Grant agreement no. 243964

QWeCI

Quantifying Weather and Climate Impacts on Health in Developing Countries

D2.1b Report on the performance of the dynamic and semi-dynamical modelling approaches for selected diseases for regions in Africa

Start date of project: 1st February 2010 Duration: 42 months

Lead contractor :	UNILIV
Coordinator of deliverable :	UNILIV
Evolution of deliverable	
Due date :	M28
Date of first draft :	1 July 2013
Start of review :	10 July 2013
Deliverable accepted :	30 July 2013

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	PU
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

General Introduction

This deliverable provides information about the state of the art modelling outputs produced by the various QWeCI partners involved in the WP2.1 "Development of dynamic disease models". Some of those results have already been published in peer-reviewed journals (Tompkins et al., 2012; Ermert et al., 2012) or are in preparation for submission (as a consequence this deliverable should not be made publicly available to avoid copyright issues with peer-reviewed publications). This deliverable was delayed (delivered at the end of the project instead of M24 as originally planned) in order to allow the various QWeCI partners to develop, test, parameterize and validate their modelling approaches with various sources of observations. The QWeCI teams have mainly focused on two vector-borne diseases: malaria and Rift Valley Fever for different regions in Africa (from the continental scale to the local scale, such as the Barkedji area in northern Senegal).

Malaria is caused in humans by infection with the protozoan Plasmodium, and is transmitted between humans by female mosquito vectors from the Anopheles spp. The first symptoms are relatively similar to those of seasonal flu, with fever, sore throat, pain, chills and aches; and sometimes nausea and diarrhoea that can lead to more serious health issues. Infection with the most severe form of the parasite, Plasmodium falciparum, if not promptly treated, may lead to kidney failure, seizures, mental confusion, coma, and death. Epidemics of the disease can be triggered by factors affecting human, vector or parasite populations including abnormal meteorological conditions, changes in anti-malarial programs, population movement, and environmental changes (Nájera et al., 1998). The mosquitoes' breeding sites (ponds) and the lifecycle of the malaria parasite are both strongly connected to climatic variability, especially rainfall and temperature. Climate-driven models of malaria provide a quantitative method of considering the impact of climate on malaria transmission solely. Investment in eradication programmes over the last decade has resulted in significant progress in reducing the worldwide burden of malaria, with half the affected counties set to reach the World Health Assembly and Roll Back Malaria goal of a reduction of 75% of malaria cases between 2000 and 2015 (WHO, 2012). Mortality due to malaria in the WHO African region decreased 33% between 2000 and 2010; however, sub-Saharan African still bears the largest burden of the disease, with 91% of the 660,000 worldwide deaths occurring in the region (WHO 2012), and where malaria accounts for 15% of post-neonatal deaths (Liu et al., 2012).

Rift Valley fever (RVF) is a viral zoonosis that affects domestic animals and humans by causing an acute fever. This disease is caused by the RVF virus that belongs to the genus *Phlebovirus* in the family *Bunyaviridae*. The virus is transmitted to vertebrate hosts by the bite of infected mosquitoes, typically by the *Aedes* and *Culex* species. RVF mainly affects domestic animals (cattle, goats, sheep and camels, among others). It generally causes high mortality and abortions in pregnant females (this is how a RVF outbreak in animals is generally suspected). Human infections are mainly caused by direct or indirect contact with viraemic animal blood or infected organs during butchering, slaughtering or veterinary procedures. The human symptoms are characterized by the onset of high fever, headache, generalised weakness and liver abnormalities. In a small percentage of the infected human population RVF can cause haemorrhagic fever, encephalitis and ocular disease and this can sometimes lead to death.

The following sections of this report provide information about the various malaria modelling exercises (UNILIV [A], UOC [B], ICTP-UOC [C], IC3-ICTP-UNIMA [D], UP [H], IC3 [I]) and RVF model development (IPD [E], CSE-UCAD [F], ILRI-IC3 [G]) carried out within the QWeCI project framework. Each section is structured in the form of a scientific publication. This study covers the major WP2.1 tasks described in the QWeCI Description of Work.

[A] The Liverpool Malaria Model

Overview

The Liverpool Malaria Model (LMM 2004) (Hoshen and Morse, 2004) is a dynamic, process-based model of malaria which consists of a mosquito population model coupled with a malaria transmission model. A compartmental approach is used to numerically solve the differential equations associated with the system, with some parameters varying as a function of the daily climate time-series used to drive the model. The mosquito population is modelled using larval and adult stages, with the number of eggs deposited into breeding sites depending on the previous ten days' (dekadal) rainfall, the larval mortality rate also dependent on dekadal rainfall, an adult mosquito mortality rate dependent on temperature, and an egg-laying/biting (gonotrophic) cycle also dependent on temperature. In the malaria transmission component of the model, temperature-dependencies occur in the rate of development of the parasite within the mosquito (sporogonic cycle) and the mosquito biting rate. Both the sporogonic and gonotrophic cycles progress at a rate dependent on the number of "degree days" above a specific temperature threshold. The gonotrophic cycle takes approximately 37 degree days with a threshold of 9 °C, whereas the sporogonic cycle takes approximately 111 degree days with a threshold of 18 °C. This latter threshold is one of the most critical areas of sensitivity in the model, and below it no parasite development can occur. Further details of the model are given elsewhere (Hoshen and Morse, 2004). The LMM, driven by climate reanalysis, has been shown to be capable of simulating the interannual variability of malaria in Botswana, as compared against a 20-year anomaly index of the disease derived from malaria observations, and has been used to evaluate the potential for malaria early warning using seasonal climate forecasts (Jones and Morse, 2010; 2012).

Assessment of intra-seasonal dynamics of malaria as simulated by LMM has previously determined that the model exhibits a lag of one to two months or more, for example compared to the seasonality model of MARA (Tanser et al., 2003; comparison described in Jones, 2007). This lag is caused by the slow response of the growing (modelled) mosquito population at the start of the season to the small "trickle" of infection used to seed the model. Adjustment of the original (LMM 2004) model parameters was not able to reduce this lag without an unrealistic shift in the seasonality of the disease (Jones, 2007).

As part of this study, three adjustments to the LMM were evaluated in terms of their ability to improve model response at the start of the season, as detailed below.

Larval population response

As part of the LMM development work reported by Ermert et al. (2011a, 2011b) the larval component of the model was extended to include multiple functional forms of the relationship between the larval mosquitoes and climate inputs. The original, fixed relationship between larval daily survival rate, P_i and the daily value of 10-day accumulated rainfall in mm, R_i was amended to include additional parameters L_0 and L_F. (Equation 1).

$$P_i = L_0 + L_F \frac{R_i + 1}{R_i + 2} \qquad \text{Equation 1}$$

In this study we assessed the sensitivity of the modelled malaria incidence to this component of the model by considering the change in the response at the start of the season when the parameter settings were varied between L_F=1, L_0=0 (equivalent to the original LMM2004 model) and L_F=0,L_0=1 (larval survivorship is independent of rainfall).

Dry seasonal mosquito population

Mosquitoes and malaria infection are introduced in the standard LMM by a "trickle" of a small number of infectious mosquitoes added every 10 days. Here we investigated the impact on the dynamics of the model of the inclusion of a further inflow of uninfected mosquitoes, as might occur from mosquito populations sustained via permanent ponds or artificial water bodies such as irrigation tanks and wells.

