



Assessing the effects of air temperature and rainfall on malaria incidence: an epidemiological study across Rwanda and Uganda

Felipe J. Colón-González,^{1,2} Adrian M. Tompkins,¹ Riccardo Biondi,^{1,3} Jean Pierre Bizimana,⁴ Didacus Bambahi Namanya⁵

¹Abdus Salam International Centre for Theoretical Physics, Trieste, Italy; ²School of Environmental Sciences, University of East Anglia, Norwich, UK; ³Wegener Center for Climate and Global Change, University of Graz, Graz, Austria; ⁴Centre for Geographic Information Systems and Remote Sensing, University of Rwanda, Butare, Rwanda; ⁵Ministry of Health, Kampala, Uganda

Abstract

We investigate the short-term effects of air temperature, rainfall, and socioeconomic indicators on malaria incidence across Rwanda and Uganda from 2002 to 2011. Delayed and nonlinear effects of temperature and rainfall data are estimated using generalised additive mixed models with a distributed lag nonlinear specification. A time series cross-validation algorithm is implemented to select the best subset of socioeconomic predictors and to define the degree of smoothing of the weather variables. Our findings show that trends in malaria incidence agree well with variations in both temperature and rainfall in both countries, although factors other than climate seem to play an important role too. The estimated short-term effects of air tem-

perature and precipitation are nonlinear, in agreement with previous research and the ecology of the disease. These effects are robust to the effects of temporal correlation. The effects of socioeconomic data are difficult to ascertain and require further evaluation with longer time series. Climate-informed models had lower error estimates compared to models with no climatic information in 77 and 60% of the districts in Rwanda and Uganda, respectively. Our results highlight the importance of using climatic information in the analysis of malaria surveillance data, and show potential for the development of climate-informed malaria early warning systems.

Introduction

Despite the global contraction in range over the past century (Gething *et al.*, 2010), malaria still imposes a significant health and socioeconomic burden to many countries (WHO, 2013). The World Health Organization estimates that about 3.4 billion people are at risk of malaria (WHO, 2013). Approximately 207 million cases and 627,000 deaths occurred in 2012 worldwide (WHO, 2013). About 90% of the total mortality occurs in sub-Saharan Africa, and 77% of that percentage happens in children under 5 years of age (WHO, 2013). Two countries significantly affected by malaria are Rwanda and Uganda. Malaria has long been considered the main cause of morbidity and mortality in both countries (NISR, MOH and ICF International, 2012; UBOS and ICF International, 2012). Over the period 2002 to 2011, more than five million malaria cases were reported in Rwanda to government health facilities. The number was significantly greater in Uganda with about 100 million reports of suspected malaria cases between 2002 and 2010.

Trends in malaria incidence could be attributed to the complex interplay of a range of determinants including climatic, environmental, and socioeconomic factors (Kazembe *et al.*, 2006; Lowe *et al.*, 2013; Rulisa *et al.*, 2013). Statistical models are useful tools that allow us: i) to understand how disease outcomes change as a function of variations in their key driver; and ii) to predict disease outcomes based on the dynamics of such drivers (James *et al.*, 2013). This paper aims to investigate the ways in which malaria incidence varies as a function of short-term changes in air temperature and rainfall over the period 2002-2011 in Rwanda and Uganda. Environmental factors including climate, land-use/land cover, irrigation usage, topography and soil type largely determine the type, distribution and density of mosquito vectors and hence disease occurrence, by providing suitable conditions

Correspondence: Felipe J. Colón-González, School of Environmental Sciences, University of East Anglia, Norwich Research Park, NR4 7TJ Norwich, UK.

Tel: +44.1603.591375 - Fax: +44.1603.591327.

E-mail: F.Colon@uea.ac.uk

Key words: Malaria; Weather effects; Statistical modelling; Health.

Acknowledgements: we would like to thank two anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. This research was supported by the European Commission's Seventh Framework Research Programme under HEALTHY FUTURES (grant agreement 266327). FJCG was also part funded by the Mexican National Council for Science and Technology (CONACYT), and by The National Institute for Health Research, Health Protection Research Unit in Emergency Preparedness and Response at King's College London, London, UK.

Note: see online Appendix for supplementary figures.

Received for publication: 27 May 2015.

Accepted for publication: 11 November 2015.

©Copyright F.J. Colón-González *et al.*, 2016

Licensee PAGEPress, Italy

Geospatial Health 2016; 11(s1):379

doi:10.4081/gh.2016.379

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any non-commercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

for their development (Christophers, 1960; Gage *et al.*, 2008; Halstead, 2008; Kazembe *et al.*, 2006; Lindsay *et al.*, 1998). Of these, only climate is likely to vary significantly from year to year, although introduction of irrigation and land-use conversions, for example converting forest to agricultural use, could bring about step changes in disease transmission (Baeza *et al.*, 2011; Stryker and Bomblies, 2012). Human-related factors including migration patterns, behaviour, immunity, population density, access to health services and preventive measures, may also impact disease occurrence (Baeza *et al.*, 2011; Gage *et al.*, 2008; Naish *et al.*, 2014; Pindolia *et al.*, 2013; Reiter, 2001, 2008). Urban areas characterised by high population densities, for example, show lower malaria incidence rates than rural regions (Kelly-Hope and McKenzie, 2009). Pathogen-related factors such as parasite species also influence the spatiotemporal patterns of malaria (for example Chowell *et al.*, 2009; Lin *et al.*, 2010). Differences in the tools, criteria, and skills for diagnosis; and errors in data management may contribute to the spatiotemporal variability observed in malaria data as well (Kyabayinze *et al.*, 2012; Yeka *et al.*, 2012).

Climate data are readily available at spatial and temporal resolutions that allow their incorporation in statistical disease models (Bouزيد *et al.*, 2014; Colón-González *et al.*, 2013; Garske *et al.*, 2013; Lowe *et al.*, 2013). Socioeconomic and behavioural data, on the other hand, are often difficult to acquire at the same spatiotemporal resolution as the epidemiological data. To give one example, although important for disease transmission, datasets of internal population movements are rarely available, and where they exist, data span two to three years at most (Bengtsson *et al.*, 2011; Tompkins and McCreesh, 2016), or are derived from isolated censuses (Pindolia *et al.*, 2013). Therefore, these data are rarely included in statistical models, and must be represented by random-effects terms (Lowe *et al.*, 2013).

In this study, we primarily investigate the short-term effects of air temperature and rainfall on malaria incidence across the whole of Rwanda and Uganda, over the period 2002-2011. We use, to our knowledge, the largest panels of district-specific malaria data with a monthly temporal resolution. Our study spans a larger time frame than any previous study conducted in these two countries (Loevinsohn, 1994; Zinszer *et al.*, 2015).

Materials and Methods

Study area

Rwanda is a small landlocked country located in the Great Lakes region of eastern Africa, just below the equatorial line. It is bordered by Uganda to the north, Tanzania to the east, the Democratic Republic of the Congo to the west, and Burundi to the south. Rwanda is the most densely populated country in continental Africa with an estimated population of 11.5 million people by 2012 (World Bank, 2014) living in an area of 26,338 km². The country is divided into 30 districts within four administrative provinces (North, South, East, and West), and the City of Kigali. Rwanda is divided into three distinct geographical regions. The western and north-central region is made up mountains, the Virunga volcano range, the highlands, and the foothills of the Congo-Nile Divide (NISR, MOH and ICF International, 2012). This region has an altitude generally higher than 2000 m (4507 m at its highest point). The central region has a mountainous terrain with an elevation between 1500 and 2000 m (NISR, MOH and ICF International, 2012). The eastern plateaus region combines lowlands with some hills, and has an altitude generally below 2000 m (NISR, MOH and ICF International, 2012). These regions are intrinsically related to malaria

occurrence. Thus, malaria is mesoendemic in the lowlands, and epidemic-prone in the highlands and hills (Rulisa *et al.*, 2013; USAID, 2013a). In the endemic regions malaria transmission happens all year, and shows two peaks (in May-June, and December-January) following two distinct rainy seasons (USAID, 2013a).

Uganda is also a landlocked country in the eastern Africa region. It shares borders with South Sudan to the north, Tanzania to the south, Kenya to the east, Rwanda to the south-west, and the Democratic Republic of Congo to the west. The country has an area of approximately 241,000 km² (UBOS and ICF International, 2012). Currently, Uganda is divided into 112 districts, twice as many as the number of administrative districts it had in 2002 (56). The mean altitude of the northern part of the country reaches about 600 m, and gradually increases upward to the south reaching altitudes beyond 1200 m (MOH, 2005). Most of the south-western part of the country is composed by highlands with altitudes of approximately 1400 m (MOH, 2005). There are also some mountainous regions with altitudes of about 2000 m located in the bordering regions with Rwanda and the Democratic Republic of Congo, the Rwenzori Mountains (west), and Mount Elgon (east) (MOH, 2005). In most of the country, there are two distinct rainy seasons with heavy rains from March to May, and light rains between September and December. Malaria is reported as mesoendemic in about 95% of the country, and prone to epidemics in the south-western highlands (USAID, 2013b). Malaria transmission is largely stable throughout the year, with seasonal variations that follow the rainy seasons (UBOS and ICF International, 2009).

Malaria case data

Monthly counts of clinically diagnosed malaria cases were obtained from the corresponding Ministries of Health for the periods January 2001 to December 2011 (132 months) in Rwanda, and January 2002 to December 2010 (108 months) in Uganda. In Rwanda, malaria data was obtained at the Health Facility level. Health facility data were then aggregated at the district level based on the coordinates (latitude and longitude) of each health facility. Uganda data were obtained at the district level from the Ministry of Health, Health Management Information System. Due to changes in the number of administrative districts over time (2006 and 2010), we froze the geography of the country to its 2002 geographical distribution. Missing malaria counts were estimated using a singular value decomposition-based method proposed by Troyanskaya *et al.* (2001), and included in the *bcv* package (Perry, 2009) for R (R Development Core Team, 2010). Malaria data for the Uganda district of Kalangala were excluded from the analysis because it was not possible to retrieve air temperature estimates for this district. Malaria data for both countries were obtained through the HEALTHY FUTURES data sharing Consortium Agreement. HEALTHY FUTURES is an international and inter-institutional project funded under the European Commission's Seventh Framework Research Programme (grant agreement 266327).

Weather data

Daily global near-surface (~2 m height) mean air temperature data developed by Jones *et al.* (2010) from the NASA Advanced Microwave Scanning Radiometer on the Earth Observing System (AMSR-E) were retrieved from NASA (2013) at a 25 km×25 km resolution. We preferred the AMSR-E dataset over other satellite-derived products because it estimates global surface air temperature during day or night, even under cloudy, non-precipitating and non-frozen conditions (Chuang *et al.*, 2012). AMSR-E data were only available for the period June 2002 to September 2011. Mean daily precipitation data were obtained from the Famine Early Warning Systems Network RFE version 2.0 (FEWS-RFE



V2.0) for the whole of Africa (IRI, 2012), at a 0.1 degree resolution. Rainfall data were obtained for the period January 2002 to December 2011. District-specific values for both meteorological parameters were estimated using standard procedures included in the R *raster* package (Hijmans, 2013). Due to the missing air temperature values for the months before June 2002 and after September 2011, all datasets were subset to the same time span of the AMSR-E air temperature data (*i.e.* June 2002 to a maximum of September 2011).

Demographic data

Gridded population estimates at a 2.5 arc min resolution were obtained for the whole world from the Gridded Population of the World project (SEDAC, 2012) at five-year intervals. Monthly population estimates for each grid box were derived using linear interpolation. Population estimates for each district were then computed using standard methods included in the R *raster* package (Hijmans, 2013).

Socioeconomic data

In Rwanda, the average number of annual female outpatient visits to health facilities, female literacy rate, percentage of women aged 15-49 who reported serious problems in accessing health care, percentage of children aged 6-59 months with any kind of anaemia, and the percentage of households with at least one mosquito net (treated or untreated) were retrieved from the National Institute of Statistics of Rwanda (NISR, MOH and ICF International, 2012). These data comprised one observation per district corresponding to the mean value for year 2010. Socioeconomic data for Uganda included the percentage of the population living in urban areas by 2002, poverty rate (%), and the average rural population density per km² (both of them for year 2005). Data were retrieved from the World Resources Institute (WRI, 2012). Urban areas in Uganda were defined using the 2002 National Population and Housing Census (UBOS, 2002) divided into categories of gazetted cities, municipalities, and town councils. Data were aggregated at the district level using standard methods within the R *raster* package as before (Hijmans, 2013).

Best subset selection of socioeconomic predictors

We aimed to investigate relationships between malaria, air temperature and rainfall. Consequently, both air temperature and rainfall were present in all our model specifications. The best subset of socioeconomic predictors, on the other hand, was determined fitting a series of generalised additive mixed models (GAMM) (Wood, 2006). A time series cross-validation (TSCV) algorithm was implemented following James *et al.* (2013) and Hyndman and Athanasopoulos (2014) to select the set of socioeconomic predictors raising the lowest test error. TSCV was used rather than k-fold or leave-one-out cross-validation methods because these last two methods assume that the data are independent and identically distributed, an assumption violated by epidemiological surveillance time series which are typically serially correlated (Bergmeir and Benítez, 2012; Opsomer *et al.*, 2001).

We fitted models using all socioeconomic predictors in isolation, as well as all possible combinations. Thus, we progressively fitted all possible models containing only one socioeconomic predictor, then two of these predictors, and so on, with all predictors included in the final model. The mean absolute error (MAE) was selected as the measure for test error. We preferred MAE over the root mean square error (RMSE), because MAE is a natural and unambiguous measure of average error magnitude (Willmott and Matsuura, 2005).

TSCV was implemented dividing the dataset into a training and a test set. Models were fitted on a training dataset comprising a fixed

number of observations (m) equivalent to 60% of the total observations. Thus, at time $t=1$, the training set comprised observations for times $t=1, \dots, m$; at $t=2$ it comprised observations for $t=2, \dots, m+1$, and so on. The test set comprised the first observations per district immediately after that at the end of the training set. Thus, when the training set comprised observations at times $t=1, \dots, m$, the test set contained the observation $t=m+1$; when the training set included observations at times $t=2, \dots, m+1$, then the test set contained the observation $t=m+2$, and so on until the test set contained the observation at time $t=n$; where n is the total number of observations in the dataset. TSCV was implemented for all socioeconomic predictors. Error estimates were computed for every time step $t=1, \dots, T$, and for each subset of predictors $h=1, \dots, H$ as in the following matrix:

$$MAE_{t,h} = \begin{bmatrix} MAE_{1,1} & MAE_{1,2} & MAE_{1,3} & \dots & MAE_{1,H} \\ MAE_{2,1} & MAE_{2,2} & MAE_{2,3} & \dots & MAE_{2,H} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ MAE_{T,1} & MAE_{T,2} & MAE_{T,3} & \dots & MAE_{T,H} \end{bmatrix}$$

The error estimate for each subset (henceforth $MAE_{t,h}$) was computed averaging the subset-specific values (the columns in the matrix above) for all time steps. We aimed to identify the subset of socioeconomic predictors raising the lowest $MAE_{t,h}$ error.

