

Malaria seasonality and rainfall seasonality in Sri Lanka are correlated in space

Olivier J.T. Briët^{1,2}, Penelope Vounatsou², Priyanie H. Amerasinghe³

¹International Water Management Institute, P.O. Box 2075, Colombo, Sri Lanka; ²Swiss Tropical Institute, Socinstrasse 57, P.O. Box CH-4002, Basel, Switzerland; ³International Water Management Institute Sub Regional Office for South Asia, c/o ICRISAT, Patancheru, AP 502 324, Andhra Pradesh, India

Abstract. Correlation in space between seasonality of malaria and seasonality of rainfall was studied in Sri Lanka. A simple seasonality index was developed by making use of the bimodal seasonality of both malaria and rainfall. The malaria seasonality index was regressed against the rainfall seasonality index taking spatial autocorrelation into account. Despite the presence of spatial autocorrelation, the coefficient for the rainfall seasonality index in explaining the malaria seasonality index was found to be significant. The results suggest that rainfall is an important driver of malaria seasonality.

Keywords: malaria, rain, seasonality, spatial autocorrelation, Sri Lanka.

Introduction

In the field of malaria transmission and prediction modelling there is not only an interest in risk *per se* but also an increasing interest in identifying the seasonality of malaria over larger geographical areas (Mabaso et al., 2005, 2007; Childs et al., 2006; Gemperli et al., 2006; Grover-Kopec et al., 2006). Malaria case time-series in Sri Lanka show both strong long-term fluctuations and seasonality (Fig. 1). The long-term fluctuations are generally attributed to the impact of malaria control strategies and the development of insecticide resistance. Seasonality, on the other hand, is generally attributed to climatic factors, in particular rainfall which provides the breeding habitats for the malaria vector mosquitoes and sustains the aquatic, immature stages of their life cycle. The optimum amount and frequency of rainfall depend on the physical nature of the breed-

ing site and the requirements of the vector in question. It should be remembered that rainfall may also have an indirect impact on mosquito breeding, e.g. when it occurs upstream or when rain water is transported for irrigation purposes. Hence, rainfall impinges on mosquito population dynamics in a rather complex way. For example, a large amount of rain within a short period of time may wash away aquatic stages as well as adults, while continuous,

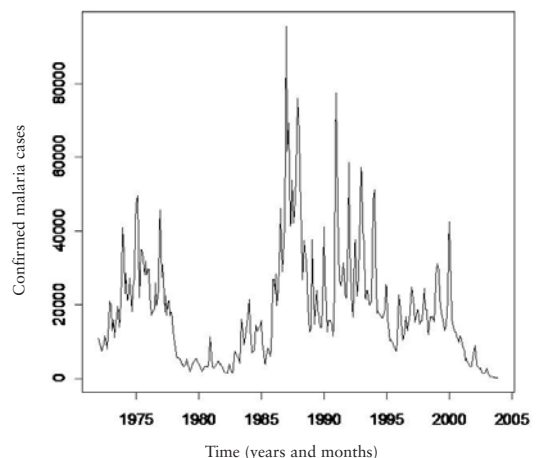


Fig. 1. Confirmed monthly malaria cases in Sri Lanka for the period January 1972 - December 2003.

Corresponding author:

Olivier J.T. Briët

International Water Management Institute

P.O. Box 2075, Colombo, Sri Lanka

Tel. +94 112 787 404; Fax +94 112 786 854

Email: o.briet@cgiar.org; olivier.briet@unibas.ch;

o.briet@gmail.com

low-volume rainfall may not be optimal for colonizing mosquito species that require temporary breeding sites. Although malaria-case time-series (from which long term non-seasonal trends were removed) and rainfall time-series appear to have strong cross-correlations, a large part of these correlations can be explained by both series being cyclical with a similar periodicity (Briët et al., 2008). Many biological processes follow annual cycles and high cross-correlations do not necessarily infer a causal link. However, there is spatial information that suggests that rainfall seasonality could be a driver of malaria seasonality and although the island of Sri Lanka is only 65,610 km², in total, it shows strong spatial variability in climate (de Silva and Fernando, 1997). The south-western part of the country (often described as the