Reservoir of infection from the immune population

Existing versions of the model (Hoshen and Morse, 2004; Ermert et al, 2011s, 2011b) do not explicitly include immunity within the human population. Acquired immunity to malaria is built up by repeated exposure of the population, leaving only young children and pregnant women (who lose their immunity during pregnancy) at risk of severe disease. In endemic areas, LMM is therefore considered to represent only non-immune proportion of the population, but this assumes that the infection of the immune or partially-immune population, who may harbour low-level, chronic infections lasting many months, does not impact on the dynamics of infection in the susceptible population. In reality, some authors suggest that gametocytogenesis (production of the sexual form of the parasite responsible for infection of the mosquito) in partially-immune "carriers" in the adult population are the source of infection in seasonally endemic areas, and even that gametocytogenesis might be triggered by factors such as an increased mosquito biting rate at the start of the season (Drakeley et al., 2006; Paul et al., 2004).

Although agent-based simulation models for immunity have been developed (McKenzie and Bossert, 2005), they are incredibly complex and not easily transferable to the LMM, which uses the simpler, less computationally-intensive compartmental approach. Here, we do not yet aim to simulate either the intra-seasonal or inter-seasonal dynamics of immunity but instead consider the maximum potential impact of an immune human "reservoir" of infection on the model response for those susceptible at the start of the season. This is achieved by adding a fourth category to the human population in the model: that of a permanently immune, chronically-infected human population which interacts with the other categories only by infection of the vector population. Future development of the LMM will consider the inter-seasonal dynamics of the immune category, which could, for example, in an early-warning context, consist of estimating the immune proportion of the human population at the start of the season as a function of previous years' observed (or simulated) malaria incidence.

Disease model input dataset and disease observation dataset

Daily rainfall and temperature are used to drive the LMM. Various sets of "observed" gridded daily rainfall products were employed in this study. Mixed satellite and rain gauges observations from the Global Precipitation Climatology Project (GPCP) dataset (Huffman et al, 2001) and from the NASA Goddard Space Flight Center Tropical Rainfall measuring mission (TRMM) dataset (Huffman et al, 2001) were employed to drive the LMM. Rainfall products

based on NCEP-NCAR (Kalnay et al, 1996) and ERA Interim (ERAI, Dee et al., 2011) reanalysis (blend of climate model outputs and various sources of observations using complex assimilation methods) were also used.

Daily temperatures were estimated using NCEP-NCAR (Kalnay et al, 1996) and ERAI (Dee et al., 2011) reanalysis. A one year spin-up was first performed, based on the daily climatology of rainfall and temperature, before running the model with each dataset. The model was run using ERAI and NCEP rainfall and temperature, and two hybrid runs using GPCP/TRMM satellite rainfall estimates and ERAI temperatures were also been produced (in this case the temperature data has been interpolated on the rainfall data grid). Table 1 summarizes the different model runs (time periods, spatial resolution, etc.). For assessment of the different LMM parameter settings, only the TRMM-ERAI input data were used.

Malaria Simulation	Period	Spatial Resolution	Comments
NCEP	1980-2010	2°x2° (~200km^2)	Coarse datasets. Large rainfall biases over West Africa.
ERAI	1989-2010	1°x1° (~100km^2)	Better reanalysis product than NCEP and former ERA40 reanalysis products for West Africa.
GPCP-ERAI	1997-2007	1°x1° (~100km^2)	Rainfall from GPCP satellite estimates and temperature from ERAI reanalysis.
TRMM-ERAI	1998-2010	0.25°x0.25° (~25km^2)	Rainfall from TRMM satellite estimates and temperature from ERAI reanalysis. Most relevant dataset to use at the country level for impact applications (higher spatial resolution).

Table1: Summary of input data for malaria simulations performed with the LMM (lmm std version).

Model outputs for different input data sources were compared with an external control in the form of mapped malaria endemicity from the Malaria Atlas Project, MAP (estimates for 2010; Gething et al., 2011). Note that the MAP data is based on a statistical Bayesian model which incorporates environmental predictors with real census data to provide "best guess" estimates of P. Falciparum parasite prevalence rate and endemicity classes in the most vulnerable population, the age category 2-10 years old. The original data (1km x 1km resolution) was interpolated to a 0.25°x0.25° grid (about 25km x 25km at the equator).

Two observed malaria datasets (time series) were also used for LMM performance assessment at a regional scale:

1. Senegal – Monthly recorded cases of malaria for two nearby towns in the Louga region of north-central Senegal: Linguère (January 2001-May 2010) and Dahra (January 2003-May 2010).
2. South Africa – Monthly recorded cases of malaria for Limpopo province for the period January 1999 to November 2005. (Source: Ministry of Health).

Performance assessment

The LMM performance is assessed using the following double-stranded approach:

- LMM-simulated transmission and high-variability epidemic "fringe" for the West Africa region and for the African continent are compared with mapped malaria endemicity estimates from the Malaria Atlas Project (Gething et al., 2011) and outputs from the MARA-ARMA project over West Africa.

- Seasonal cycle and interannual variability of LMM simulations are compared with observed malaria data from South Africa and Senegal for different parameter settings, incorporating two new components to LMM. (Table 2)

Name	Details	Adjusted LMM parameter value(s)
lmm std	Standard LMM2004	none
lmm l0	Larval survival probability insensitive to rainfall	LR0=1 LRF=0
lmm imm	Additional transmission to/from a permanently infected, immune human population (parameter is size relative to non-immune population)	IPOP=1
lmm imm_l0	Immune transmission plus reduced larval sensitivity	LR0=1 LRF=0 IPOP=1
lmm imm_m	Immune transmission plus influx of uninfected mosquitoes per non-immune human added every 10 days	IPOP=1 UIMOS=1000

Table 2: Summary of LMM parameter settings assessed against observed malaria data.