A general description of the modelling approach:

The expected number of malaria cases $E(Y_{it})$ for district $i=1, \dots, I$ (where $I=30$ in Rwanda, and $I=55$ in Uganda), at time $t=1, \dots, T$ was assumed to follow an overdispersed Poisson distribution. To account for such overdispersion, we fitted quasi-maximum likelihood Poisson models allowing the scale parameter to be different from the mean (Cameron and Trivedi, 1990). The Poisson quasi-maximum likelihood estimator is the standard consistent estimators for count variables when we cannot assume that the Poisson distribution is true (Wooldridge, 2002). This quasi-maximum likelihood estimator is robust to distributional misspecification, and is efficient even when data distribution is not Poisson (Wooldridge, 2002). Some elements were constant to all models to account for specific aspects of the disease dynamics.

First, air temperature and precipitation were included as the main predictors of changes in malaria incidence. Second, we included the logarithm of the population/month at risk as an offset to adjust the malaria data by population to estimate relations on the crude incidence rate (CIR) rather than on the total number of cases as in Colón-González *et al.* (2013). Third, whenever significant, we incorporated a smooth interaction function between the latitude and longitude coordinates of each district's centroid to model any spatial dependence in the data (Lowe *et al.*, 2013; Wood, 2006). Fourth, we used cubic spline functions for time to control for long-term trends and seasonality in malaria incidence following Bhaskaran *et al.* (2013). Long-term trends were controlled for because it is impossible to ascertain whether these trends are due to real changes in malaria incidence over time, or due to changes in reporting or coverage. Similarly, seasonality was accounted for because seasonal patterns in malaria incidence could be related to factors other than weather such as irrigation practices (Baeza *et al.*, 2011). Finally, we incorporated district-specific random effects to account for the effects of unknown or unobserved variables (Johnston and DiNardo, 1997) in the model such as diagnostic performance variability, immunity, and intervention measures following Lowe *et al.* (2013).

Accounting for seasonality

Malaria shows distinct seasonal patterns that could vary between regions as a response to climate forcing (Dery *et al.*, 2010; Haque *et al.*, 2010). Meteorological factors are characterised by being spatially coherent at large scales (Conrad, 1946), and so climate-related seasonal changes in malaria incidence are also expected to be coherent across spatial scales larger than the typical administrative unit at which health data are recorded. Non-climatic factors such as irrigation practices (see for example Baeza *et al.*, 2011) could also modulate seasonal changes in malaria incidence; however, these changes are likely to occur at a finer spatial scale.

The potential differences in the seasonal patterns of malaria incidence across space pose the question as to whether seasonality should be modelled using one spline function of time for the whole country, or it would be better to use a spline function for each of the regions where malaria incidence shows a similar seasonal pattern. To test this hypothesis, we fitted models with a single spline function of time for the whole country, and then models with a spline function of time for each distinct region, and compared their MAE_{t,h} errors. The model with the lowest MAE_{t,h} error was selected. To define regions with similar seasonal malaria dynamics, we used the output of the fully dynamical Vector-borne disease community model of the International Centre for Theoretical Physics, Trieste (VECTRI) malaria model, version 1.3 (Tompkins and Ermert, 2013). A full description of the VECTRI model can be found elsewhere (Tompkins and Ermert, 2013). Briefly, VECTRI is a high-resolution dynamical malaria model that explicitly accounts for the effects of temperature on the key cycles of malaria transmission, and the effects of rainfall on seasonal temporary water bodies in a pond model in addition to the effects of population density on the calculation of daily biting rates. The VECTRI model neglects the effects of non-climatic influential factors such as host immunity, migration and interventions. We preferred using VECTRI-derived outputs over observed malaria reports because, by being essentially climate-driven, the VECTRI outputs are spatially homogeneous. Epidemiological surveillance data, on the other hand, are rarely if ever spatially homogeneous due to the influence of local non-climatic factors.

Daily entomological inoculation rates or EIR (number of infectious mosquito bites per person per unit time) were estimated at a 10 km×10 km for the period 1 January 2001 to 31 December 2012, and then averaged at the monthly scale. The logarithm of the EIR (log-EIR) was then computed and used as our malaria metric because previous research indicates that its behaviour is correlated to that of malaria incidence (Beier *et al.*, 1999; Smith *et al.*, 1998). Monthly log-EIR estimates were retrieved for both countries using the R *raster* package (Hijmans, 2013) and aggregated at the district level.

Districts were allocated to a climate-influenced region using Ward's Minimum Variance Clustering (WMVC) on the mean annual cycle of the log-EIR. The mean annual cycle consisted of district-specific mean log-EIR estimates for each month of the year. WMVC has been extensively described elsewhere (Borcard *et al.*, 2011). Briefly, WMVC is a clustering method based on the linear model sum of squares criterion to produce groups that minimise the within-group sum of squares (Borcard *et al.*, 2011).

District-specific log-EIR annual cycles were normalised (with zero mean and unit variance) to adjust for different scales between districts as in Shannon *et al.* (2003). Membership to a climatic region was binary such that a district could only belong to one region after the partition process (Legendre and Rogers, 1972). We selected WMVC because it is less susceptible to noise and outliers than other methods such as k-means (Tan *et al.*, 2005). WMVC was performed using the R *statistics* package.

Model specification

We specified the expected number of malaria cases in province i and time t as follows:

$$g(\mu_{it}) = \alpha + \log(\xi_{it}) + h(\text{Lon}_i, \text{Lat}_i) + t' + d_i + \sum_{q=1}^Q \beta(u_{it}) + \sum_{p=1}^P f(x_{it,p})$$

where $g(\mu_{it})$ is a log link function of the expectation $E(Y_{it} \equiv \mu_{it})$, α denotes the intercept; $\log(\xi_{it})$ is the logarithm of the population at risk for district i and time t included as an offset to adjust the malaria data by population; $h(\text{Lat}_i, \text{Lon}_i)$ is a smooth function of the interaction between the district-specific latitude and longitude centroid coordinates; t' is a cubic regression spline function of time t to control for long-term trends and seasonality over the whole data period. The spline function of time had $k=1, \dots, K$ degrees of freedom (number of calendar years $\times S$ seasons) to control for seasonality in the data. The number of seasons per year S was selected fitting the model with two or four degrees of freedom per year. Whenever it helped to reduce the MAE_{t,h} error compared to a single spline for the whole country, a spline function of time was fitted for each climate-influenced region $j=1, \dots, J$ to account for region-specific long-term trends and seasonality. The term d_i denotes district-specific random effects; u_{it} are socioeconomic variables with linear effects specified by the related coefficients β ; $f(\cdot)$ denotes smoothed relationships between the meteorological parameters x_{it} and malaria incidence at lags $l=1, \dots, L$.

The relationship between $g(\mu_{it})$ and $f(x_{it})$ is defined with a series of cross-basis functions describing simultaneously the nonlinear form of the relationship along x_{it} and its distributed lag effects (Gasparrini *et al.*, 2010; Gasparrini, 2011). Cross-basis functions were implemented using a so-called distributed-lag nonlinear model method (Gasparrini *et al.*, 2010; Gasparrini, 2011), which effectively avoids the collinearity issues resulting from the inclusion of a parameter for each lag of the exposure variable. Detailed algebraic definitions of the cross-basis functions can be found elsewhere (Gasparrini *et al.*, 2010; Gasparrini, 2011). The exposure dimension of the cross-basis functions (exposure-response relationships) for both temperature and precipitation was defined with natural cubic splines with two degrees of freedom. The lag dimension (lag-response relationship) was defined with linear functions. We considered lags from zero to two months to account for biologically and physically plausible lagged effects. Cross-basis functions were defined using the R *dlnm* package (Gasparrini, 2011).

Selection of the degree of smoothing for the exposure-response dimension

After we defined the best subset of socioeconomic predictors, we defined the most adequate specification for the degree of smoothing of the exposure dimension of the $f(x_{it})$ cross-basis functions (see equation 1). Specifically, we modified the number of degrees of freedom (ranging from two to six) of the natural cubic splines to model the nonlinear and delayed effects of the weather predictors (Brit *et al.*, 2008; Haque *et al.*, 2010; Lowe *et al.*, 2013). The lag-response dimension was defined with linear terms. Models were specified as in equation 1. All possible model specifications were tested using the TSCV algorithm described. We aimed to identify the model specification raising the lowest MAE_{t,h} error estimate.

Contribution of the weather predictors

We evaluated the contribution of air temperature and rainfall to our preferred models' predictive power by comparing their ability to reduce



the model's MAE_{ti} error. We fitted these models incorporating all weather and socioeconomic predictors in isolation, as well as all possible combinations. Thus, as in the best subset selection step, we fitted all possible models containing one predictor, then two predictors, and so on until all predictors were included back into model. The TSCV algorithm was used to compute the error estimates for each model specification. Error estimates for all model specifications were then compared against a null model with no weather or socioeconomic predictor.

Results and Discussion

Descriptive statistics at the national level are reported in Table 1. After the imputation of missing values, a total of 5,104,341 malaria reports were estimated in Rwanda over the period January 2001 to December 2011. In Uganda a total of 101,636,920 malaria reports were estimated to have occurred over the period January 2002 to December 2010. None of the months reported zero malaria cases at the district level. It is recalled that in this study, malaria cases refer to clinically diagnosed (suspected) malaria occurrences.

The average monthly CIR in Rwanda was relatively low with only about five cases occurring for every 1000 people partly due to the high mean altitude of the country (~1800 m above mean sea level). The CIR in Uganda was considerably greater with about 39 cases per 1000 people occurring each month. Near-surface air temperature was suitable for malaria transmission in all districts and months in both countries (range=18–33°C in Rwanda, and 20–41°C in Uganda), meanwhile precipitation seems to be an important limiting factor with some months experiencing low or zero rainfall (range=0–13 mm day⁻¹ in Rwanda, and 0–12

mm day⁻¹ in Uganda). Table 1 also indicates that there are moderate to large disparities in the socioeconomic conditions of the districts in both countries but does not provide information about their spatial distribution. In Rwanda, the largest disparities were related to the problems experienced by female patients for accessing health care, and to the possession of mosquito nets for preventing malaria transmission. In Uganda, greatest variability was observed in urbanisation where percentages ranged between one and 100%. However, it must be pointed out that most of the districts show urbanisation levels between one and twelve percent, four districts show levels between 15 and 25%, and only Kampala was reported with an urbanisation level of 100%.

Figure 1 shows the spatial patterns observed in malaria incidence. Greater incidence rates were observed in low-lying districts in Rwanda compared to high-elevation areas. Previous reports indicate a greater rate of change in the number of malaria cases in the highlands compared against mid or low altitudes in Rwanda (Loevinsohn, 1994). However, the absolute number of malaria cases is highest in the eastern, southern, and south-western parts of the country, characterised for mid to low altitudes as reported by Henninger (2013); USAID (2013a) and Rulisa *et al.* (2013). It is noted that these patterns are highly similar to those of air temperature and rainfall, with greater mean temperature and lower rainfall in those districts where malaria incidence is greater. Both temperature and rainfall largely determine the disease niche through several biological mechanisms. Warming temperatures, for example, shorten the sporogonic cycle of the parasite within the vector (Craig *et al.*, 1999) increasing the likelihood of transmission. Also, the duration of the mosquito development from egg to adult, the length of the gonotrophic cycle length and larvae mortality in the immature stages are steered by temperature to complicate this picture (Paaijmans *et al.*, 2008a, 2008b). Rainfall, on the other hand, may increase the number of breeding sites for the vector (Gage *et al.*, 2008). It is recalled that malaria transmission

Table 1. Descriptive statistics for the weather, demographic, and socioeconomic variables considered in the study. The number of district-specific observations retrieved for each variable is indicated by *n*. The number of observations for all socioeconomic variables equals the number of districts in each country because data could only be retrieved aggregated annually, and for one particular year.

	<i>n</i>	Mean	Median Rwanda	SD	Range
Cases	3960	1289	1000	1435.4	4–24479
Crude incidence rate	3960	4.9	3.7	5.6	0–108.6
Air temperature (°C)	3360	26.0	26.7	3.5	18.4–33.0
Precipitation (mm.day ⁻¹)	3960	2.4	2.3	1.8	0–12.7
Female outpatients (visits.year ⁻¹)	30	1.8	1.6	0.7	0.9–3.9
Child anaemia (%)	30	37.6	35.6	8.1	23–52.7
Female literacy (%)	30	76.9	77.4	7.8	61.7–92.0
Mosquito net possession (%)	30	82.4	87.4	10.9	59.9–94.8
Problems to access health care (%)	30	61.6	61.0	12.9	33.8–87.9
Uganda					
Cases	5940	17079	14448	10819	24–148162
Crude incidence rate	5940	38.8	34.1	22.7	0.1–357.8
Air temperature (°C)	5940	29.2	29.5	3.2	19.8–41.3
Precipitation (mm.day ⁻¹)	5940	3.0	2.8	1.6	0–11.9
Poverty rate (%)	55	32.4	28.3	18.5	0–77.9
Population density (people/km ²)	55	18.2	14.7	14.6	0–61.1
Urbanisation (%)	55	8.9	6.2	13.3	1.1–100

SD, standard deviation.

shows a highly non-linear relationship with these two drivers (Craig *et al.*, 1999; Lunde *et al.*, 2013; Tompkins and Ermert, 2013), and so both high temperatures and heavy rainfall may increase vector mortality resulting in reduced transmission (Thomson *et al.*, 2005).

The observed spatial heterogeneity in malaria incidence agrees well with previous research in the country (Karema *et al.*, 2012; Rulisa *et al.*, 2013), and is likely due to the interplay of several factors including the local environment, bed net use, and host genetic polymorphism (Clark *et al.*, 2008). It is noted that child anaemia shows a somewhat similar geographical distribution to that of malaria incidence with a greater proportion of anaemic children in the eastern and south-western parts of the country. However, the patterns on the rest of the country are not that similar at a visual inspection, and this is evidenced in a low correlation coefficient between these two variables (Spearman $\rho=0.3$). Some correlation between these two variables is to be expected, however, because malaria is frequently associated with anaemia (Douglas *et al.*, 2012; Menendez *et al.*, 2000). The other socioeconomic indicators showed patterns that differ from that of malaria incidence. It is difficult to ascertain whether these patterns reflect the situation for the overall study period because whilst we have malaria data for over a decade, socioeconomic data could only be obtained for a single year.