wet zone since it receives more than 1900 mm of rain annually) is affected by two periods of monsoon rains with peaks in May and October, whereas rainfall peaks in November/December with a very minor, almost imperceptible peak in April (Fig. 2) in the so called dry zone in the north-east which receives less than 1900 mm rain annually. Corresponding to this distribution of rainfall, malaria-case time-series show a strong bimodal seasonality in the south-western part of the island, whereas the malaria time-series become more monomodal in nature towards the north and east with the second peak in the middle of the year being much less important (Fig. 3). In order to establish whether there is a correlation between rainfall seasonality and malaria seasonality over space, a regression analysis was carried out.

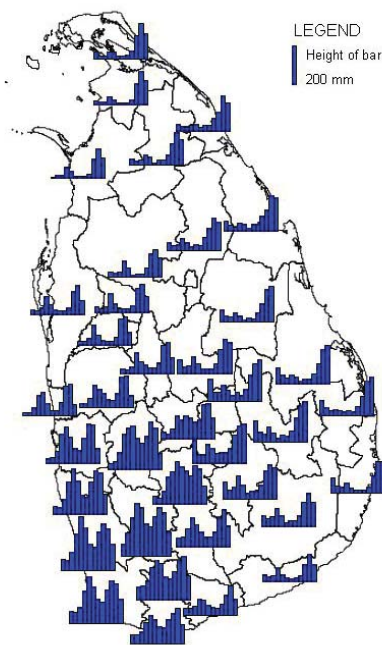


Fig. 2. Seasonal variability of rainfall (geometric means) for the period January 1972 - December 2003 in Sri Lanka. Legend: the height of the bars should be compared to that of the standard bar which corresponds to 200 mm of rain. The bar to the far left in each icon represents the geometric mean of the monthly rainfall in January and that to the far right the geometric mean of the monthly rainfall in December.

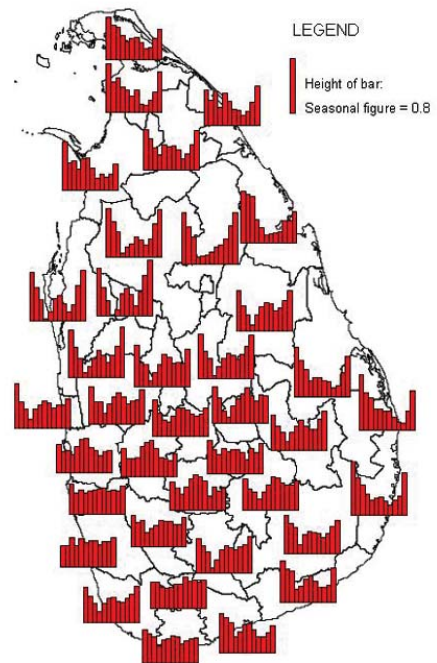


Fig. 3. The seasonal distribution of malaria cases in Sri Lanka. Legend: the compound number of monthly malaria cases (logarithmically transformed) from January (bar far left) to December (bar far right) from 1972 to 2003. The value of 0.6 has been added to all data to make them positive.