Results

Comparison between LMM and the MAP 2010 mapped malaria risk

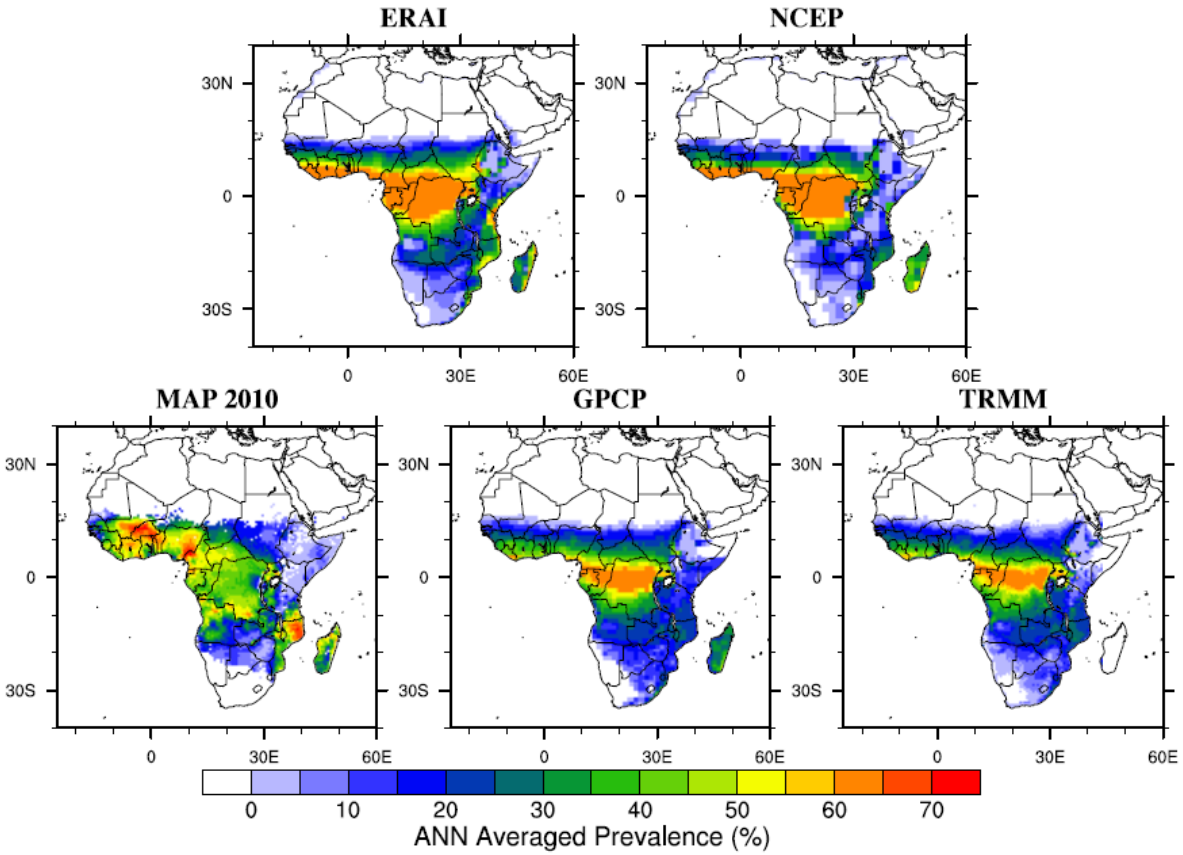


Figure 1.1 Mean annual prevalence as simulated by the LMM driven by different gridded climate datasets over Africa (the climatology is calculated over the periods shown in table 1). The MAP 2010 Parasite Prevalence rate map is shown on the lower left panel.

Figure 1.1 & 1.2 depict the mean annual malaria prevalence as simulated by the LMM driven by different gridded climate datasets and *P. falciparum* parasite prevalence rate based on the MAP 2010 dataset. The LMM prevalence distribution exhibits a zonal pattern, with large values simulated over central tropical Africa (DRC, Gabon, CAR, and Cameroon), over the coasts of the Gulf of Guinea, and over the eastern coasts of Kenya, Tanzania and Mozambique for the reanalysis-driven runs. Simulated prevalence then decreases as a function of the latitude from the equator to the Sahel and from the equator to southern Africa (Fig 1.1).
The standard version of the LMM seems to simulate the northern edge of low malaria prevalence too far south with respect to the MAP and the MARA data. The southern boundary is extending too far south over southern Africa with respect to the MAP2010 analysis. The prevalence hot spots shown over southern Mali, south-western Nigeria and northern Ivory Coast for MAP2010 are not reproduced by the LMM (Fig 1.2). The LMM simulates larger prevalence values over southern Ghana with respect to northern Ghana, while the opposite is shown by the MARA and MAP2010 datasets over Ghana. Over eastern Africa, the areas free of malaria are relatively well captured by the LMM whereas the large prevalence values simulated over the eastern coasts of Kenya look unrealistic (a similar problem is seen with

Vectri; see section C). The LMM reproduces the prevalence hot spots over Mozambique and south-eastern Tanzania when driven by the ERAI climate data. The LMM prevalence distribution is really similar to the MAP estimates over Angola, Zambia, Botswana and Namibia.

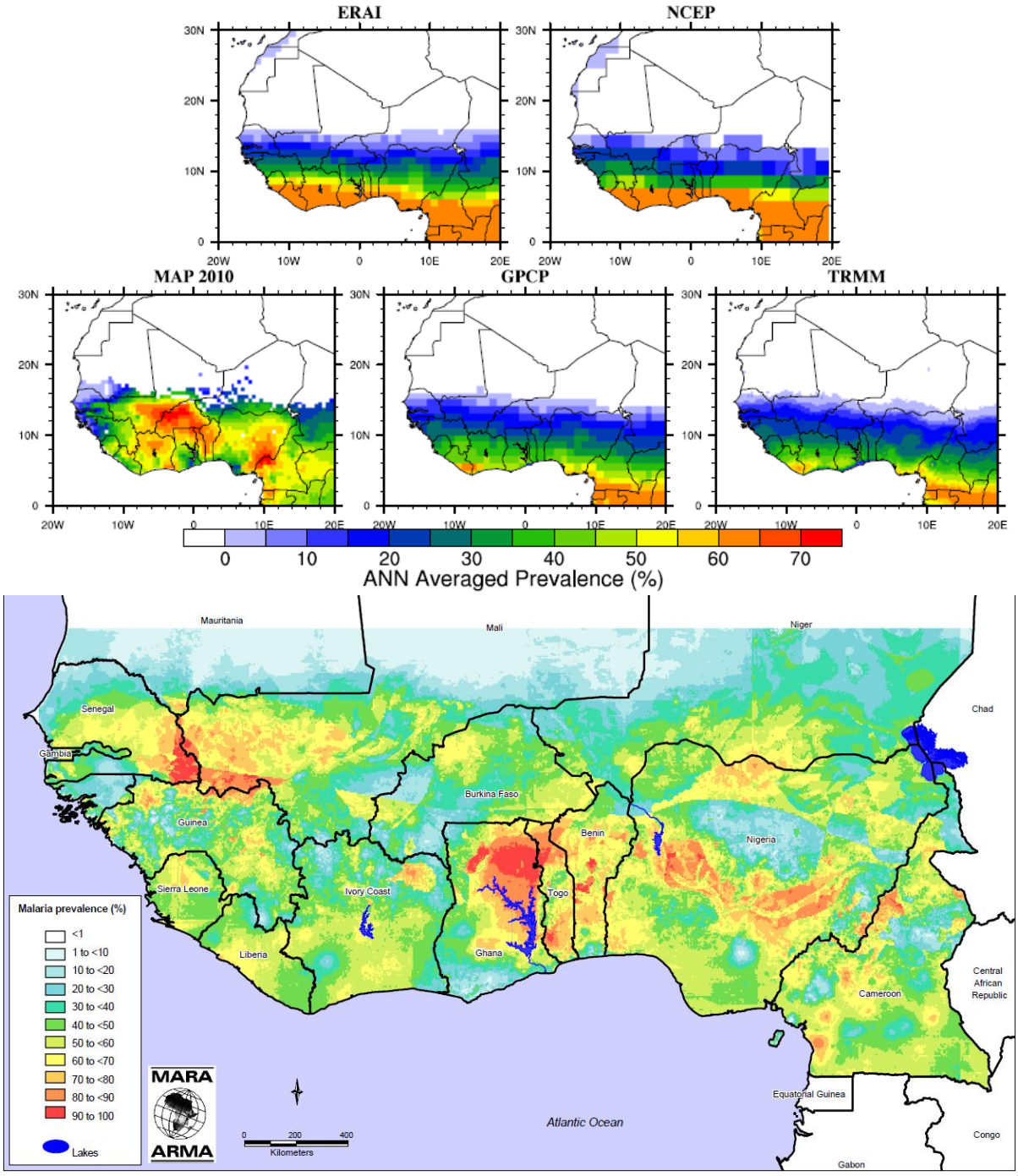


Figure 1.2 Same as Figure 1.1 for the West African region. The lower panel depicts simulated Malaria Prevalence by the MARA statistical model over West Africa (Tanser et al., 2003).