Previous studies conducted in Uganda suggest a homogeneous spatial pattern with most of the country showing high to very high transmission levels, and a few, well defined epidemic prone areas located in the highlands (UBOS and ICF International, 2009; USAID, 2013b). On the contrary, our data show large spatial heterogeneity in the distribution of malaria cases per 1000 people (Figure 2) with no clear spatial patterns. Factors such as human behaviour (*e.g.* irrigation activities), immunity levels, and access to preventive measures such as sanitation and window screening (Baeza *et al.*, 2011; Pindolia *et al.*, 2013; Reiter, 2001, 2008), may all contribute to such heterogeneity in addition to differences in the criteria, and skills to use diagnostic tools (Kyabayinze *et al.*, 2012; Yeka *et al.*, 2012). It is noted, however, that malaria incidence tends to be greater in the northern and eastern half of the country where the mean temperature, poverty rate, and rural population densities are greater, and the mean altitude is lower in agreement with previous reports (UBOS and ICF International, 2009; USAID, 2013b). The levels of urbanisation do not show a clear spatial pattern. However, some of the districts with the greater levels of urbanisation in central Uganda show some of the lowest levels of malaria incidence as observed in other African countries such as Ghana (Klinkenberg *et al.*, 2008).

Climate-influenced regions

Figure 3 illustrates the location of the climate-influenced regions obtained through WMVC. The WMVC method applied to the VECTRI-estimated log-EIR produced a series of compact clusters of districts of variable size that were coherent with the spatial patterns observed in malaria incidence. We emphasise that the clustering algorithm identifies differences in malaria seasonality, not in the absolute number of malaria cases. In Rwanda, districts are assigned to four regions, one of which covers the east and south of the country where incidence is greatest. In Uganda, districts were allocated to six regions. It is noted that the south-western highlands, where malaria incidence has been continuously reported as being lower (UBOS and ICF International, 2009; USAID, 2013b), is assigned to an independent region. In all regions of both countries the annual season cycle is bimodal in response to the two rainy seasons at these latitudes, with maximum transmission occurring in January and then May-June. Malaria transmission is at its lowest in August in Rwanda, and September in Uganda (Figure 4). It is notable that the highest variability tends to occur in the months directly preceding the peak transmission. For example, in

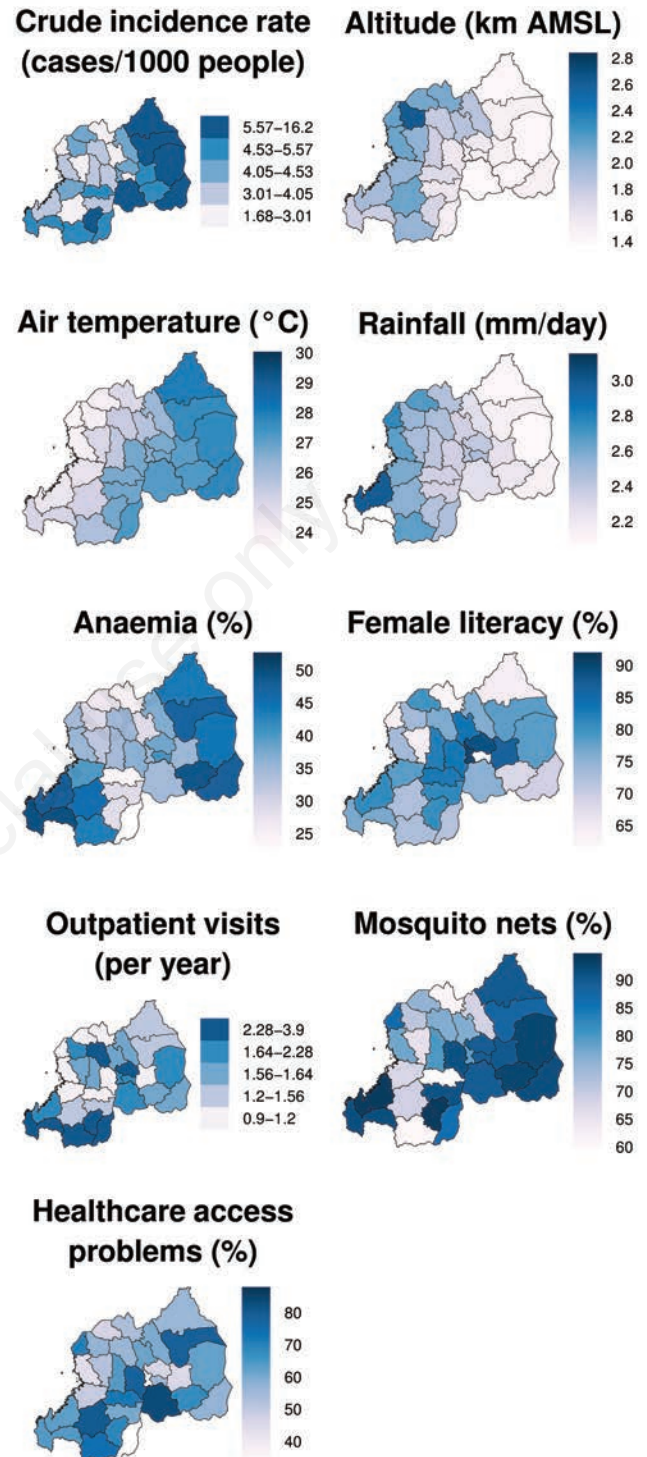


Figure 1. Mean monthly crude incidence rate per 1000 people (quantiles), mean altitude in km above mean sea level, mean air surface temperature ($^{\circ}\text{C}$), mean daily precipitation (mm day^{-1}), percentage of children aged 6-59 months with any kind of anaemia, female literacy rate (%), average number of female outpatient visits to health facilities *per annum* (quantiles), percentage of households with at least one mosquito net (treated or untreated), and percentage of women aged 15-49 reporting healthcare access problems across Rwanda.



regions 3 and 4 in both Rwanda and Uganda, the greatest year-to-year variability occurs in December (represented as a very large range between the upper and lower whiskers of the December the box-plot) just before the transmission peak in January during which variability is lower (Figures 4 and 5). This behaviour is likely to be produced by the variability in the onset of rainfall. In other words, transmission is always underway by January in this zone, while transmission in December only occurs in years in which the rains begin unusually early

in October and early November. Thus, even in areas where malaria is mesoendemic or hyperendemic, malaria transmission can still be highly variable in certain months of the year, implying potential usefulness of malaria early warning systems (Tompkins and Di Giuseppe, 2015).

A conspicuous downward trend in malaria incidence is observed over the period 2006-2010 in some areas of Rwanda. Figure 6 reveals that this decrease was not homogeneous across the country (notice that the figure uses a 12-month rolling mean). In regions 2 and 4 (Figure 6) characterised by high altitudes, malaria incidence decreased in 2006 and showed a light increase with a peak in late 2010. In regions 1 and 3 (Figure 6) on the other hand, the significant decrease in incidence observed in 2006 was followed by a significant rise in incidence with a peak in late 2010, and a further decrease thereafter. Previous research suggested that these decreases in malaria incidence were related to a scale-up of intervention measures implemented by the government over the same period (Karema *et al.*, 2012; Rulisa *et al.*, 2013). Also, the conspicuous increase observed in malaria incidence has been attributed to a temporary decrease in the distribution of insecticide-treated mosquito nets (Karema *et al.*, 2012). Whilst these hypotheses may be correct, we found that malaria incidence over the 2006-2010 period, particularly in Region 3, starts rising at the same time as air temperature (Figure 6).

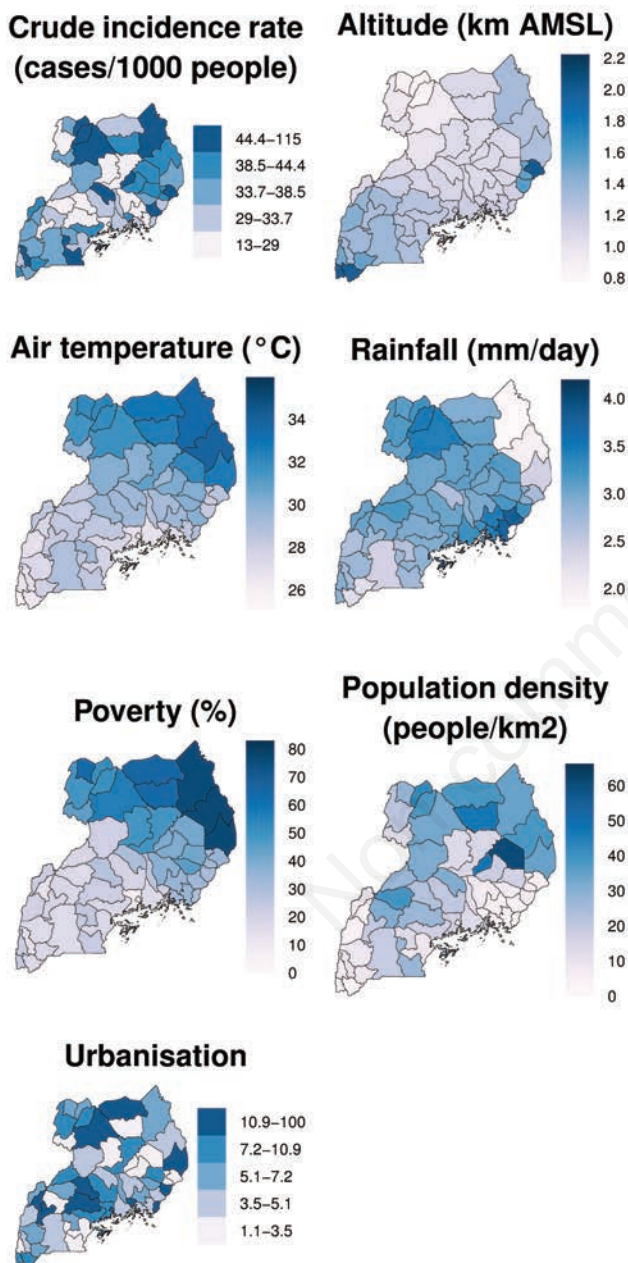


Figure 2. Mean monthly crude incidence rate per 1000 people (quantiles), mean altitude in km above mean sea level, mean air surface temperature (°C), mean daily precipitation (mm day⁻¹), poverty rate (%), rural population density (people/km²), and percentage of the population living in urbanised areas (quantiles) across Uganda.

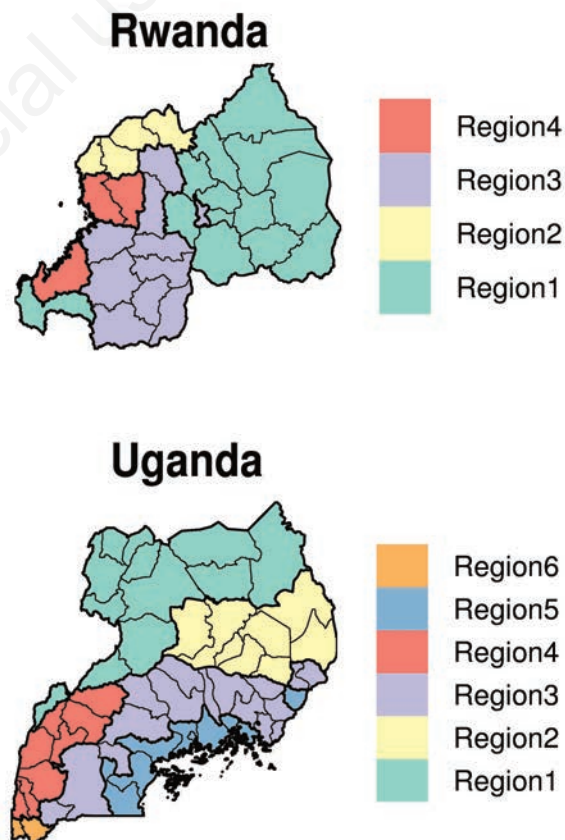


Figure 3. Spatial location of climate-influenced regions defined using the spatiotemporal dynamics of the Vector-borne disease community model of the International Centre for Theoretical Physics, Trieste-estimated logarithm of the entomological inoculation rate in Rwanda and Uganda. Clusters were computed using Ward's minimum variance clustering.

This observation poses doubts as to whether changes in the intervention measures are the only cause of the observed variations in malaria incidence, and highlights the need for further investigation. In Uganda, most regions (*i.e.* regions 1, 2, 3, and 5) show a significant rising trend in malaria over the study period (Figure 7). These rising trends agree well with a recent study conducted in Uganda indicating that malaria in Uganda has risen and continues to rise in some regions (Kamya *et al.*, 2015). As in Rwanda, variations in malaria transmission in Uganda are in agreement with changes in air temperature and rainfall. This agreement is more evident in Regions 1, 4 and 5 where the long-term trends in malaria transmission closely follow the trends in air temperature. It is noted that in Region 6 there is a disagreement between the malaria, temperature and rainfall trends towards the end of the period (2010). This disagreement could be due to intervention measures to control the disease in the area. This hypothesis could not be assessed due to the lack of data on interventions.

Best subset

After having tested 62 model specifications based on the maximum number of predictor combinations, the TSCV algorithm applied to the

malaria data in Rwanda favoured a GAMM including air temperature, rainfall, average number of annual female outpatient visits to health facilities, the percentage of children 6-59 months with any kind of anaemia, and the percentage of households with at least one mosquito net in addition to the fixed elements of the model (*i.e.* spatial and temporal trends, random effects and a population offset). The MAE_{t,h} was 3.24 cases per 1000 people.

For simplicity, we selected a more parsimonious model including only air temperature, rainfall, and the annual female outpatient visits that produced the same MAE_{t,h} than the previous model (3.24 cases per 1000 people). The lowest MAE_{t,h} was achieved using a spline function of time with two degrees of freedom per year for each climatic region to account for potential long-term trends and seasonality in the data. Table 2 reports the variables included in the selected model. It is noted that while the reporting and interpretation of the non-climatic linear effects is straightforward, the results of the smoothed non-linear dependencies cannot be summarised using a set of regression parameters (Beck and Jackman, 1998; Gasparrini *et al.*, 2010). Graphical methods are used here to interpret the estimated nonlinearities (see sub-section *Resulting relationships*).