Materials and methods

Malaria and rainfall data

Records of the total count of blood films examined for malaria, and how many of these are positive for malaria, are reported monthly by government health facilities and aggregated by the medical officer responsible for each so called health area (a sub-district health administrative division). This study is based on information regarding such blood film counts provided by the Anti Malaria Campaign (AMC) of Sri Lanka for the period 1972 - 2003. For some of the records, the number of blood films examined was marked as "not received" (and therefore classified as missing). For 14.9% of the records, the value was given as zero or left blank. For the latter records (the blanks), there was ambiguity as to whether the data were missing due to problems in data recording or whether they could also be taken as zero, i.e. no patients presented themselves for examination in that particular area in that particular month. In the data cleaning procedure (see statistical methods below), 1.4% of the records were declared as not available (NA). This included the records where the place for entering the number of blood films examined was marked as "not received" (0.95% of all the records) and the records for which the number of the blood films could be classified as a lower additive outlier (Burman and Otto, 1988) (0.44% of all records). The data from the districts in the north and east, where data gathering and reporting was affected by the armed conflict, had the largest percentage of NA labels: Jaffna (5.4%), Mannar (26.1%), Vavuniya (8.9%), Kilinochchi (2.0%), Trincomalee (2.0%) and Ampara (5.4%). Over time, some health areas changed boundaries or split into two, e.g. in 1972 and 2003, the number of health areas was 98 and 230, respectively, and were therefore deemed unsuitable for temporal data aggregation. For the purpose of this study, health areas with variable boundaries were aggregated into larger areas corresponding to malaria data for which the catchment area did not change over the

1972 - 2003 period. Thus, the surface of Sri Lanka was divided into 37 areas (Figs. 2 and 3) and the "cleaned" monthly malaria-positive data were aggregated accordingly. Precipitation records, collected by 342 stations across the island, were purchased from the Meteorological Department of Sri Lanka and monthly rainfall surfaces were created through spatial prediction using kriging. Three stations with consistently aberrant rainfall records, detected through cross-validation using kriging (Ribeiro and Diggle, 2001), were removed from the dataset. From each monthly rainfall surface, the average value of rainfall was extracted for each area.

Statistical methods

In a data cleaning procedure, the time series of blood film counts in health areas were logarithmically transformed to normality (after the value one was added to the data). Under the null hypothesis, each observation was assumed to be part of a seasonal autoregressive integrated moving average (SARIMA) process (Box and Jenkins, 1968) with parameters $P = 0$, $d = 1$, $q = 1$, $P = 0$, $D = 1$, and $Q = 1$. Observations were marked as additive outliers if the likelihood ratio test statistic (for an additive outlier) for the observation was below a threshold of 6 (Burman and Otto, 1988) and classified as NA. For each of these and other NA observations that were not at the beginning or at the end of a series, values for the number of malaria-positive blood films were estimated through a one-step-ahead SARIMA forecasting model on the original series and on the reversed series. These two estimates were then averaged. This approach has been discussed by Mwaniki et al. (2005). Finally, the health area data series were aggregated to district resolution before analysis, as these spatial units remained constant over the study period, whereas for many health areas boundaries changed (within the district boundaries) over the study period.

For each area and for each calendar month of the year, the 34-year (the period matching the malaria

data available) mean rainfall was calculated and the values logarithmically transformed. For each area, the first rainfall peak was calculated as the sum of the rainfall climatology during the calendar months March - August, and the second rainfall peak was calculated as the sum of the rainfall climatology during the calendar months September - February. The logarithmically transformed ratio of the two peaks was used as an index of rainfall seasonality. For malaria-case count time-series, a similar procedure was applied except that the long-term trends were calculated using a 13-point moving average filter with the coefficients at the extremes given half weight (Chatfield, 2004) and removed. Also, the first malaria peak was calculated as the sum of the seasonal figure during the calendar months April - September and the second malaria peak as the sum of the seasonal figure during the calendar months October - March.

The distribution of the malaria seasonality index was tested for normality using the Shapiro-Wilk test (Royston, 1982). The presence of spatial autocorrelation of the malaria seasonality index among areas was tested with the Moran's I test (Moran, 1950).