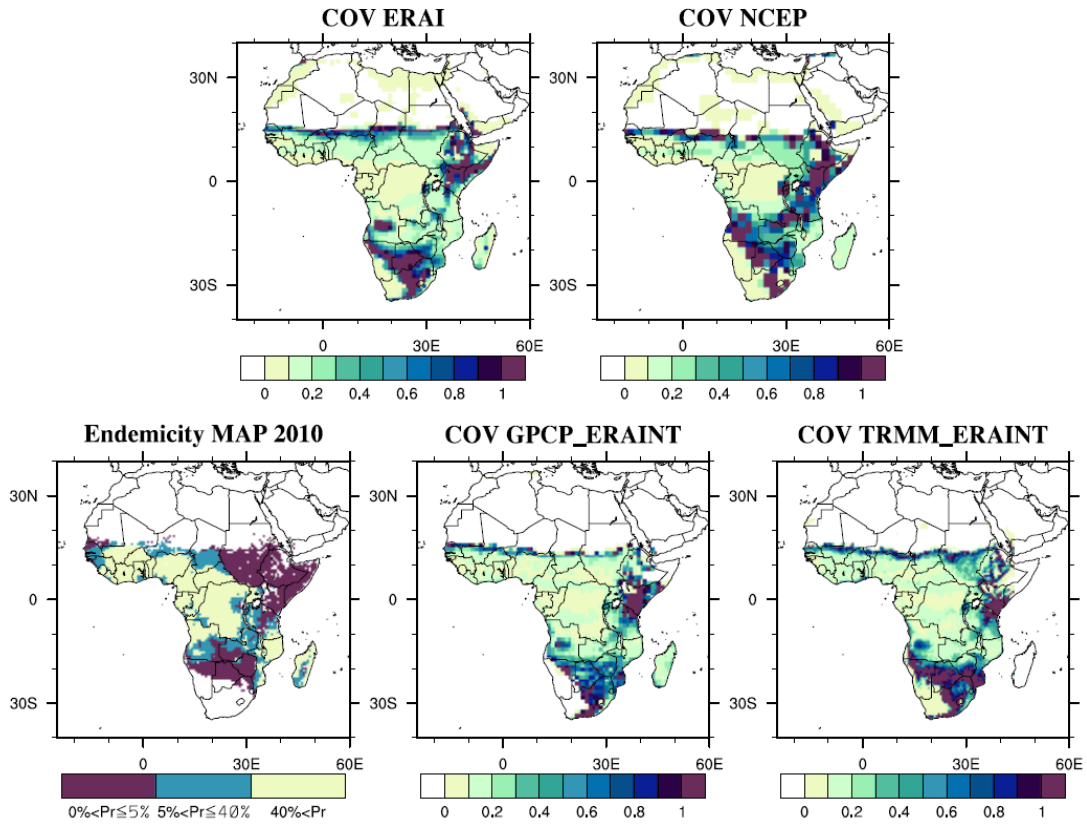


Figure 1.3 Annual coefficient of variation (COV) for Malaria Incidence as simulated by the LMM driven by different gridded climate datasets over Africa (the COV is the ratio between the standard deviation divided by the mean calculated over the different time period shown in table 1). The MAP 2010 endemicity classes map is shown on the lower left panel. Note that this is another output of the MAP statistical model which is not based on the mean prevalence categories depicted in Fig 1.1 and Fig 1.2.

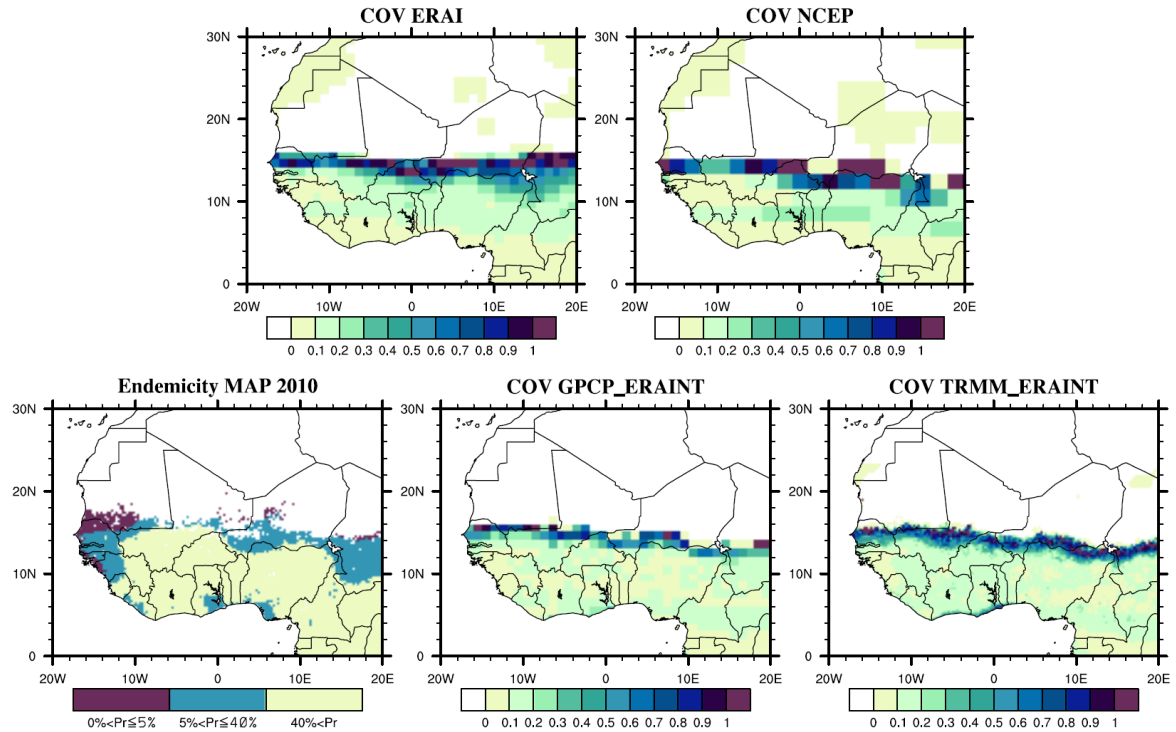


Figure 1.4 Same as Figure 1.3 for the West African region.

Figures 1.3 and 1.4 discriminate epidemic and endemic malaria areas as simulated by the LMM (using the coefficient of variation which characterizes large year to year variability in simulated malaria incidence) and as modelled by the MAP2010 analysis. The LMM epidemic belt is placed at the northern edge of the Sahel, crossing northern Senegal, southern-central Mali, and southern Niger towards central Sudan. The Sahelian epidemic belt is simulated too far south with respect to the MAP2010 data. Most of Senegal, Gambia and Guinea Bissau are shown to be epidemic area by MAP2010 while this is not reproduced by the LMM. A large epidemic area is simulated by the LMM over eastern Africa including most of Ethiopia, northern Kenya, Somalia and northern Tanzania. This is relatively consistent with outputs from MAP2010 (excepting over southern Sudan which is highly epidemic according to MAP2010). The endemic/epidemic areas are also well discriminated by the LMM (satellite driven runs e.g. TRMM/GPCP) over Botswana, Namibia and Angola when comparing it with MAP2010. Only the Limpopo region is shown to be epidemic over South Africa by MAP2010, while this covers a larger area over South Africa by the LMM.

The MAP2010 data is an analysis (statistical model) which merges malaria survey data with different environmental & socio-economic predictors to provide estimates of malaria endemicity on a high resolution gridded map. The number of surveys in this analysis is critical to provide robust estimates. The survey network is dense enough over West Africa, the western coasts of central to southern Africa and eastern Africa to have confidence in those estimates (Fig 2A from Gething et al. 2011). The values simulated over central Africa mainly rely on the predictors (as the survey data do not exist or just a few points are available), in other words the MAP data is more a model output (such as LMM) over central Africa.

The differences between the LMM and the MAP2010 analysis are sometimes large (especially over West Africa and South Africa). This is not surprising as the standard version of the LMM only considers the impact of climate on malaria transmission; while the MAP2010 analysis indirectly includes some measures of intervention (through the incorporation of survey data) and other critical socio-economic factors (poverty, demography, urbanisation...). As an example, large control measures almost eradicate malaria from South Africa excepting over the north-east (mainly Limpopo). Interventions (control measures) are not taken into account by the LMM.