In Uganda, a total of 14 model specifications were tested. TSCV

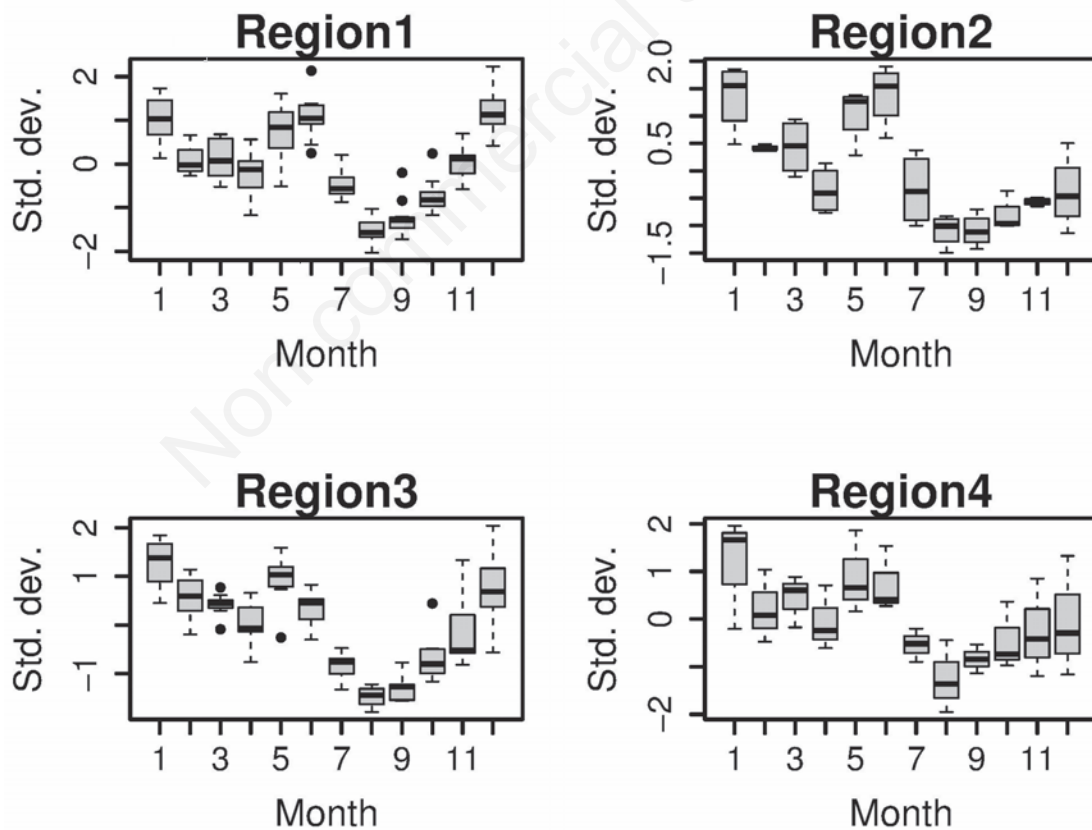


Figure 4. Box-plots of the mean seasonal cycle of malaria incidence in four regions across Rwanda. The dispersion between the different parts of the box-plot helps to show the interannual variability. The thick black horizontal lines inside the boxes indicate the median value for each month. The top ends of the boxes indicate the upper (75%) quartiles. The lower ends of the boxes mark the lower (25%) quartiles. The upper and lower whiskers represent the maximum and minimum values for each month (excluding outliers which are indicated with black dots). Outliers are values beyond ± 1.5 times the interquartile range.

favoured a GAMM including air temperature, rainfall, and urbanisation (Table 2). The lowest $MAE_{i,h}$ (12.5 cases per 1000 people) was obtained when using a single spline function of time with four degrees of freedom per year for the whole country to control for long-term trends and seasonality.

Degree of smoothing of the exposure-response relationships

We tested nine different model specifications with different degrees of smoothing in the relationship between malaria and each of the weather predictors. In Rwanda, TSCV favoured a natural cubic spline with six degrees of freedom ($MAE_{i,h}=3.17$ cases per 1000 people) to represent the relations between malaria incidence and air temperature

lagged zero to two months ($temp_{0,2}$). The results for rainfall lagged zero to two months ($rain_{0,2}$) were represented with a natural cubic spline with two degrees of freedom. Considering the plausibility of the estimated exposure-response relationships, and the compromise between complexity, generalisability and ease of interpretation (Gasparrini, 2011), we selected a model where both $temp_{0,2}$ and $rain_{0,2}$ were represented by natural cubic splines with two degrees of freedom. The simpler model was easier to interpret because did not require to clarify the whole complexity of nonlinearities estimated when using a more complex spline function. Similar approaches have been proposed elsewhere (see for example Armstrong, 2006).

The selected model explained approximately 65% of the deviance in the malaria data. The $MAE_{i,h}$ of the simpler model (3.24 cases per 1000 people) did not differ greatly from that of the more complex model. In

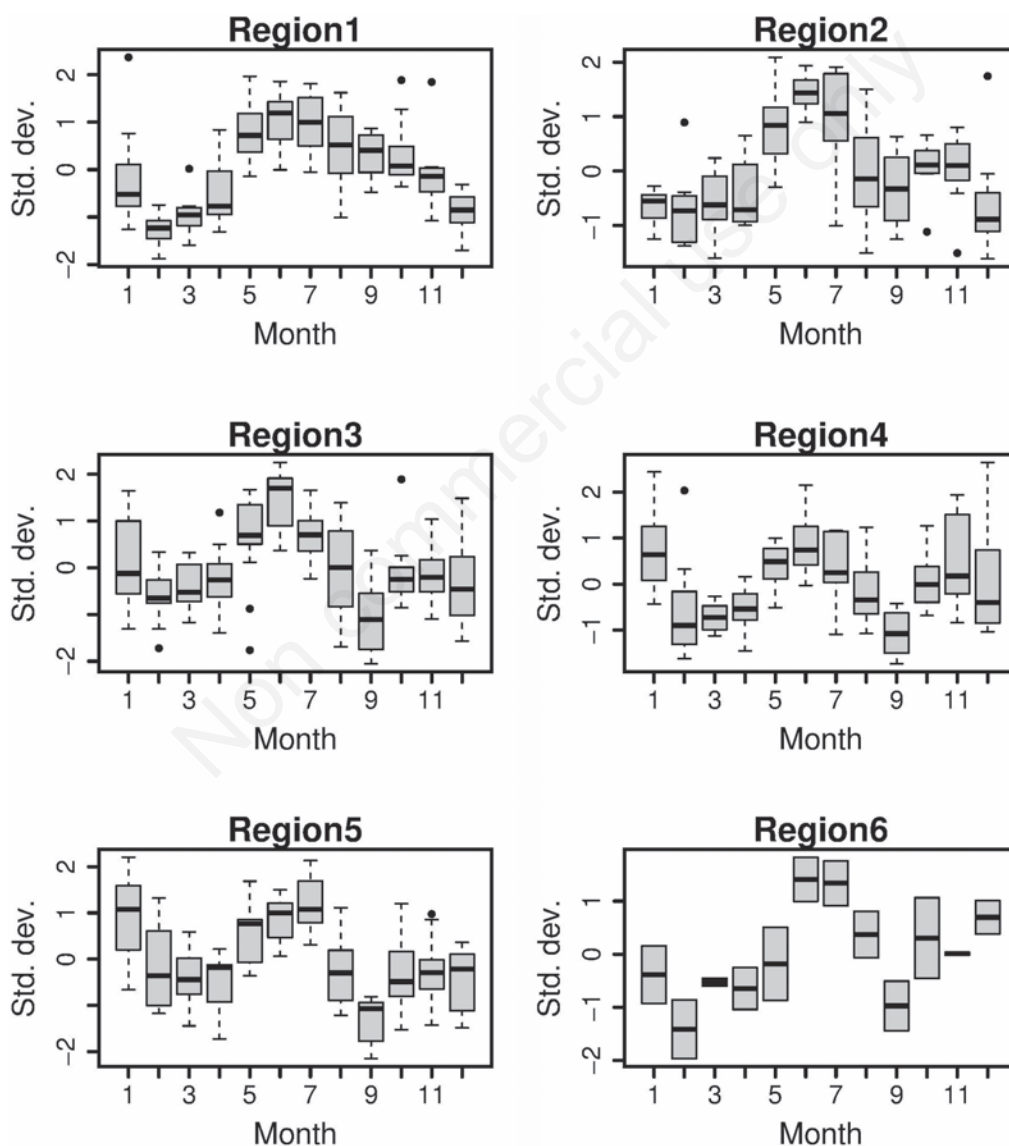


Figure 5. Box-plots of the mean seasonal cycle of malaria incidence in six regions across Uganda. The dispersion between the different parts of the box-plot helps to show the interannual variability. The thick black horizontal lines inside the boxes indicate the median value for each month. The top ends of the boxes indicate the upper (75%) quartiles. The lower ends of the boxes mark the lower (25%) quartiles. The upper and lower whiskers represent the maximum and minimum values for each month (excluding outliers which are indicated with black dots). Outliers are values beyond ± 1.5 times the interquartile range.

Uganda, TSCV favoured a cubic spline with two degrees of freedom to represent the relationships between malaria incidence, $temp_{0.2}$ and $rain_{0.2}$.

The MAE_{t,h} raised by this model was 12.5 cases per 1000 people. The model fit on the Uganda dataset explained 67% of the deviance in the malaria dataset. We emphasise that these spline functions were applied only to the exposure-response dimension whilst the lag-response dimension was defined as a linear function.

Resulting relationships

Figure 8 depicts the cumulative response functions for $temp_{0.2}$, $rain_{0.2}$, the average annual female outpatient visits and urbanisation levels computed based on the selected models. The relationships between malaria incidence and the meteorological predictors are interpreted as their accumulated effect on malaria incidence over the current month, and up to a lag of two months, inclusive, under the hypothetical assumption that the predictor remains constant over that period. The relative change in malaria incidence per 1000 people or relative risk, for both variables, is compared against the mean values of both

Table 2. Variables included in the selected Poisson generalised additive mixed models used to assess the effects of weather and non-climatic predictors on malaria incidence in Rwanda and Uganda.

Country	Predictor	Type
Rwanda	$Temp_{0.2}$	Distributed lag smooth term
	$Rain_{0.2}$	Distributed lag smooth term
	Female outpatient visits	Linear term
	Latitude and longitude	Smooth term
	Time	Smooth term per climate-influenced region
	Log-population	Offset
	Unobserved variables	Random effects
Uganda	Predictor	Type
	$Temp_{0.2}$	Distributed lag smooth term
	$Rain_{0.2}$	Distributed lag smooth term
	Urbanisation	Linear term
	Latitude and longitude	Smooth term
	Time	Smooth term
	Log-population	Offset
Unobserved variables	Random effects	

$temp_{0.2}$, air temperature lagged zero to two months; $rain_{0.2}$, results for rainfall lagged zero to two months.

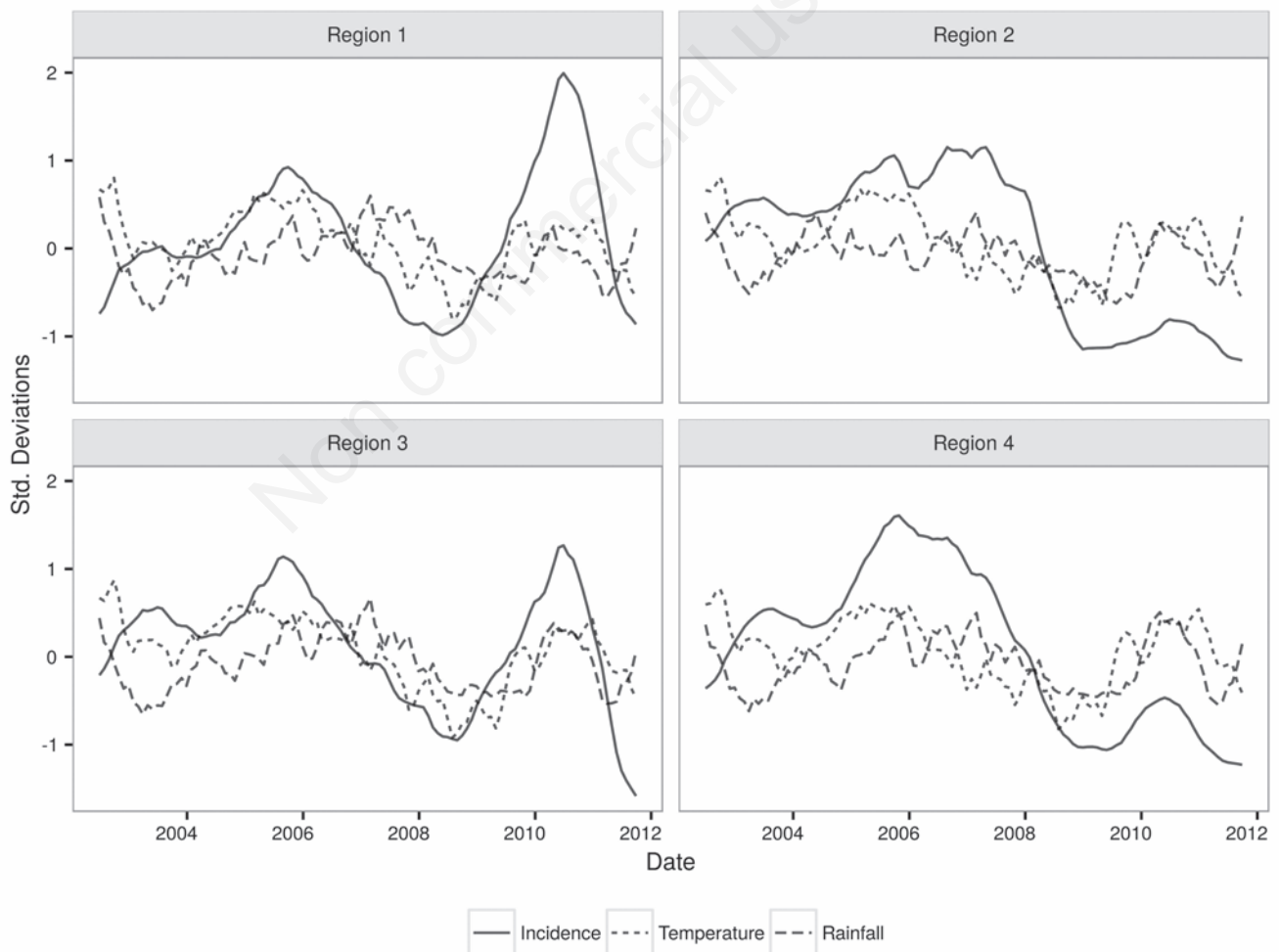


Figure 6. Time series (12 month-rolling mean) of malaria incidence (solid lines), air temperature (dotted lines), and precipitation (dashed lines) in four regions in Rwanda. The y axis indicates changes in all three variables measured in standard deviations.



$temp_{0.2}$ and $rain_{0.2}$. Changes in the relative risk of malaria incidence as a function of the number of female outpatient visits per year and urbanisation levels, on the other hand, are compared exclusively against their value in year 2010 and 2002 respectively.