Let y_i be the malaria seasonality index in area i , $i = 1, \dots, 37$. It was assumed that $y_i \sim$ arise from a normal distribution with mean μ_i and precision parameter τ , that is $y_i \sim N(\mu_i, \tau)$. We considered $\mu_i = \beta_0 + \beta_1 x_i + \phi_i$ where x_i measures rainfall seasonality. β_0, β_1 are regression coefficients and ϕ_i is an area-specific random effect taking into account the spatial correlation introduced by the spatial structure of unobserved covariates. For the ϕ_i 's, a CAR(γ, τ_s) model was adopted which assumes that

$$\phi_i | \phi_{-i}, \tau_s, \gamma \sim N \left(\frac{\gamma \sum \phi_{-i}}{n_i}, \frac{1}{n_i \tau_s} \right)$$

where γ is a spatial correlation parameter, n_i is the number of areas bordering area i and $\sigma_i^2 = \frac{1}{\tau_s}$ measures between area variation. To facilitate model fit a Bayesian modelling framework was used. The following prior distributions were chosen for the parameters: $\beta_0, \beta_1 \sim U(-\infty, \infty)$, $\tau, \tau_s \sim \text{Ga}(0.005, 0.005)$, $\gamma \sim U(a, b)$ with limits a, b specified as

described by Gelfand and Vounatsou (2003). The CAR(γ, τ_s) was considered as a prior distribution for the ϕ_i 's. The effect of this prior distribution is to shrink the observed value (in this case the malaria seasonality index) of an area to that of the local mean, where the local mean is the mean of all contiguous areas excluding the area itself. The posterior distribution of the seasonality index of an area is therefore a compromise between the prior, which is based on the seasonality index of neighbouring areas, and the data for the area. Two spatial models were fitted: (i) a CAR(γ, τ_s) and (ii) a CAR(1, τ_s). The latter model assumes maximum spatial correlation although it does not give a proper distribution for the ϕ_i 's (Sun et al., 2000). The former model gives a well-defined proper distribution. Bayesian CAR models have been widely used in malaria mapping (Kazembe et al., 2006; Kleinschmidt et al., 2002; Mabaso et al., 2005). A non spatial model was also applied. The deviance information criterion (DIC) (Spiegelhalter et al., 2002) was used to determine the best fitting model. The models were estimated using a Markov chain Monte Carlo process using three chains, and 1,500,000 iterations (including a burn-in of 500,000 iterations), with a thinning rate of 100. Convergence was assessed by studying plots of the Gelman-Rubin convergence statistic as modified by Brooks and Gelman (1998).

The including analysis described above was repeated including zone as a regressor. Also, because there is some concern that cases in areas with historically low transmission may primarily have been acquired elsewhere (and patterns would therefore not represent local transmission) the analysis (without zone as regressor) was repeated excluding nine areas with a geometric mean annual case-load of less than 600 over the study period (all of these except Chilaw were situated in the wet zone and some were situated at high elevations): Colombo district, Kalutara district, Galle district, the northern part of Matara district, the western part of Ratnapura district, Nuwara Eliya district (comprising of two areas), eastern part of Kandy district, and Chilaw (the southern part of Puttalam district).

All data management and analysis was performed in the software environment R (2007). The Bayesian regression analysis was carried out in the software package “WinBUGS” (Lunn et al., 2000), which can be called from R using the R package “R2WinBUGS”.

Results

There was no evidence that the malaria seasonality index, nor the residuals of any of the regression analyses, were not Gaussian distributed, according to the Shapiro-Wilk test. A scatter plot of the malaria seasonality index against the rainfall seasonality index (Fig. 4) showed that there are no clear outliers. The Moran’s I test for the complete dataset (n = 37, I = 0.46, 95% credible interval = 0.45 - 0.52) and the reduced dataset (n = 28, I = 0.45, 95% credible interval = 0.42 - 0.54) showed that there was significant spatial autocorrelation in the malaria seasonality index, thus nearby pairs of districts had a more similar malaria seasonality index than distant pairs.

Estimates for the mean of the coefficients in the

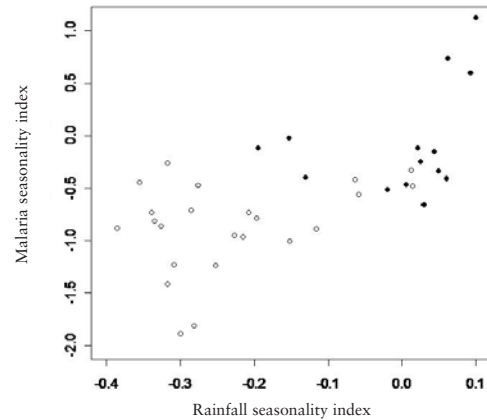


Fig. 4. Scatter-plot of malaria seasonality versus rainfall seasonality in Sri Lanka.