However, the epidemic fringe distribution is relatively well simulated by the LMM; and those epidemic areas are critical as they include susceptible population with low immunity to P. Falciparum (where malaria has the largest negative impact on human well-being). Furthermore malaria seasonal forecasts performed with the LMM have been shown to be skilful over epidemic areas in West Africa (Jones et al., 2012) and over Botswana (Jones et al., 2010).

Comparison against observed malaria

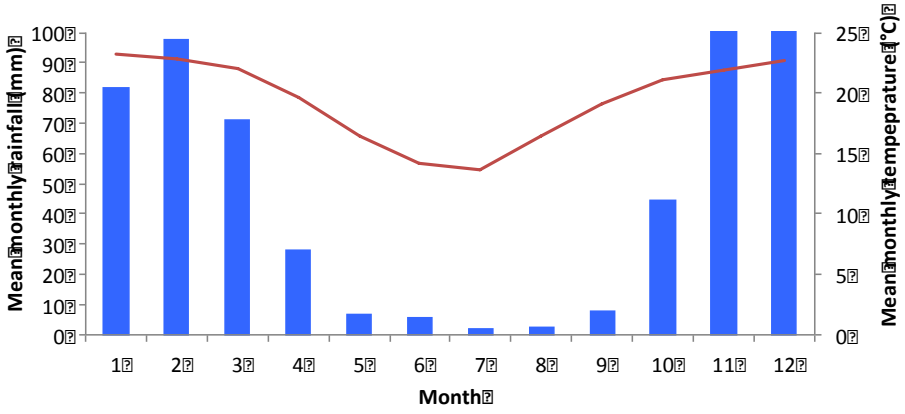
Limpopo Province, South Africa

The seasonal cycle of the climate data used to drive the LMM for a single location in Limpopo province – the capital, Polokwane, are shown in Figure 2.1a. The corresponding seasonal cycles in simulated incidence are compared for different parameter settings with that calculated from Ministry of Health records in Figure 2.1b. The dry season extends from May to September, with rainfall increasing in October and peaking in November/December and then again in February. On average temperatures are below the sporogonic threshold of 18 °C between May and August inclusive, with a mean temperature of 19.1 °C in September.

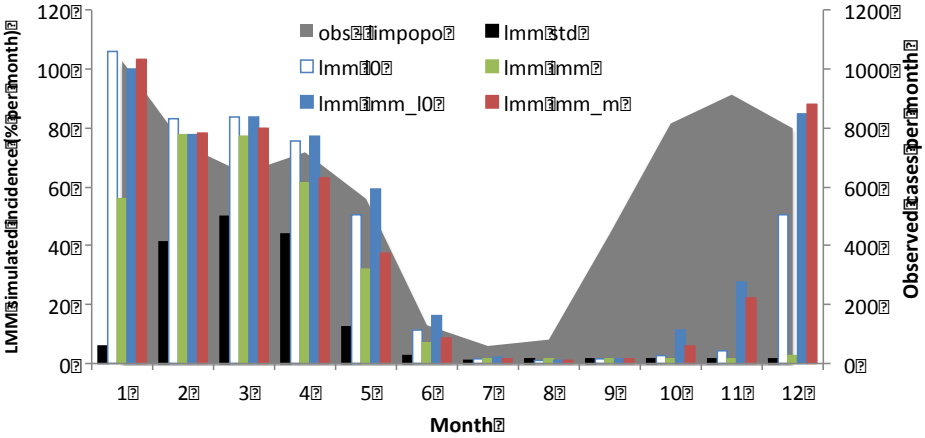
The observed seasonal malaria begins to increase in September, before the rainfall for Polokwane. The corresponding seasonal cycle in simulated incidence is delayed with respect to the cycle of observed malaria cases. For the original LMM parameter setting and that with only an adjusted human reservoir of infection (lmm std and lmm imm in Figure 2.1b), the peak malaria occurs in February/March, three to four months after the first observed peak in cases (and the first peak in TRMM rainfall). This delay is reduced for the other three parameter settings which all peak in January. These settings also exhibit a small secondary peak in March. A secondary peak in observed cases occurs in April. The end of the season in May/June seems well-represented in the model simulations, where temperature below 18 °C prevent development of the parasite within the mosquito population. Observed variability is fairly uniform over the season (Figure 2.1c), whereas model variability tends to be a maximum at the start and end of the season.

Modelled interannual variability in malaria incidence does not correspond to that found in the observations (Figure 2.2e). The observed seasonal totals of malaria cases (where the season is defined to run from August to July) show the lowest number of cases for the 2004/2005 season (Figure 2.2e) despite the highest number of degree days above 18 °C occurring during that year (Figure 2.2d), resulting in the year of highest malaria incidence for all model settings (Figure 2.2e). The LMM response is also high for 1999, when seasonal total rainfall is highest, and 2001 when SON rainfall is highest, but this is not reflected in the observations. The observed peak year in observed cases corresponds to only moderate rainfall and temperature conditions, and only a moderate response in LMM-simulated incidence.

a)



b)



c)

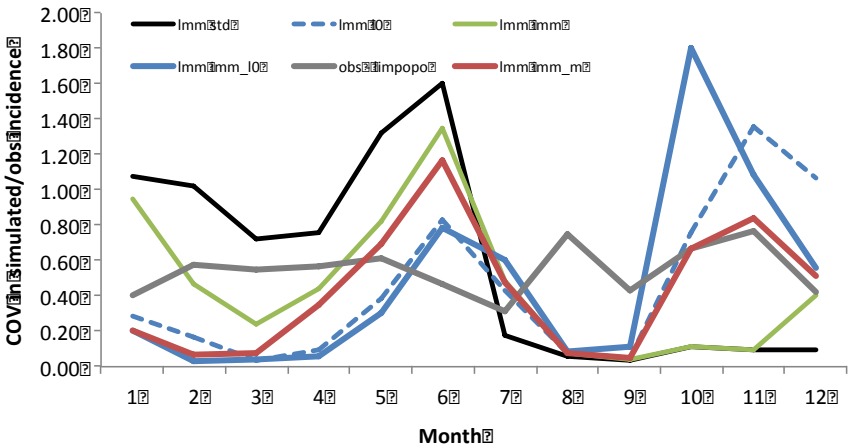
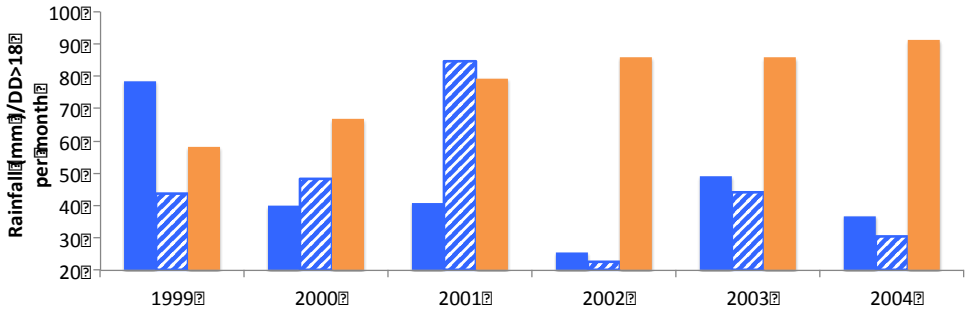


Figure 2.1 a) Mean seasonal cycles for rainfall and temperature for Polokwane, South Africa, for the period 1998-2008 (TRMM rainfall and ERA Interim temperature). b) Mean LMM seasonal cycle for the same period for multiple parameter settings, compared to seasonal cycle in observed cases for Limpopo province for 1999-2005. c) Coefficient of variation in LMM simulated incidence compared to that for observations for the same periods as a).