In Rwanda, our preferred model estimates an almost linear positive relationship between malaria incidence and $temp_{0.2}$ with temperatures below the mean producing a lower relative risk than temperatures above the mean. This relationship indicates that as $temp_{0.2}$ rises, so does malaria incidence. The estimated relation may be the result of rising temperatures increasing the development rate of the vector, and shortening the length of the sporogonic and gonotrophic cycles resulting in increased transmission (Craig *et al.*, 1999). Our mean estimated relation between malaria incidence and $temp_{0.2}$ does not agree with the estimated inverted U shaped relations proposed by Craig *et al.* (1999) and Lunde *et al.* (2013) showing a peak in transmission at about 25–30°C. However, the lower bound of the 95% confidence interval does

suggest a decrease in the relative risk of malaria at high temperatures with a peak at approximately 27°C, in agreement with these studies. In Uganda, on the other hand, the estimated effects of temperature on malaria incidence show the inverted U shape suggested by Craig *et al.* (1999) and Lunde *et al.* (2013) with a peak at about 27°C. This effect suggests that malaria transmission in Uganda is constrained by relatively low or high temperatures and agrees with the known ecology of the disease. Our estimated relationships between malaria and temperature in Rwanda, however, are in strong agreement with those recently estimated by Lowe *et al.* (2013) in Malawi, by Haque *et al.* (2010) in Bangladesh, and by Blanford *et al.* (2013) in some districts of Kenya, and suggest that variations in $temp_{0.2}$ may not exceed the threshold at which parasite development is negatively affected to restrict malaria transmission. It is recalled that although variations in the AMSR-E temperature data correlate well with temperature records from meteorological stations (Chuang *et al.*, 2012), the actual temperature may be cold-

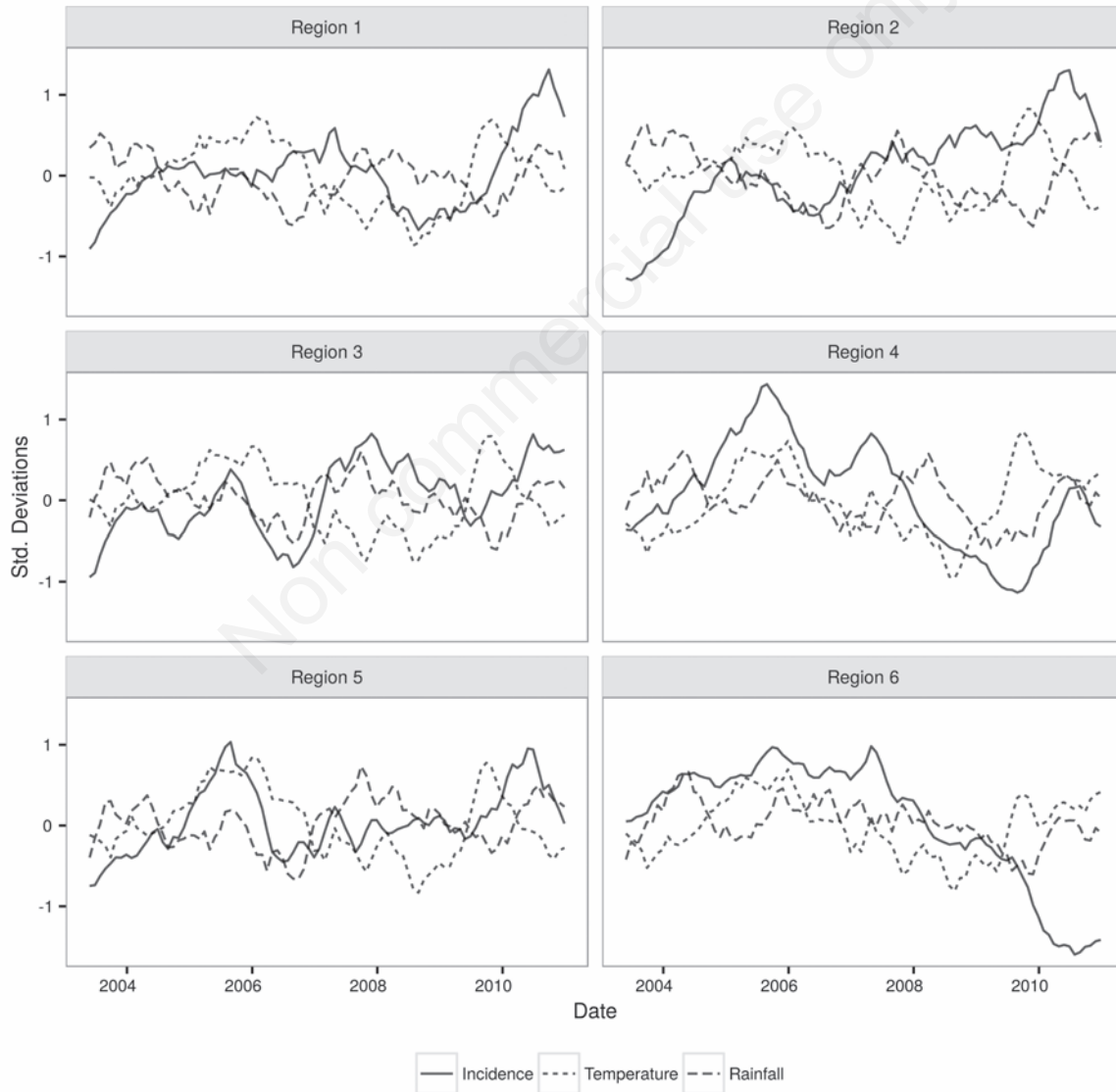


Figure 7. Time series (12 month-rolling mean) of malaria incidence (solid lines), air temperature (dotted lines), and precipitation (dashed lines) in six regions in Uganda. The y axis indicates changes in all three variables measured in standard deviations.

er or warmer than the satellite-derived data. Some anopheline species (e.g. *Anopheles gambiae* and *A. funestus*) spend considerable time indoors where they shelter from the high temperatures experienced outdoors (Faye *et al.*, 1997; Githeko *et al.*, 1996; Mnzava *et al.*, 1995). Other species may seek refuge under a variety of structures such as leaves, tree holes, and bridges, which may have their own micro-climate (Githeko *et al.*, 1996; Paaijmans and Thomas, 2011). Whether temperature measurements recorded by remote sensing devices are representative of the conditions experienced by mosquitoes in these micro-climates is uncertain (Paaijmans and Thomas, 2011). Consequently, even though our data suggest a temperature range between 18–33°C across the country for the study period, the temperatures experienced by the vector might be significantly colder, and malaria transmission is thus not restricted by warm temperatures.

The relative risk of malaria in both countries increases with rising $rain_{0.2}$ until a maximum is attained at a mean rainfall rate of 5–7

mm•day⁻¹, decreasing thereafter. This non-linear relationship between transmission and rainfall agrees with previous work in Botswana (Thomson *et al.*, 2005) and Malawi (Lowe *et al.*, 2013) and is likely to be partly a result of increasing rainfall contributing to the creation of rain-filled breeding sites, whereas the observed decay observed may be related to the flushing of larvae occurring in the presence of extreme rain events (Paaijmans *et al.*, 2007). Topography, soil type and land cover will also impact breeding site flushing.

The average number of female outpatient visits to health facilities per annum (2010) showed a positive log-linear relationship with malaria incidence, and so malaria incidence increases with rising average annual visits. It has been previously observed that about 50% of all outpatient visits in endemic countries are related to malaria (Guinovart *et al.*, 2008; WHO, 2011). Thus, this relationship is expected since increasing malaria case numbers are likely to result in an increasing number of outpatient visits. Likewise, a rising number of visits per year

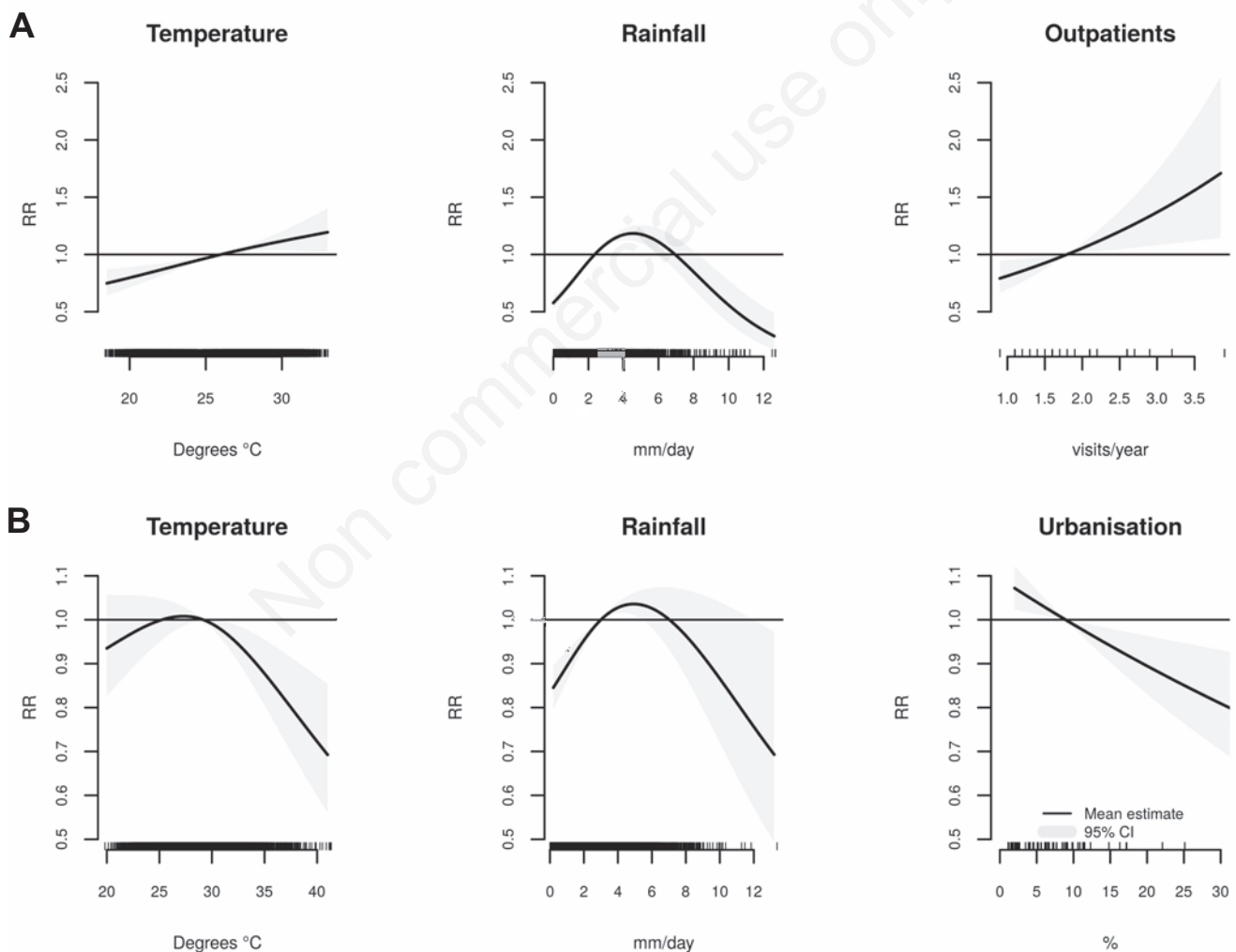


Figure 8. Generalised additive mixed model-estimated relationships between average monthly malaria incidence, monthly average near-surface air temperature and precipitation lagged 0 to 2 months from a distributed lag nonlinear model, average number of female outpatient visits per year, and urbanisation in Rwanda (A) and Uganda (B). The y axis [relative risk (RR)] represents the excess in malaria incidence with respect to the mean values of each covariate. The thick smooth lines indicate the maximum likelihood estimates, and the grey shaded areas represent the 95% confidence intervals. The rug at the bottom of each figure indicates observed values for each covariate.



increases the probability of observing malaria or suspected malaria in outpatients. Such positive relationship may also be a proxy for increasing access to health facilities which is likely to result in increased reporting as pointed out by Lowe *et al.* (2013).

We estimated a negative effect of urbanisation on malaria transmission with more urbanised districts showing a lower relative risk of malaria than rural districts. This effect agrees well with previous research indicating that urban areas show lower malaria rates than rural regions (for example Kelly-Hope and McKenzie, 2009), a situation possibly due to a dilution effect with large population densities reducing the likelihood of transmission. Appropriate treatment-seeking behaviour in health facilities is also more likely to occur urban than in rural areas due to geographically more accessible facilities in urban dwellings (Okeke and Okeibunor, 2010). Moreover, some anopheline species do not breed well in polluted water bodies associated with urban environments (Sinka *et al.*, 2010), which may lead to lower transmission levels due to a lower number of circulating adults in these

areas. Also, malaria transmission is largely influenced by housing type, housing materials, and access to preventive measures (*e.g.* screens, insecticide and bed nets), which are likely to be better in urban areas resulting in a lower risk (Robert *et al.*, 2003).

Model validation

The outputs of the models presented in Table 2 fitted on the whole time series effectively capture the spatiotemporal distribution of the observed crude malaria incidence rate in both countries. Figure 9 shows that the time series of GAMM-estimated malaria incidence captures quite closely the temporal variations observed in the surveillance data with some overestimations, and several underestimations over several periods in Rwanda (*e.g.* 2009-2011). The model estimates for Uganda, on the other hand, followed variations in the malaria data more closely with some minor underestimations over the period 2009-2010. It is noted that the estimated data effectively reflects the temporal behaviour in malaria incidence observed by Karema *et al.* (2012) in