Legend: the malaria seasonality index on the vertical axis is viewed against the rainfall seasonality index on the horizontal axis for areas in the wet zone receiving more than 1900 mm rainfall annually (solid circles) and in the dry zone (open circles).

non-spatial, CAR(1, τ_s) and CAR(γ , τ_s) models (Table 1) were very similar. The standard deviations of the coefficients for the rainfall seasonality index in explaining malaria seasonality were larger (as

Table 1. Results of regression of the malaria seasonality index against the rainfall seasonality index in malarious areas in Sri Lanka.

| Model | Parameter | Mean | Standard deviation | 95% Credible interval | |
|-----------------------------|-----------|-------|--------------------|-----------------------|--------|
| Non spatial | β_0 | -0.20 | 0.11 | -0.41, | 0.01 |
| | β_1 | 2.53 | 0.51 | 1.54, | 3.55 |
| | τ | 4.65 | 1.10 | 2.68, | 6.92 |
| | Moran’s I | 0.16 | 0.04 | 0.12, | 0.27 |
| | DIC | 52.18 | | | |
| CAR (1, τ_s) | β_0 | -0.18 | 0.14 | -0.44, | 0.09 |
| | β_1 | 2.69 | 0.73 | 1.29, | 4.19 |
| | τ | 10.34 | 23.34 | 2.99, | 52.50 |
| | τ_s | 40.90 | 80.37 | 1.00, | 276.70 |
| | Moran’s I | 0.11 | 0.11 | -0.10, | 0.36 |
| | DIC | 45.47 | | | |
| CAR (γ , τ_s) | β_0 | -0.19 | 0.13 | -0.44, | 0.05 |
| | β_1 | 2.61 | 0.56 | 1.53, | 3.73 |
| | τ | 10.71 | 32.20 | 2.90, | 61.91 |
| | τ_s | 53.50 | 93.04 | 0.94, | 314.21 |
| | γ | -0.03 | 0.70 | -1.37, | 0.96 |
| | Moran’s I | 0.12 | 0.09 | -0.09, | 0.27 |
| | DIC | 41.10 | | | |

Legend: tau = precision = 1/variance; gamma = spatial auto correlation coefficient; DIC = deviance information criterion.

expected) in the spatial models as compared to the non-spatial model. The 95% credible interval of the posterior distribution of the rainfall seasonality index coefficient did not include zero. The CAR (γ , τ_s) was the best fitting model because the DIC had the smallest value.

In the CAR(γ , τ_s) model with both the rainfall index and the zone as covariate, the 95% credible interval of the posterior distribution of the rainfall coefficient was larger but still did not include zero (results not shown). It should be noted that the zone coefficient was highly significant in explaining the rainfall seasonality index ($P < 0.00001$), thus collinearity may have played an important role. There was no evidence for zone-rainfall seasonality interaction (analysis not shown). After accounting for the effect of the rainfall seasonality (and the zone effect), the residuals in the non-spatial model showed a much weaker, albeit significant, spatial autocorrelation based on the Moran's I.

Discussion

Spatial autocorrelation in the rainfall seasonality index accounted for most of the spatial autocorrelation in the malaria seasonality index, as shown by the comparison of the Moran's I of the raw data with the Moran's I of the residuals of a non-spatial model including the rainfall seasonality index. The additional spatial correlation could be due to unobserved variables that change gradually over space, such as those related to soil conditions and altitude (temperature), or factors intrinsic to malaria transmission (as nearby districts may influence other districts) or due to cross-border reporting of cases. However, the analysis excluding those areas that were thought particularly affected by cross-border reporting yielded a similar spatial correlation index. The CAR (γ , τ_s) model showed the lowest DIC. However, there are some doubts as to whether this model really performed this well as the gamma parameter was not significantly different from zero. Nevertheless, since the regression

models indicated that the 95% credibility intervals for the coefficient for rainfall did not include zero in the spatial models, there was evidence for a significant correlation between seasonality of rainfall and malaria.