d)



e)

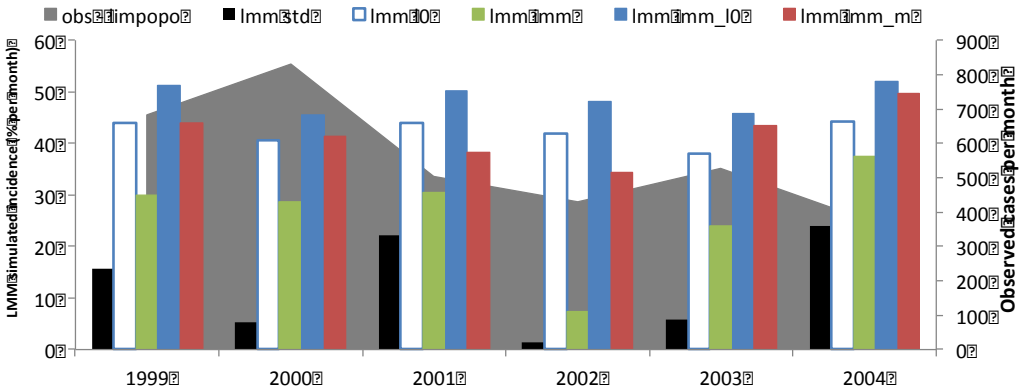


Figure 2.1 contd. Interannual variability in d) climate variables used to drive LMM (August to July rainfall: solid blue bars; SON rainfall: hatched blue bars; degree days above 18 °C: solid orange bars) and e) simulated LMM incidence compared to observations for Limpopo province.

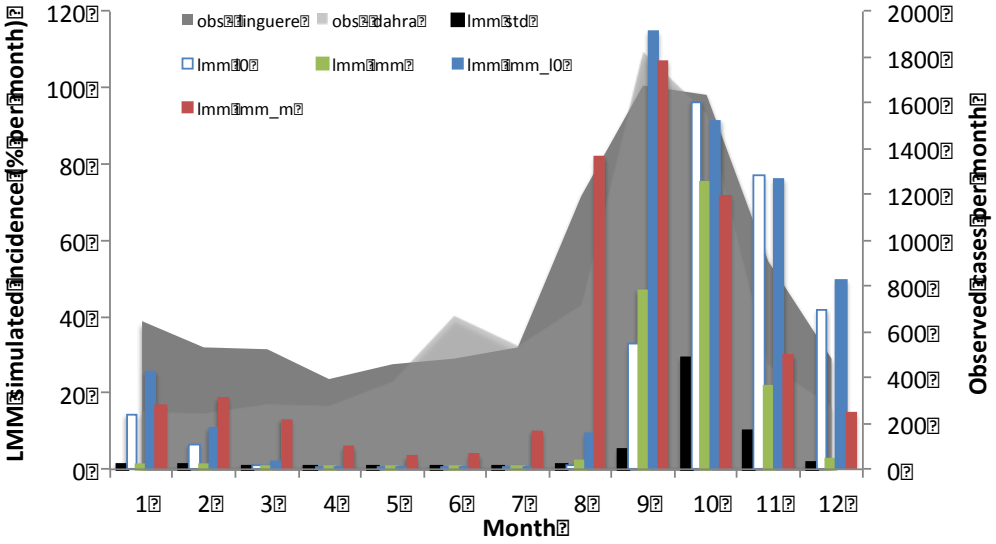
Louga Region, Senegal

Plots of the seasonal cycle for different LMM parameter settings are given in Figure 2.2a, with the corresponding variability per month shown in Figure 2.2b. Compared to the observed cycles, which peak in September, the peak of LMM-simulated incidence with standard settings (lmm std) is delayed by one month. The magnitude of model response in September is improved by including an immune category of transmission in the human population (lmm imm), although the peak remains in October. The addition of a more permissive larval setting (lmm imm_l0) succeeds in moving the peak to September, but the simulated decline in incidence towards the end of the season for such a setting is too slow, extending into November and December while the observed cases drop off rapidly after October. The closest fit to the observed cycle is obtained by adjusting both the immune category and introducing a influx of uninfected mosquitoes (lmm imm_m; red bars), and this setting is the only one for which there is dry season transmission, as evident in the observed data.

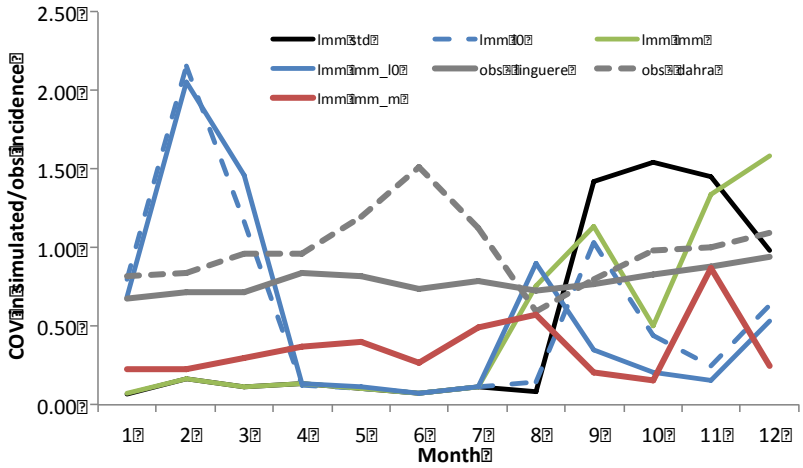
One impact of the lmm imm_m parameter setting is a reduction of variability in the peak months of September and October (Figure 2.2b) below that found in the observations. No model settings reproduce the June peak in variability during the dry season evident in the observations for Dahra; however this value is skewed by the single very high incidence of 3000 cases recorded for Dahra in June 2005 (not shown).

Analysis of interannual variability in seasonal totals of malaria cases (where the season is defined as the 12 months starting in May - the first month after the minimum in the observed seasonal cycle for Linguère) reveals a negative trend in cases over the later data period, with lower numbers of cases in 2008 and 2009 (and 2007 for Linguère) compared to previous years (Figure 2.2c). None of the LMM parameter settings reproduce this effect. Categorising the six remaining years (2001-2006) into two groups of the three highest and three lowest years, all of the LMM parameter settings correctly bin at least two out of the three years in each group, as measured against the observed cases for Linguère. The imm_m setting correctly assigns all three highest years of 2005, 2003 and 2001, although the difference in mean incidence values between 2001 and 2004 is only marginal at 0.085 cases per 100 people per month.

a)



b)



c)

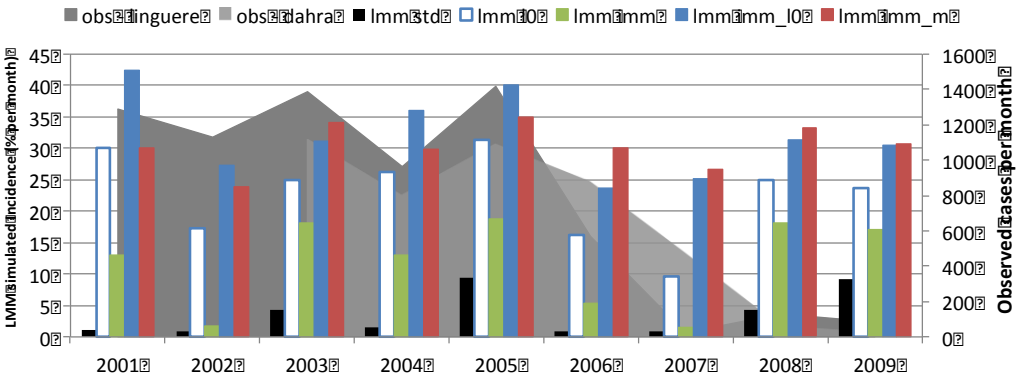


Figure 2.2 Effect of different model parameter settings on a) Mean LMM seasonal cycle and b) Coefficient of variation in LMM incidence for the period 1998-2010 compared to observed cases for Dahra (2003-2010) and Linguère , Senegal (2001-2010), and c) Interannual variability for LMM-simulated and observed incidence for a May to April malaria seasons. ERA-Interim temperatures together with TRMM rainfall were used to drive LMM for the nearest grid point (15.12°N,