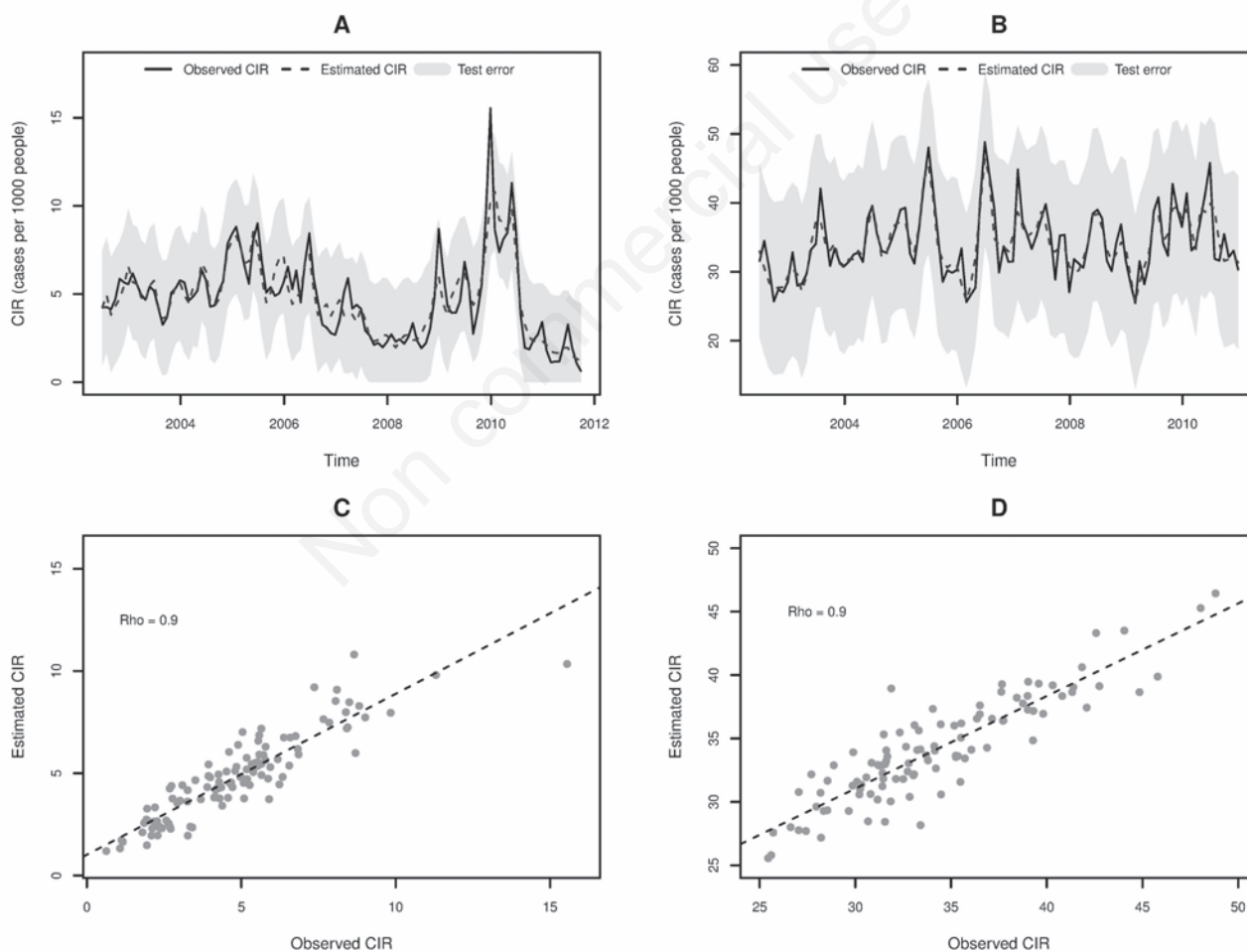


Figure 9. A, B) Time series of observed and generalised additive mixed model (GAMM)-estimated malaria crude incidence rate (CIR; cases per 1000 people) aggregated at the national level in Rwanda (A) and Uganda (B). The solid line represents the observed malaria incidence, and the dashed line corresponds to the GAMM-estimated mean malaria incidence. The shaded area represents the error estimates of the model. C, D) Scatter plot of observed vs GAMM-estimated malaria crude incidence rate aggregated at the national level in Rwanda (C) and Uganda (D). The dashed line shows the intercept/slope from a linear regression model fitted to both variables.

Rwanda with lower incidence over the period 2006-2008, a conspicuous increase in 2009, and a further decrease from 2010 onwards.

The Spearman rank correlation between observed and predicted malaria incidence was 0.9 in both countries supporting the notion that the outputs from our models closely represent the observed variations in malaria incidence. As observed in Figure 9, observed and estimated incidence rates for Rwanda are almost linear, with some outliers, the greatest being related to a peak in observed malaria incidence in 2010 that is largely underestimated by our model. In Uganda, the relationship is stronger than in Rwanda, and this is reflected in a small scatter of points around the regression line. As expected, due to the presence of district-specific random effects in the models, the estimated spatial

patterns of malaria incidence largely followed those of the observed incidence (Figures 1, 2, and Appendix Figure 1). In Rwanda, greater CIR were estimated in the low-lying districts located in the eastern and southern parts of the country, and lower values estimated in high-elevation districts, mainly located in western and central Rwanda, in agreement with Henninger (2013); Rulisa *et al.* (2013) and USAID (2013a). In Uganda, the greatest incidence rates were estimated in the northern part of the country characterised by lowlands as suggested by UBOS and ICF International (2009) and USAID (2013b). However, the lowest rates were estimated in the central region as observed in our malaria dataset, and not in the highland regions as previously reported (UBOS and ICF International, 2009; USAID, 2013b; Yeka *et al.*, 2012).

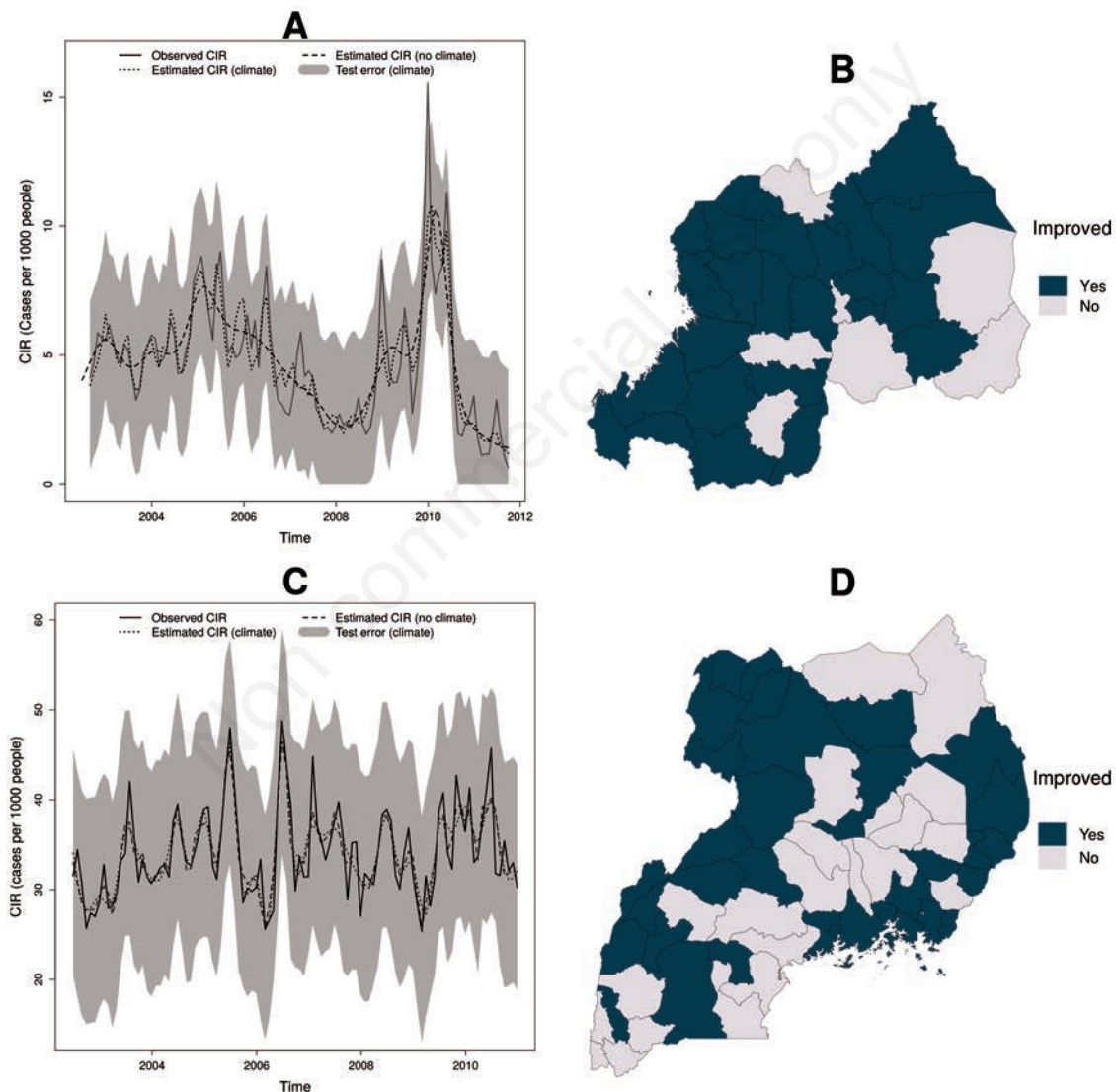


Figure 10. A, C) Time series of observed and generalised additive mixed model (GAMM)-estimated malaria crude incidence rate (CIR; cases per 1000 people) aggregated at the national level in Rwanda (A) and Uganda (C). The solid line represents the observed malaria incidence, the dashed line corresponds to the GAMM-estimated mean malaria incidence based on a model with socioeconomic without climatic information in it, and the dotted line indicates the GAMM-estimated malaria incidence based on a model with both weather and socioeconomic data. The shaded area represents the error estimates of the climate-informed model. B, D) Geographical location of the districts [Rwanda (B) and Uganda (D)] where the incorporation of weather data into a Poisson GAMM improved the predictive ability of the model compared against a further model without climatic information in it.



Contribution of the predictors

Table 3 shows the results of the analysis conducted to assess the contribution of weather to the predictive ability of the model. As can be observed, compared against a model with no climatic or socioeconomic information ($MAE_{t,h}=3.40$ cases per 1000 people), the incorporation of both $temp_{0.2}$ and $rain_{0.2}$ ($MAE_{t,h}=3.25$ cases per 1000 people) into a single model fitted to the Rwanda data does contribute to reducing the test error whilst the incorporation of annual female outpatient visits has a negligible effect on reducing such error ($MAE_{t,h}=3.40$ cases per 1000 people). The climate-informed model raised a similar error estimate as the full model ($MAE_{t,h}=3.24$ cases per 1000 people). The small contribution of outpatient data to the predictive ability of the model is not surprising due to the lack of long time series of data and the presence of random effects in the model. Similarly, the weather-informed model fitted on the Uganda data showed a lower $MAE_{t,h}$ (12.5 cases per 1000 people) than the models without $temp_{0.2}$ or $rain_{0.2}$ information in them ($MAE_{t,h}=12.7$ cases per 1000 people). As in Rwanda, the contribution of the socioeconomic predictor (urbanisation) to the predictive ability of the model was negligible likely due to the lack of long-term time series and the inclusion of random effects in the model.

As shown on Figure 10, the climate-informed models were able to capture most of the inter- and intra-annual variability in malaria incidence. The non-climate-informed model fitted on the Rwanda data was only able to capture some of this variation due to the incorporation of long-term and seasonal trend terms in the model. The non-climate-informed model fitted on the Uganda data, on the other hand, closely followed the behaviour of the climate-informed model suggesting that the overall contribution of $temp_{0.2}$ and $rain_{0.2}$ to the model's fit is low after having accounted by long-term trends and seasonality.

The incorporation of weather predictors into the model improved the model's predictive ability in 77% of the districts in Rwanda (comprising about 77% of the total population), and 60% of the districts in Uganda (containing about 62% of the total population). The districts where weather improved the predictive ability of the model are highlighted in Figure 10. It can be observed from the figure that four out the seven Rwanda districts where weather was unable to improve the model's predictive ability correspond to endemic districts where transmission is

highest. This finding is expected because, as suggested by Lowe *et al.* (2013) and Tompkins and Di Giuseppe (2015), climate variability may be less likely to impact malaria transmission in endemic districts where climate is conducive most of the year and variations in incidence may be mainly due to control interventions or data errors.

In epidemic-prone districts where transmission is usually low, on the other hand, high temperatures and precipitation may intermittently provide suitable conditions for mosquito development that may result in unusually high malaria transmission, and so, these drivers may be more likely to contribute to the estimation of malaria risk. In the Uganda model, on the other hand, weather improved the model's predictive abil-

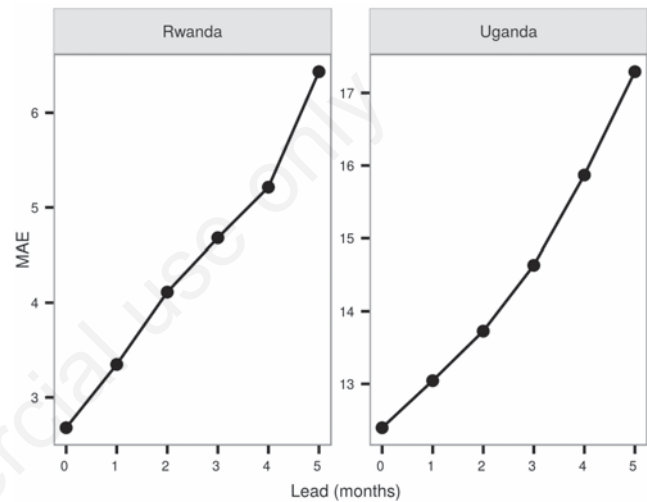


Figure 11. Error estimates of the generalised additive mixed model (GAMM)-estimated mean monthly malaria incidence on the current month, as well as one, two, three, four and five months ahead using a time series cross-validation algorithm in Rwanda and Uganda. MAE, mean absolute error.

Table 3. Error estimates for Poisson generalised additive mixed models fitted to assess the contribution of weather and non-climatic predictors to the predictive power of the generalised additive mixed model in Rwanda and Uganda.

