in other districts of former Bengal (results not shown). Future work should examine the basis for a positive effect of rainfall in other regions (e.g. Punjab), for longer lags than that of the dilution effect itself (e.g. Madras; Ruiz-Moreno *et al.* 2007), and at longer (interannual) temporal scales as indicated by the recently reported association between floods and cholera transmission in Matlab, Bangladesh (Koelle *et al.* 2005).

As mentioned in §1, sharp transitions between regimes have been observed in many ecological and epidemiological systems. For epidemiological systems, the identification of thresholds is important for the development of climate-based alert systems (Curriero *et al.*, 2001; Charron *et al.*, 2004; Grover-Kopec *et al.* 2005; Platonov 2006); for ecological systems, to evaluate the impact of environmental changes on ecological dynamics. More generally, a characterization of the nature of the transition between states may serve as a basis for the development of more mechanistic models for intervention analysis.

The threshold model we have developed is static in the sense that it does not involve the temporal dynamics of the system itself. As such, the model is concerned with the estimation of functional responses to environmental drivers and differs in scope from the dynamical models of Grenfell *et al.* (1998) and Stenseth *et al.* (2004). The latter are time-series autoregressive models in which the threshold is not a function of an environmental covariate but of the state of the system itself, specifically of the (log) population density in the previous year. The population evolves in time according to a piecewise linear equation whose parameters and additive noise level vary across the threshold. Future developments of our model should investigate two possible directions. The first one is to incorporate probabilistic threshold responses to environmental covariates within time-series models for ecological dynamics. This could be done within the framework of new inference methods for stochastic dynamical systems (e.g. Ionides *et al.* 2006) to examine, for example, the association of residuals with environmental covariates. The second one is to incorporate the type of threshold considered here, involving a mixture of stochastic processes and changes in the variance, within time-series models in which the autoregressive variable itself exhibits a threshold, as in Grenfell *et al.* (1998) and Stenseth *et al.* (2004). More generally, the resulting approaches should be applicable to threshold dynamics in other types of ecological systems (e.g. Scheffer & Carpenter 2003).

REFERENCES

Anderson, R. & May, R. 1992 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.

Charron, D., Thomas, M., Waltner-Toews, D., Aramini, J., Edge, T., Kent, R., Maarouf, A. & Wilson, J. 2004 Vulnerability of waterborne diseases to climate change in Canada: a review. *J. Toxicol. Environ. Health A* **67**, 1667–1677. (doi:10.1080/15287390490492313)

Codeço, C. T. 2001 Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC Infect. Dis.* **1**, 1–39. (doi:10.1186/1471-2334-1-1)

Curriero, F. C., Patz, J. A., Rose, J. B. & Lele, S. 2001 The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. *Am. J. Public Health* **91**, 1194–1199.

Efron, B. & Tibshirani, R. J. 1993 *An introduction to the bootstrap*. New York, NY: Chapman and Hall/CRC.

Ellis, A. M. & Post, E. 2004 Population response to climate change: linear vs. non-linear modeling approaches. *BMC Ecol.* **4**, 2. (http//www.biomedcentral.com/1472-6785/4/2)

Engel, R., Normand, N., Horowitz, J. & Peleg, M. 2001 A model of microbial contamination of a water reservoir. *Bull. Math. Biol.* **63**, 1005–1023. (doi:10.1006/bulm.2001.0238)

Faruque, S. M., Naser, I. B., Islam, M. J., Faruque, A. S. G., Ghosh, A. N., Nair, G. B., Sack, D. A. & Mekalanos, J. J. 2005 Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. *PNAS* **102**, 1702–1707. (doi:10.1073/pnas.0408992102)

Flannery, B. *et al.* 2001 Referral pattern of leptospirosis cases during a large urban epidemic of dengue. *Am. J. Trop. Med. Hyg.* **5**, 657–663.

Godambe, V. P. (ed.) 1991 *Estimating functions*. Oxford, UK: Oxford University Press.

Grenfell, B. T., Wilson, K., Finkenstadt, B. F., Coulson, T. N., Murray, S., Albon, S. D., Pemberton, J. M., Clutton-Brock, T. H. & Crawley, M. J. 1998 Noise and determinism in synchronized sheep dynamics. *Nature* **394**, 674–677. (doi:10.1038/29291)

Grover, J. P. 1997 *Resource competition*. New York, NY: Springer.