15.12°W). The LMM parameter settings are: lmm std: standard settings, lmm l0: larval stage adjusted, lmm imm: immune category adjusted, lmm imm_l0: both larval stage and immune category adjusted, lmm imm_m both immune category and uninfected mosquito influx adjusted

Summary and conclusions

In Senegal, the observed seasonal cycle in malaria transmission could only be simulated in LMM by incorporating both a human reservoir of infection at the start of the season and an influx of mosquitoes, for example from permanent water bodies. LMM did not simulate the observed variability during the dry season, nor the decline in malaria recorded during the later data period, suggesting these effects lie outside the natural climate processes representing by the model, and which in the latter case may be due to non-environmental factors (no data adjustment has been made for human interventions, for example). The human reservoir/mosquito influx setting also provided the marginally closest match in terms of the categorisation of the remaining years between high and low malaria categories, although all LMM settings produced similar results in terms of interannual variability.

In South Africa, The simulated seasonal cycle was delayed with respect to observations. The delay could be reduced to two months by introducing a reservoir of infection in the human population in combination with either in influx of uninfected mosquitoes or a reduced sensitivity of larval mosquito survival probability to rainfall. The apparently rapid response of the observed cases to temperatures only marginally above 18°C and before the onset of the TRMM rainfall may indicate that the climate data used here does not adequately represent the climate experienced by the population susceptible to malaria, or that there is sufficient permanent mosquito habitat (beyond what has been simulated here by the lmm_m parameter setting) for transmission to occur as soon as temperatures rise above the sporogonic threshold. It may also indicate that the standard model threshold of 18°C is too high. The LMM simulations were not able to reproduce the observed interannual variability in seasonal totals of malaria cases.

The new LMM components trialled in this study resulted in an improved ability of the model to represent the observed seasonal cycle in transmission. Adjustment of the larval scheme improved the response at the start of the season, but resulted in a delay to the end of the season for the primarily rainfall-driven region in Senegal. In South Africa further analysis is required to determine whether the discrepancies between observed and simulated malaria is due to the climate data used to drive the model, or the model processes themselves.

References

Dee, D. P et al., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc*., **137**: 553–597. doi: 10.1002/qj.828

Drakeley, C., C. Sutherland, J. T. Bousema, R. W. Sauerwein, and G. A. T. Targett, 2006. The epidemiology of Plasmodium falciparum gametocytes: weapons of mass dispersion. *Trends in Parasitology*, **22**:424-430.

Ermert V, Fink AH, Jones AE, Morse AP. 2011a. A new version of the Liverpool Malaria Model. I. Review of the parameter setting and model structures. Malar J, 10:35; doi:10.1186/1475-2875-10-35.

Ermert V, Fink AH, Jones AE, Morse AP. 2011b. A new version of the Liverpool Malaria Model. II. Calibration

Gething P.W., Patil A.P., Smith D.L., Elyazar, J.R.F., G.L. Tatema and S.I. Hay, 2011. A new world malaria map: Plasmodium falcipraum endemicity in 2010. *Malaria Journal*, **10**:378

Hoshen M.B. and A. P. Morse, **2004**. A weather-driven model of malaria transmission. *Malaria Journal,* **3**:32, doi:10.1186/1475-2875-3-32.

Huffman, G.J., R.F. Adler, M. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B McGavock, J. Susskind, 2001. Global Precipitation at One-Degree Daily Resolution from Multi-Satellite Observations. *J. Hydrometeor*., **2**(1): 36-50.

Jones A., 2007. Seasonal ensemble prediction of malaria in Africa, University of Liverpool: Liverpool PhD Thesis.

Jones, A. and A.P. Morse, 2010. Application and validation of a seasonal ensemble prediction system using a dynamic malaria model. *Journal of Climate*, **23**:4202–4215

Jones, A.E. and Morse A.P., 2012. Skill of ENSEMBLES seasonal re-forecasts for epidemic malaria predictions in West Africa. *Geophysical Research Letters***, 39:**5, L23707

Kalney E. et al., 1996. The NCEP/NCAR 40-Year Reanalysis project. *Bull of Amer. Met. Soc.* **17**(3): 437-471.

Liu L, et al., 2012. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. The Lancet. 9832:2151-2161.

McKenzie FE, WH Bossert, 2005. An integrated model of Plasmodium falciparum dynamics. *Journal of Theoretical Biology* **232** (3): 411-426

Morse A.P., Doblas-Reyes F., M.B. Hagedorn and Palmer T.N., 2005. A forecast quality assessment of end to end probabilistic multi-model seasonal forecast system using a malaria model. Tellus A 57(3):464-475.

Nájera, J. A., Kouznetsov, R. L., and C. Delacollette, 1998. *Malaria epidemics: Detection and*

control, forecasting and prevention. WHO/MAL/98.1084. World Health Organization, Geneva, Switzerland. [http://www.who.int/malaria/docs/najera_epidemics/naj_toc.htm]

Paul, R.E., M. Diallo, and P. T. Brey, 2004. Mosquitoes and transmission of malaria parasites - not just vectors. *Malaria Journal*, **3**:39. doi 10.1186/1475-2875-3-39.

Tanser F.C., B. Sharp, D. Le Sueur, 2003. Potential effect of climate change on malaria transmission in Africa. *Lancet* **362**: 1792-1798.

WHO, 2012. World Malaria Report 2012. World Health Organization. ISBN 978 92 4 156453 3.

[B] Validation of an integrated weather - malaria model (REMO-LMM$_{2010}$-S$_{2005}$)

Overview

Climate exerts a strong influence on the spread and transmission intensity of malaria in Africa. In this study, the performance of an integrated weather–disease model is assessed.

Simulated were mosquito biting rates using the *2010 version of the Liverpool Malaria Model* (LMM$_{2010}$). The input data for the LMM were bias-corrected temperature and precipitation data from the *Regional Model* (REMO) on a 0.5° latitude–longitude grid. A *Plasmodium falciparum infection model* (S$_{2005}$ model) expands the LMM simulations to incorporate information on the malaria parasite infection rate among children. Malaria simulations were carried out with this integrated weather–disease model (using the following model chain: REMO-LMM2010-S$_{2005}$) for 1960 to 2000.

The simulated 41-year parasite ratios of children of the integrated weather-disease malaria model were quantitatively compared with the predicted spatial distribution of *P. falciparum* malaria endemicity of 2007 from the *Malaria Atlas Project* (MAP). The geographic malaria extent of the model chain is comparable to that of the MAP analysis. Most differences regarding the values of parasite ratios vanish when the uncertainty of the MAP model is considered. Differences are found for the north- eastern part of Somalia, where the integrated weather–disease model underestimates the malaria occurrence. The model overestimates the infection rate of children in parts of Senegal, Chad, Sudan, Ethiopia, and Kenya, which is likely attributed to nonmeteorological factors such as malaria control.