The work most similar to the work presented here is that by Mabaso et al. (2005, 2007) who made use of a seasonal concentration index to summarize seasonality in malaria incidence and entomological inoculation rate (EIR). The seasonal concentration index is based on vector representation (i.e., both magnitude and direction) of the mean monthly values in a given year. Mabaso et al. (2007) found that the seasonal concentration index of rainfall was significant in explaining the seasonal concentration index of EIR across Africa. The EIR is more closely related to environmental variables and is a better measure of the risk of inoculation than reported malaria cases, which is confounded by immunological processes. When protective immunity is high as a result of high levels of transmission, as is the case in many African settings, it strongly confounds the relationship between the (all age) incidence time-series and transmission (Smith et al., 1993). When immunity plays an important role in disease transmission, it may create and maintain so called "endogenous cycles" in incidence time-series, even when the vectorial capacity (the mosquito vector population's potential to transmit malaria) is at a constant level. Cyclical patterns in incidence time-series may thus partly be caused by immunity dynamics and this may confound the relationship between incidence and extrinsic drivers such as weather (Pascual et al., 2008). Because of extremely low sporozoite rates in vector mosquitoes in Sri Lanka, it is difficult to measure the EIR which is otherwise a good measure of the risk of inoculation. In the absence of reliable EIR estimates, the malaria-case incidence might be a satisfactory measure of parasite transmission, provided that protective and anti disease immunity is low. Although there are strong similarities between the seasonal concentration index and the

seasonality index presented here, there are also important differences. For example, neither differentiate between bimodal systems with (evenly spaced) peaks of similar height and systems without seasonality. However, the seasonality index is continuous, whereas the seasonal concentration index is contained in the zero-one (zero-100 if expressed as percentage) interval. Most importantly, in bimodal systems, the seasonality index allow differentiating between a situation with the first peak being higher than a second (and the reverse situation), whereas the seasonal concentration index does not differentiate.

In this paper, seasonality in temperature was not studied as it was considered of less importance than rainfall seasonality in the Sri Lankan context. The country is situated close to the equator and the temperature therefore fluctuates over a narrow range. Also, a large part of its temporal variability is governed by rainfall. Moreover, for most malarious areas (except in the hilly part of the country situated just south of the centre of the island), the temperature (varying between 22 and 32 degrees Celsius) is well within the range suitable for malaria transmission (Craig et al., 1999). A study in Ethiopia found temperature to be generally not significant in explaining malaria for districts below 1,650 m (Teklehaimanot et al., 2004). However, it merits investigation whether temperature influences malaria seasonality in the hills.

This paper provides evidence that, (even) after correction for spatial autocorrelation in the data, rainfall seasonality is significant in explaining malaria seasonality in space. This suggests that high cross-correlations between rainfall time-series and malaria time-series found elsewhere (Briët et al., 2008) are not accidental but that rainfall is a driver of seasonality of malaria cases. Rainfall seasonality could thus in theory be used as a predictor of the seasonality of malaria transmission in the absence of malaria case data or EIR data in areas of low transmission and temperatures which are conducive to malaria transmission year round.

Acknowledgements

The author acknowledges Dr. G.N.L. Galappaththy and the Directorate of the AMC and Regional Malaria Officers and their teams for making surveillance data available, and Dr. D.M. Gunawardena for helpful comments to the earlier draft versions of this paper. The author is funded through the National Oceanic and Atmospheric Administration (NOAA), National Science Foundation (NSF), Environmental Protection Agency (EPA) and Electric Power Research Institute (EPRI) Joint Program on Climate Variability and Human Health.