Country	Model	Weather and socioeconomic predictors	$MAE_{t,h}$
Rwanda	1	None	3.40
	2	$Rain_{0.2}$	3.29
	3	$Temp_{0.2}$	3.59
	4	Female outpatient visits	3.40
	5	$Rain_{0.2}$ and $Temp_{0.2}$	3.25
	6	$Rain_{0.2}$ and female outpatient visits	3.28
	7	$Temp_{0.2}$ and female outpatients visits	3.58
	8	$Rain_{0.2}$, $Temp_{0.2}$ and female outpatient visits	3.24
Uganda	1	None	12.70
	2	$Rain_{0.2}$	12.05
	3	$Temp_{0.2}$	12.94
	4	Urbanisation	12.70
	5	$Rain_{0.2}$ and $Temp_{0.2}$	12.52
	6	$Rain_{0.2}$ and urbanisation	12.06
	7	$Temp_{0.2}$ and urbanisation	12.95
	8	$Rain_{0.2}$, $Temp_{0.2}$ and urbanisation	12.52

$MAE_{t,h}$, mean absolute error for each subset; $rain_{0.2}$, results for rainfall lagged zero to two months; $temp_{0.2}$, air temperature lagged zero to two months.

ity mainly in areas of high endemicity located in Regions 1, 2 and 4 (Figure 10). Compared to the results for Rwanda, this finding seems counter-intuitive. However, it is recalled that based on our data, malaria is endemic across the whole of Uganda, and so no epidemic-prone regions were detected. It is noted that the lack of epidemic-prone areas in Uganda greatly differs from the malaria endemicity maps proposed by MOH (2005) where highland areas, mainly located in the south-western part of the country, are prone to epidemics. Moreover, Regions 1, 2 and 4 in Uganda show a large year-to-year variability in malaria data in certain months of the year (see for example the box-plots for July, August and September from Figure 5), which agree well with the variations observed

in rainfall in these regions (Appendix Figure 2). Therefore, it is likely that $temp_{0,2}$ and $rain_{0,2}$ improve the predictive ability of the model in areas of great climate-driven variability. From Table 3 we noted that models including $temp_{0,2}$ resulted in lower $MAE_{t,h}$ estimates than those from models with no climatic information (models 1 and 4) only if such models also included $rain_{0,2}$ (models 5 and 8). When $temp_{0,2}$ was used in isolation or in combination with socioeconomic variables (models 3 and 7), the $MAE_{t,h}$ estimates were consistently greater than those from the models with no weather information. This finding suggests that once seasonality and long-term trends are controlled for, AMSR-E derived $temp_{0,2}$ may not be the best predictor for explaining the remaining residual variation

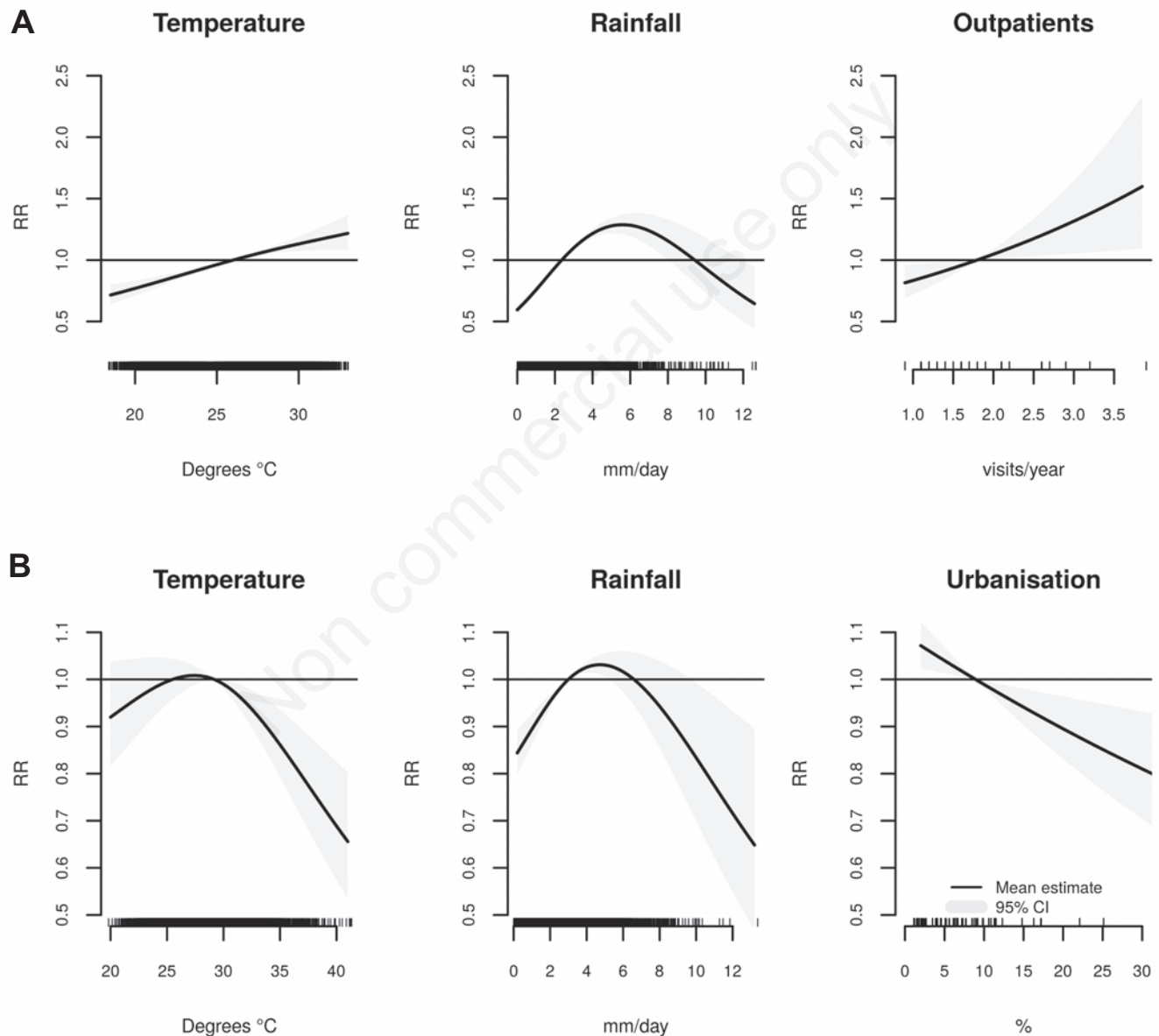


Figure 12. Generalised additive mixed model-estimated relationships among average monthly malaria incidence, monthly average near-surface air temperature, and precipitation lagged 0 to 2 months from a distributed lag nonlinear model, average number of female outpatient visits per year and urbanisation levels in Rwanda (A) and Uganda (B) after having controlled for temporal correlation in the residuals. The smooth lines indicate the restricted maximum likelihood estimates, and the shaded areas represent the 95% confidence intervals.



in malaria incidence, and poses the question as to whether other remote-sensing products would explain a larger amount of such residual variation. However, given that the reduction in the predictive ability of the model was not largely compromised by keeping $temp_{0.2}$ in, we decided to retain this variable in the model specification.

Temporal correlation

We detected significant temporal correlation (autocorrelation) in the deviance residuals at a lag of 1 month (autocorrelation function was 0.8 in Rwanda and 0.4 in Uganda), which represents a technical violation of the Poisson model. Such first order autocorrelation was evident when we computed the $MAE_{i,h}$ for the current month, as well as for test sets containing observations for one, two, three, four and five months ahead of the last observation contained in the training set. As can be observed from Figure 11, the error estimates substantially increase as we move ahead in time with a lead of zero and one months raising the smaller error estimates. This finding supports the autocorrelation function results indicating that there is a temporal autocorrelation in the malaria data of the order of about one month.

We explicitly modelled the temporal correlation by including the one month lagged deviance residuals in the model as suggested by Brumback *et al.* (2000) and Bhaskaran *et al.* (2013), and re-estimated the relationships between malaria incidence and our selected predictors. This approach significantly reduced the autocorrelation (autocorrelation function=0.3 in Rwanda, and 0.1 in Uganda). The estimated exposure-outcome relations changed moderately in Rwanda and remained the same in Uganda as can be observed in Figure 12. In Rwanda, little variation is observed in the confidence intervals of both $temp_{0.2}$ and female outpatients, with no major changes in the mean response. It is noted that whilst the confidence intervals for $temp_{0.2}$ reduced their range, the opposite occurs with the number of female outpatient visits indicating that this covariate marginally explains some of the variation in malaria incidence. We emphasise that this result is likely to be due to the very short dataset available for the female outpatients predictor and the incorporation of random effects in the model. More significant changes are observed for $rain_{0.2}$. Specifically, the relative risk of malaria incidence increased from about 1.2 to 1.3 for rainfall values between 5-7 mm day⁻¹. Additionally, the decrease in risk became slightly more gradual after having controlled by temporal correlation.

Conclusions

Here, we presented an analysis of the overall relations between malaria incidence, air temperature, and rainfall across Rwanda and Uganda. For this study we brought together, to our knowledge, the longest and more spatially diverse malaria datasets yet assembled for both countries, only closely matched in Rwanda by that of Karema *et al.* (2012). Our statistical approach based on a Poisson generalised additive mixed model with a distributed lag nonlinear specification is specifically designed to analyse the potential delayed and non-linear effects of some predictors on the outcome variable whilst accounting for the effects of unobserved variables in the model. Our TSCV algorithm has the advantage, relative to more conventional methods such as AIC, BIC, and R², that it can provide direct estimates of the test error whilst the other methods need to make adjustments in the training error to account for potential biases due to over-fitting (James *et al.*, 2013). One limitation of this study is that our malaria data is likely subject to both under-reporting (due to passive epidemiological surveil-

lance that is unable to identify asymptomatic cases, and those who do not seek medical attention), and over-reporting (due to misdiagnosis caused by lack of expertise or equipment). Thus, the estimated incidence rates may differ from those occurring in the population. This situation has been previously detected and acknowledged in a range of settings (Choge *et al.*, 2014; Kyabayinze *et al.*, 2012; Nankabirwa *et al.*, 2009; WHO, 2013; Yeka *et al.*, 2012). However, according to WHO (2013), epidemiological surveillance systems do not need to detect all cases to allow the assessment of malaria incidence. Epidemiological surveillance data constitute, to date, the only source of systematic and continuous empirical data across large spatial and temporal scales for many countries. Clinically diagnosed (suspected) malaria cases are the most common and, in many cases, the only available indicator of malaria occurrence in several African countries (Wimberly *et al.*, 2012). Laboratory confirmed cases would have been preferred for conducting this study; however, only time series of clinically diagnosed data were available to cover the entirety of both countries with a time span of about a decade. Laboratory confirmation is still limited in the region in terms of tools, skills and infrastructure although efforts are being taken to improve coverage (Kyabayinze *et al.*, 2012; Yeka *et al.*, 2012). For example, laboratory confirmation is routinely conducted in six sentinel sites across Uganda (Okello *et al.*, 2006; Zinszer *et al.*, 2015); however, the geographical and temporal coverage of the related dataset is considerably low with data freely available only for the period October 2006 to March 2013 for the longest time series. Our dataset covers a greater geographical area and spans for a larger period, allowing more robust estimations of malaria relations than the sentinel site set. A further limitation is related to the use of remote sensing products as a source for climate data. Although remote sensing indices may not necessarily correlate with the measurements obtained at weather stations (Vancutsem *et al.*, 2010), it has been demonstrated that both AMSR-E derived air temperatures and FEWS rainfall data closely correlate with observations recorded at meteorological stations in some regions (Chuang *et al.*, 2012; Cohen-Liechti *et al.*, 2012).

In conclusion, our results provide evidence that both air temperature and rainfall modulate the risk of malaria occurrence in Rwanda and Uganda. While the effects of air temperature on malaria were positive and almost linear in Rwanda, the effects of the same predictor in the Uganda model were highly nonlinear as reported by previous studies. The effects of precipitation were also significantly nonlinear as expected; and these results are in strong agreement with previous research in the field. The effects of socioeconomic development on malaria incidence are, however, difficult to ascertain due to the scarce amount of data available. This paper contributes to the field in that although the effects of both temperature and rainfall had had been previously investigated for other countries in the region (Blanford *et al.*, 2013; Garske *et al.*, 2013; Lowe *et al.*, 2013), nation-wide studies of these effects across the whole of Rwanda or Uganda have not been previously undertaken. A novel contribution of this study is that we implement a TSCV algorithm to evaluate the predictive performance of our model, and to prevent over-fitting caused by including too many degrees of freedom. TSCV provides a relatively unbiased measure of the true error of the model based on new data.

References

- Armstrong B, 2006. Models for the relationship between ambient temperature and daily mortality. *Epidemiology* 17:624-31.
- Baeza A, Bouma MJ, Dobson AP, Dhiman R, Srivastava HC, Pascual M,

2011. Climate forcing and desert malaria: the effect of irrigation. *Malaria J* 10:190.
- Beck N, Jackman S, 1998. Beyond linearity by default: generalized additive models. *Am J Polit Sci* 42:596-627.
- Beier J, Killeen G, Githure J, 1999. Short report: entomologic inoculation rates and *Plasmodium falciparum* malaria prevalence in Africa. *Am J Trop Med Hyg* 61:109-13.
- Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J, 2011. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS Med* 8:e1001083.
- Bergmeir C, Benítez JM, 2012. On the use of cross-validation for time series predictor evaluation. *Inform Sciences* 191:192-213.
- Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B, 2013. Time series regression studies in environmental epidemiology. *Int J Epidemiol* 42:1187-95.
- Blanford JI, Blanford S, Crane RG, Mann ME, Paaijmans KP, Schreiber KV, Thomas MB, 2013. Implications of temperature variation for malaria parasite development across Africa. *Sci Rep* 3:1300.
- Borcard D, Gillet F, Legendre P, 2011. *Numerical ecology with R*. Springer, New York, NY, USA.
- Bouzid M, Colón-González FJ, Lung T, Lake IR, Hunter PR, 2014. Climate change and the emergence of vector-borne diseases in Europe: case study of dengue fever. *BMC Public Health* 14:781.
- Brit O, Vounatsou P, Gunawardena D, Galappaththy G, Amerasinghe P, 2008. Temporal correlation between malaria and rainfall in Sri Lanka. *Malaria J* 7:77.
- Brumback B, Ryan L, Schwartz J, Neas L, Stark P, Burge H, 2000. Transitional regression models, with application to environmental time series. *J Am Stat Assoc* 95:16-27.
- Cameron AC, Trivedi PK, 1990. Regression-based tests for overdispersion in the Poisson model. *J Econom* 46:347-64.
- Choge JK, Magak NG, Akhwale W, Koech J, Ngeiywa MM, Oyoo-Okoth E, Esamai F, Osano O, Khayeka-Wandabwa C, Kweka EJ, 2014. Symptomatic malaria diagnosis overestimate malaria prevalence, but underestimate anaemia burdens in children: results of a follow up study in Kenya. *BMC Public Health* 14:332.
- Chowell G, Munayco CV, Escalante AA, McKenzie FE, 2009. The spatial and temporal patterns of *Falciparum* and *vivax* malaria in Peru: 1994-2006. *Malaria J* 8:142.
- Christophers SR, 1960. *Aedes aegypti*. The yellow fever mosquito. Its life history, bionomics and structure. Cambridge University Press, Cambridge, UK.
- Chuang TW, Henebry GM, Kimball JS, VanRoekel-Patton DL, Hildreth MB, Wimberly MC, 2012. Satellite microwave remote sensing for environmental modeling of mosquito population dynamics. *Remote Sens Environ* 125:147-56.
- Clark TD, Greenhouse B, Njama-Meya D, Nzarubara B, Maiteki-Sebuguzi C, Staedke SG, Seto E, Kamya MR, Rosenthal PJ, Dorsey G, 2008. Factors determining the heterogeneity of malaria incidence in children in Kampala, Uganda. *J Infect Dis* 198:393-400.
- Cohen-Liechti T, Matos JP, Boillat JL, Schleiss AJ, 2012. Comparison and evaluation of satellite derived precipitation products for hydrological modeling of the Zambezi River basin. *Hydrol Earth Syst Sc* 16:489-500.
- Colón-González FJ, Fezzi C, Lake IR, Hunter PR, 2013. The effects of weather and climate change on dengue. *PLoS Neglect Trop D* 7:e2503.
- Conrad V, 1946. *Methods in climatology*. Harvard University Press, Cambridge, MA, USA.
- Craig MH, Snow RW, Le Sueur D, 1999. A climate-based distribution model of malaria transmission in Sub-Saharan Africa. *Parasitol Today* 15:105-11.
- Dery DB, Brown C, Asante KP, Adams M, Dosoo D, Amenga-Etego S, Wilson M, Chandramohan D, Greenwood B, Owusu-Agyei S, 2010. Patterns and seasonality of malaria transmission in the forest-savannah transitional zones of Ghana. *Malaria J* 9:314.
- Douglas NM, Anstey NM, Buffet PA, Poesoprodjo JR, Yeo TW, White NJ, Price RN, 2012. The anaemia of *Plasmodium vivax* malaria. *Malaria J* 11:135.
- Faye O, Konate L, Mouchet J, Fontenille D, Sy N, Hebrard G, Herve JP, 1997. Indoor resting by outdoor biting females of *Anopheles gambiae complex* (Diptera: Culicidae) in the Sahel of northern Senegal. *J Med Entomol* 34:285-9.
- Gage K, Burkot T, Eisen R, Hayes E, 2008. Climate and vectorborne diseases. *Am J Prev Med* 35:436-50.
- Garske T, Ferguson NM, Ghani AC, 2013. Estimating air temperature and its influence on malaria transmission across Africa. *PLoS One* 8:e56487.
- Gasparrini A, 2011. Distributed lag linear and non-linear models in R: the package *dlnm*. *J Stat Softw* 43:1-20.
- Gasparrini A, Armstrong B, Kenward M, 2010. Distributed lag non-linear models. *Stat Med* 29:2224-34.
- Gething PW, Smith DL, Patil AP, Tatem AJ, Snow RW, Hay SI, 2010. Climate change and the global malaria recession. *Nature* 465:342-5.
- Githeko AK, Service MW, Mbogo CM, Atieli FK, 1996. Resting behaviour, ecology and genetics of malaria vectors in large-scale agricultural areas of western Kenya. *Parassitologia* 38:481-9.
- Guinovart C, Bassat Q, Sigauque B, Aide P, Sacarlal J, Nhampossa T, Bardaji A, Nhacolo A, Macete E, Mandomando I, Aponte JJ, Menéndez C, Alonso PL, 2008. Malaria in rural Mozambique. Part I: Children attending the outpatient clinic. *Malaria J* 7:36.
- Halstead SB, 2008. Dengue virus-mosquito interactions. *Annu Rev Entomol* 53:273-91.
- Haque U, Hashizume M, Glass G, Dewan A, Overgaard H, Yamamoto T, 2010. The role of climate variability in the spread of malaria in Bangladeshi highlands. *PLoS One* 5:e14341.
- Henninger SM, 2013. Local climate changes and the spread of malaria in Rwanda. *Sci Res* 5:728-34.
- Hijmans RJ, 2013. *Raster: geographic data analysis and modeling*. Available from: <http://CRAN.R-project.org/package=raster>
- Hyndman RJ, Athanasopoulos G, 2014. *Forecasting: principles and practice*. Available from: <https://www.otexts.org/book/fpp>
- IRI, 2012. Estimated daily precipitation (RFEv2) in Africa from the famine early warning system. Available from: <http://iridl.ldeo.columbia.edu/SOURCES/NOAA/NCEP/CPC/FEWS/Africa/DAILY/RFEv2>
- James G, Witten D, Hastie T, Tibshirani R, 2013. *An introduction to statistical learning with applications in R*. Springer, New York, NY, USA.
- Johnston J, DiNardo J, 1997. *Econometric methods*. McGraw-Hill International, Singapore.
- Jones LA, Ferguson CR, Kimball JS, Zhang K, Chan S, McDonald KC, Njoku EG, Wood EF, 2010. Satellite microwave remote sensing of daily land surface air temperature minima and maxima from AMSR-E. *IEEE J Sel Top Appl* 3:111-23.
- Kamya MR, Arinaitwe E, Wanzira H, Katureebe A, Barusya C, Kigozi SP, Kilama M, Tatem AJ, Rosenthal PJ, Drakeley C, Lindsay SW, Staedke SG, Smith DL, Greenhouse B, Dorsey G, 2015. Malaria transmission, infection, and disease at three sites with varied transmission intensity in Uganda: implications for malaria control. *Am J Trop Med Hyg* 92:903-12.
- Karema C, Aregawi MW, Rukundo A, Kabayiza A, Mulindahabi M, Fall