Grover-Kopec, E., Kawano, M., Klaver, R. W., Blumenthal, B., Ceccato, P. & Connor, S. J. 2005 An online operational rainfall-monitoring resource for epidemic malaria early warning systems in Africa. *Malar. J.* **4**, 6. (doi:10.1186/1475-2875-4-6)

Huq, A., West, P. A., Small, E. B., Huq, M. I. & Colwell, R. 1984 Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio cholerae* serovar 01 associated with live copepods in laboratory microcosms. *Appl. Environ. Microbiol.* **48**, 420–424.

Ionides, E. L., Bretó, C. & King, A. 2006 A inference for nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **103**, 18 438–18 443. (doi:10.1073/pnas.0603181103)

Jacobson, A. R., Provenzale, A., Hardenberg, A. V., Bassano, B. & Festa-Bianchet, M. 2004 Climate forcing and density dependence in a mountain ungulae population. *Ecology* **85**, 1598–1610.

Karande, S., Bhatt, M., Kelkar, A., Kulkarni, M., De, A. & Varaiya, A. 2003 An observational study to detect leptospirosis in Mumbai, India, 2000. *Arch. Dis. Child.* **12**, 1070–1075. (doi:10.1136/adc.88.12.1070)

Koelle, K., Rodo, X., Pascual, M., Yunus, M. & Mostafa, G. 2005 Refractory periods and climate forcing in cholera dynamics. *Nature* **436**, 696–700. (doi:10.1038/nature03820)

Lele, S. & Taper, M. L. 2002 A composite likelihood approach to estimation of (co)variance components. *J. Stat. Plan. Inference* **103**, 117–135. (doi:10.1016/S0378-3758(01)00215-4)

Lloyd-Smith, J. O., Cross, P. C., Briggs, C. J., Daugherty, M., Getz, W. M., Latto, J., Sanchez, M. S., Smith, A. B. & Swei, A. 2005 Should we expect population thresholds for wildlife disease? *Trends Ecol. Evol.* **20**, 511–519.

Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faruque, A. S. G. & Colwell, R. 2000 Climate and infectious disease: use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *PNAS* **97**, 1438–1443. (doi:10.1073/pnas.97.4.1438)

McLachlan, G. & Peel, D. A. 2000 *Finite mixture-models*. Manitoba, Canada: Wiley.

Miller, C. J., Drasar, B. S. & Feachem, R. G. 1982 Cholera and estuarine salinity in Calcutta and London. *Lancet* **319**, 1216–1218. (doi:10.1016/S0140-6736(82)92340-6)

Pascual, M., Bouma, M. J. & Dobson, A. P. 2002 Cholera and climate: revisiting the quantitative evidence. *Microbes Infect.* **4**, 237–245. (doi:10.1016/S1286-4579(01)01533-7)

Platonov, A. E. 2006 The influence of weather conditions on the epidemiology of vector-borne diseases by the example of West Nile fever in Russia. *Vestn. Akad. Med. Nauk SSSR* **2**, 25–29.

Ruiz-Moreno, D., Pascual, M., Bouma, M., Dobson, A. & Cash, B. 2007 Cholera seasonality in madras (1901–1940): dual role for rainfall in endemic and epidemic regions. *EcoHealth* **4**, 52–62. (doi:10.1007/s10393-006-0079-8)

Sarkar, U. *et al.* 2002 Population-based case-control investigation of risk factors for leptospirosis during an urban epidemic. *Am. J. Trop. Med. Hyg.* **66**, 605–610.

Scheffer, M. & Carpenter, S. R. 2003 Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends Ecol. Evol.* **18**, 648–656. (doi:10.1016/j.tree.2003.09.002)

Smith, V. H. 1983 Low nitrogen to phosphorus ratios favor dominance by blue-green algae in lake phytoplankton. *Science* **221**, 669–671. (doi:10.1126/science.221.4611.669)

Stenseth, N. C., Chan, K. S., Tavecchia, G., Coulson, T., Mysterud, A., Clutton-Brock, T. & Grenfell, B. 2004 Modelling non-additive and nonlinear signals from climatic noise in ecological time series: soy sheep as an example. *Proc. Biol. Sci.* **271**, 1985–1993. (doi:10.1098/rspb.2004.2794)

WHO 2005 Flooding and communicable diseases fact sheet. *Weekly Epidem. Record* **80**, 21–28.