References

- Box GEP, Jenkins GM, 1968. Some recent advances in forecasting and control. *Appl Statist* 17, 91-109.
- Briët OJT, Vounatsou P, Gunawardena DM, Galappaththy GNL, Amerasinghe PH, 2008. Temporal correlation between malaria and rainfall in Sri Lanka. *Malar J*, in press.
- Brooks SP, Gelman A, 1998. Alternative methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 7, 434-455.
- Burman JP, Otto MC, 1988. Outliers in time series. SRD Research Report Number: CENSUS/SRD/RR-88114.
- Chatfield C, 2004. *The analysis of time series: an introduction*, Chapman & Hall/CRC, 6th ed., pp. 333.
- Childs DZ, Cattadori IM, Suwonkerd W, Prajakwong S, Boots M, 2006. Spatiotemporal patterns of malaria incidence in northern Thailand. *Trans R Soc Trop Med Hyg* 100, 623-631.
- Craig MH, Snow RW, Le Sueur D, 1999. A climate based distribution model of malaria transmission in sub-Saharan Africa. *Parasitol Today* 15, 105-111.
- de Silva MBG, Fernando TK, 1997. Climate. In: Arjuna's Atlas of Sri Lanka. Arjuna's Consulting Co., 1st edition, 16-22.
- Gelfand AE, Vounatsou P, 2003. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* 4, 11-25.
- Gemperli A, Sogoba N, Fondjo E, Mabaso M, Bagayoko M, Briët OJ, Anderegg D, Liebe J, Smith T, Vounatsou P, 2006. Mapping malaria transmission in West and Central Africa. *Trop Med Int Health* 11, 1032-1046.

- Grover-Kopec EK, Blumenthal MB, Ceccato P, Dinku T, Omumbo JA, Connor SJ, 2006. Web-based climate information resources for malaria control in Africa. *Malar J* 5, 38.
- Kazembe LN, Kleinschmidt I, Sharp BL, 2006. Patterns of malaria-related hospital admissions and mortality among Malawian children: an example of spatial modelling of hospital register data. *Malar J* 5, 93.
- Kleinschmidt I, Sharp B, Mueller I, Vounatsou P, 2002. Rise in malaria incidence rates in South Africa: a small-area spatial analysis of variation in time trends. *Am J Epidemiol* 155, 257-264.
- Lunn DJ, Thomas A, Best N, Spiegelhalter D, 2000. WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 10, 325-327.
- Mabaso ML, Craig M, Ross A, Smith T, 2007. Environmental predictors of the seasonality of malaria transmission in Africa: the challenge. *Am J Trop Med Hyg* 76, 33-38.
- Mabaso ML, Craig M, Vounatsou P, Smith T, 2005. Towards empirical description of malaria seasonality in southern Africa: the example of Zimbabwe. *Trop Med Int Health* 10, 909-918.
- Moran PAP, 1950. Notes on continuous stochastic phenomena. *Biometrika* 37, 17-23.
- Mwaniki KJ, Odhiambo OR, Charles W, 2005. Seasonal time series data imputation: comparison between feedforward neural networks and parametric approaches. *East Afr J Stat* 1, 68-83.
- 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.
- Ribeiro Jr. PJ, Diggle PJ, 2001. geoR: a package for geostatistical analysis. *R-News* 1.
- Royston P, 1982. An extension of Shapiro and Wilk's *W* test for normality to large samples. *Appl Stat* 31, 115-124.
- Smith T, Charlwood JD, Kihonda J, Mwankusye S, Billingsley P, Meuwissen J, Lyimo E, Takken W, Teuscher T, Tanner M, 1993. Absence of seasonal variation in malaria parasitaemia in an area of intense seasonal transmission. *Acta Trop* 54, 55-72.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A, 2002. Bayesian measures of model complexity and fit (with discussion). *J R Stat Soc Ser B* 64, 583-640.
- Sun D, Tsutakawa RK, Kim H, He Z, 2000. Spatio-temporal interaction with disease mapping. *Stat Med* 19, 2015-2035.
- 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 effects reflect biological mechanisms. *Malar J* 3, 41.