- IS, Gausi K, Williams RO, Lynch M, Cibulskis R, Fidele N, Nyemazi JP, Ngamije D, Umulisa I, Newman R, Binagwaho A, 2012. Trends in malaria cases, hospital admissions and deaths following scale-up of anti-malarial interventions, 2000-2010, Rwanda. *Malaria J* 11:236.
- Kazembe LN, Kleinschmidt I, Holtz TH, Sharp BL, 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. *Int J Health Geogr* 5:41.
- Kelly-Hope LA, McKenzie FE, 2009. The multiplicity of malaria transmission: a review of entomological inoculation rate measurements and methods across Sub-Saharan Africa. *Malaria J* 8:19.
- Klinkenberg E, McCall P, Wilson MD, Amerasinghe FP, Donnelly MJ, 2008. Impact of urban agriculture on malaria vectors in Accra, Ghana. *Malaria J* 7:151.
- Kyabayinze DJ, Achan J, Nakanjako D, Mpeka B, Mawejeje H, Mugizi R, Kalyango JN, D'Alessandro U, Talisuna A, Jean-Pierre Vg, 2012. Parasite-based malaria diagnosis: are health systems in Uganda equipped enough to implement the policy? *BMC Public Health* 12:695.
- Legendre P, Rogers D, 1972. Characters and clustering in taxonomy: a synthesis of two taximetric procedures. *Taxon* 21:567-606.
- Lin E, Kiniboro B, Gray L, Dobbie S, Robinson L, Laumaea A, Schöpflin S, Stanicic D, Betuela I, Blood-Zikursh M, Siba P, Felger I, Schofield L, Zimmerman P, Mueller I, 2010. Differential patterns of infection and disease with *P. falciparum* and *P. vivax* in young Papua New Guinean children. *PLoS One* 5:e9047.
- Lindsay SW, Parson L, Thomas CJ, 1998. Mapping the range and relative abundance of the two principal African malaria vectors, *Anopheles gambiae sensu stricto* and *An. arabiensis*, using climate data. *P R Soc London* 265:847-54.
- Loevinsohn ME, 1994. Climate warming and increased malaria incidence in Rwanda. *Lancet* 343:714-8.
- Lowe R, Chirombo J, Tompkins AM, 2013. Relative importance of climatic, geographic and socio-economic determinants of malaria in Malawi. *Malaria J* 12:416.
- Lunde TM, Bayoh MN, Lindtjørn B, 2013. How malaria models relate temperature to malaria transmission. *Parasite Vector* 6:20.
- Menendez C, Fleming AF, Alonso PL, 2000. Malaria-related anaemia. *Parasitol Today* 16:469-76.
- Mnzava AEP, Rwegoshora RT, Wilkes TJ, Tanner M, Curtis C, 1995. *Anopheles arabiensis* and *Anopheles gambiae* chromosomal inversion polymorphism, feeding and resting behavior in relation to insecticide house spraying in Tanzania. *Med Vet Entomol* 9:316-24.
- MOH, 2005. Uganda malaria control strategic plan 2005/06-2009/10. Ministry of Health of Uganda, Kampala, Uganda.
- Naish S, Dale P, Mackenzie JS, McBride J, Mengersen K, Tong S, 2014. Spatial and temporal patterns of locally-acquired dengue transmission in Northern Queensland, Australia, 1993-2012. *PLoS One* 9:e92524.
- Nankabirwa J, Zurovac D, Njogu JN, Rwakimari JB, Counihan H, Snow RW, Tibenderana JK, 2009. Malaria misdiagnosis in Uganda. Implications for policy change. *Malaria J* 8:66.
- NASA, 2013. Advanced microwave scanning radiometer. Earth observing system sensor on the NASA. Available from: <https://nsidc.org/data/amsre>
- NISR, MOH and ICF International, 2012. Rwanda demographic and health survey 2010. Report, National Institute of Statistics of Rwanda and Ministry of Health of Rwanda, Kigali, Rwanda, and ICF International, Calverton, ML, USA.
- Okeke TA, Okeibunor JC, 2010. Rural-urban differences in health-seeking for the treatment of childhood malaria in South-East Nigeria. *Health Policy* 95:62-8.
- Okello P, van Bortel W, Byaruhanga A, Correwyn A, Roelants P, Talisunam A, D'Alessandro U, Coosemans M, 2006. Variation in malaria transmission intensity in seven sites throughout Uganda. *Am J Trop Med Hyg* 75:219-25.
- Opsomer J, Wang Y, Yang Y, 2001. Nonparametric regression with correlated errors. *Stat Sci* 16:134-53.
- Paaijmans KP, Jacobs AFG, Takken W, Heusinkveld BG, Githeko AK, Dicke M, Holtslag AAM, 2008a. Observations and model estimates of diurnal water temperature dynamics in mosquito breeding sites in western Kenya. *Hydrol Process* 22:4789-801.
- Paaijmans KP, Takken W, Githeko AK, Jacobs AFG, 2008b. The effect of water turbidity on the near-surface water temperature of larval habitats of the malaria mosquito *Anopheles gambiae*. *Int J Biometeorol* 52:747-53.
- Paaijmans KP, Thomas MB, 2011. The influence of mosquito resting behaviour and associated microclimate for malaria risk. *Malaria J* 10:183.
- Paaijmans KP, Wandago MO, Githeko AK, Takken W, 2007. Unexpected high losses of *Anopheles gambiae* larvae due to rainfall. *PLoS One* 2:e1146.
- Perry PO, 2009. Bcv: cross-validation for the SVD (bi-cross-validation). R package version 1.0. Available from: <https://cran.r-project.org/web/packages/bcv/bcv.pdf>
- Pindolia DK, Garcia AJ, Huang Z, Smith DL, Alegana VA, Noor AM, Snow RW, Tatem AJ, 2013. The demographics of human and malaria movement and migration patterns in East Africa. *Malaria J* 12:397.
- R Development Core Team, 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Reiter P, 2001. Climate change and mosquito-borne disease. *Environ Health Persp* 109:141-61.
- Reiter P, 2008. Global warming and malaria: knowing the horse before hitching the cart. *Malaria J* 7:3.
- Robert V, Macintyre K, Keating J, McWilson W, Trappe JP, Duchemin JB, Beier JC, 2003. Malaria transmission in urban Sub-Saharan Africa. *Am J Trop Med Hyg* 69:169-76.
- Rulisa S, Kateera F, Bizimana JP, Agaba S, Dukuzumuremyi J, Baas L, de Dieu Harelimana J, Mens PF, Boer KR, de Vries PJ, 2013. Malaria prevalence, spatial clustering and risk factors in a low endemic area of eastern Rwanda: a cross sectional study. *PLoS One* 8:e69443.
- SEDAC, 2012. Gridded population of the world, version 3. Available from: <http://sedac.ciesin.columbia.edu/data/collection/gpw-v3>
- Shannon W, Culverhouse R, Duncan J, 2003. Analyzing microarray data using cluster analysis. *Pharmacogenomics* 4:41-52.
- Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, Patil AP, Temperley WH, Gething PW, Kabaria CW, Okara RM, Van Boeckel T, Godfray HC, Harbach RE, Hay SI, 2010. The dominant anopheles vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. *Parasite Vector* 3:117.
- Smith T, Charlwood JD, Kitua AY, Masanja H, Mwankusye S, Alonso PL, Tanner M, 1998. Relationships of malaria morbidity with exposure to *Plasmodium falciparum* in young children in a highly endemic area. *Am J Trop Med Hyg* 59:252-7.
- Stryker JJ, Bomblies A, 2012. The impacts of land use change on malaria vector abundance in a water-limited, highland region of Ethiopia. *Ecohealth* 9:455-70.
- Tan P, Steinbach M, Kumar V, 2005. Introduction to data mining.

- Addison-Wesley Longman, Boston, MA, USA.
- Thomson MC, Mason SJ, Phindela T, Connor SJ, 2005. Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana. *Am J Trop Med Hyg* 73:214-21.
- Tompkins AM, Di Giuseppe F, 2015. Potential predictability of malaria in Africa using ECMWF monthly and seasonal climate forecasts. *J Appl Meteorol Clim* 54:521-40.
- Tompkins AM, Ermert V, 2013. A regional-scale, high resolution dynamical malaria model that accounts for population density, climate and surface hydrology. *Malaria J* 12:65.
- Tompkins AM, McCreesh N, 2016. Migration statistics relevant for malaria transmission in Senegal derived from mobile phone data and used in an agent-based migration model. *Geospat Health* 11:408.
- Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB, 2001. Missing value estimation methods for DNA microarrays. *Bioinformatics* 17:520-5.
- UBOS, 2002. 2002 Uganda population and housing census. Uganda Bureau of Statistics, Kampala, Uganda.
- UBOS and ICF International, 2009. Uganda malaria indicator survey (MIS) 2009. Uganda Bureau of Statistics, Kampala, Uganda, and ICF International, Calverton, ML, USA.
- UBOS and ICF International, 2012. Uganda demographic and health survey 2011. Uganda Bureau of Statistics, Kampala, Uganda, and ICF International, Calverton, ML, USA.
- USAID, 2013a. President's malaria initiative: malaria operational plan FY 2013 - Rwanda. US Agency for International Development, Washington, DC, USA.
- USAID, 2013b. President's malaria initiative: malaria operational plan FY 2014 - Uganda. US Agency for International Development, Washington, DC, USA.
- Vancutsem C, Ceccato P, Dinku T, Connor SJ, 2010. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sens Environ* 114:449-65.
- WHO, 2011. World malaria report 2011. World Health Organization, Geneva, Switzerland.
- WHO, 2013. World malaria report 2013. World Health Organization, Geneva, Switzerland.
- Willmott CJ, Matsuura K, 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res* 30:79-82.
- Wimberly MC, Midekisa A, Semuniguse P, Teka H, Henebry GM, Chuang T-W, Senay GB, 2012. Spatial synchrony of malaria outbreaks in a highland region of Ethiopia. *Trop Med Int Health* 17:1192-201.
- Wood SN, 2006. Generalized additive models: an introduction with R. Chapman & Hall/CRC, FL, USA.
- Wooldridge JM, 2002. Econometric analysis of cross section and panel data. MIT Press, London, UK.
- World Bank, 2014. Population (total). Available from: <http://data.worldbank.org/indicator/SP.POPTOTL>
- WRI, 2012. Uganda GIS data. Available from: <http://www.wri.org/resources/data-sets/uganda-gis-data>
- Yeka A, Gasasira A, Mpimbaza A, Achan J, Nankabirwa J, Nsoby S, Staedke SG, Donnelly MJ, Wabwire-Mangen F, Talisuna A, Dorsey G, Kanya MR, Rosenthal PJ, 2012. Malaria in Uganda: challenges to control on the long road to elimination: I. Epidemiology and current control efforts. *Acta Trop* 121:184-95.
- Zinszer K, Kigozi R, Charland K, Dorsey G, Brewer TF, Brownstein JS, Kanya MR, Buckeridge DL, 2015. Forecasting malaria in a highly endemic country using environmental and clinical predictors. *Malaria J* 14:245.