Model description and assumptions

The *P. falciparum* infection model. A nonlinear relationship exists between the *entomological inoculation rate* (EIR) and *P. falciparum* infection in children (Smith et al. 2005). Smith et al. (2005) fitted various mathematical functions to 119 published paired *annuell EIR* (EIR$_a$) and PR$_{<15}$ observations from Africa (Hay et al. 2005) by the maximum likelihood method. The best-fitting model (the S$_{2005}$ model) assumed heterogeneous infection rates and no immunity to reinfection but includes superinfection (i.e., an infection that follows a previous infection):

$$PR_{<15} = 1 - \left(1 + \frac{b \times EIR_a}{r \times k}\right)^{-k}$$

where b represents the transmission efficiency (the probability that an infectious mosquito bite causes infection) and $1/r$ is the expected time for parasite clearance. The S$_{2005}$ model assumes heterogeneous infection rates following a □□(gamma) distribution, with a mean of 1 and variance $1/k$. The estimated parameters were b/r = 0.45, $1/k$ = 4.2. Note that the S$_{2005}$ model reveals a large uncertainty because of the large variability of the observations (Smith et al. 2005, see their figure One).

The Liverpool Malaria Model (LMM). The LMM is a weather-driven, mathematical–biological model of malaria that was originally formulated by Hoshen and Morse (2004). It simulates daily malaria transmission and infection rates based on daily mean temperature

and 10-day accu- mulated precipitation. However, the model does not include some detailed aspects of the malaria infection, for example, the immune status of humans.

Ermert et al. (2011a, 2011b) constructed the improved *2010 version of the Liverpool Malaria Model* (LMM_{2010}) to include structural changes. For example, they parameterized hydrological conditions by a fuzzy distribution model (Ermert et al. 2011a), which considers the flushing of breeding habitats by excessive precipitation. Their study calibrated and extensively validated the LMM_{2010} by numerous field observations from West Africa in terms of 11 entomological and parasitological variables (Ermert et al. 2011b). They measured, using a problem-adapted skill score, the ability of the model to capture the range of observations in the vicinity of weather stations. Their study provided evidence that the LMM_{2010} simulates realistic EIR_a values and reasonable malaria seasons. The results show that the model reproduces the strong observed interannual variability of EIR_a. The performance is somewhat weaker regarding parasitological variables. For this reason, the S_{2005} model was applied for the simulation of *P. falciparum* infection rates. The EIR_a values were passed from the LMM_{2010} to the S_{2005} model to produce meaningful $PR_{<15}$ values.

Model Input dataset and disease observation dataset

Regional model (REMO) climate simulations. The malaria model chain is driven by temperature and precipitation data from the *Regional Model* (REMO), which is a limited-area regional climate model with a horizontal grid resolution of 0.5°. An ensemble of three REMO integrations was produced for the historical period (1960-2000) to take into account the internal variability of the model. In this case, REMO was forced by the observed greenhouse gas emissions. It is found that REMO simulates a reasonable climatological pattern of annual rainfall in Africa (Paeth et al. 2009). The highest precipitation amounts are simulated for the Congo basin and at the luvward sites of mountain ranges such as the Guinean Mountains (~ 8°N, 12°W).

Bias correction. Climate models are subject to biases that, when they are systematic in nature, can be compensated for by appropriate statistical methods. Therefore, the simulated weather data was corrected to ensure realistic LMM_{2010} input data using climatological differences between the REMO data and observed monthly rainfall from the *Climatic Research Unit* (version CRU TS 1.1; University of East Anglia, Norwich, UK), as well as daily temperatures from ERA40 (*European Centre for Medium-Range Weather Forecasts Re-analysis, 40 years*).

The integrated weather-disease model. The LMM_{2010} was used for the simulation of the spread of malaria under past weather conditions. Three LMM_{2010} runs were simulated on a 0.5° grid by daily temperatures and rainfall amounts from different REMO integrations of the present-day climate of 1960-2000. Subsequently, the S_{2005} model integrations were performed by EIR_a values of single years from the LMM_{2010} runs. The model chain (REMO-LMM_{2010}-S_{2005}) is denoted as the integrated weather-disease model (see Figure 1), which output is quantified in this study by data from the Malaria Atlas Project (MAP; Hay et al. 2009).

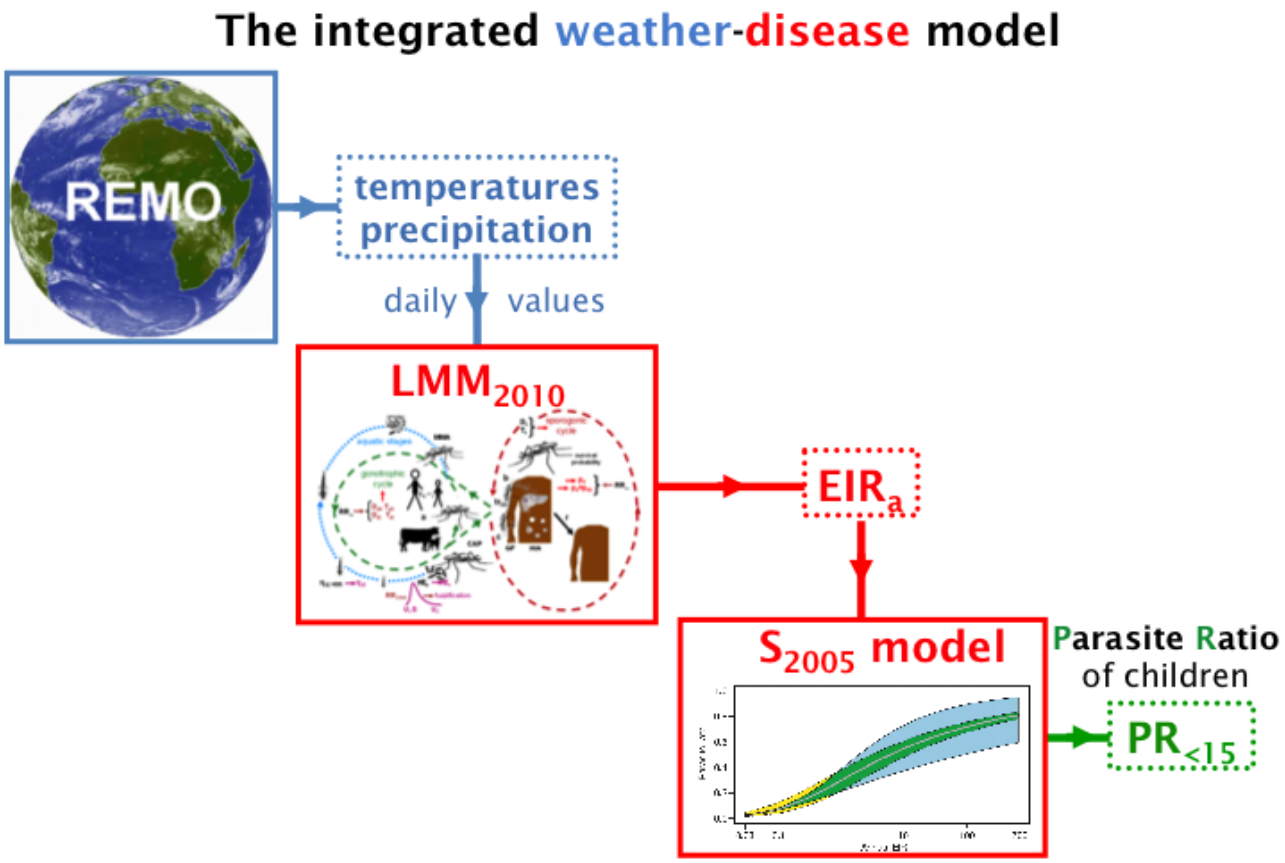


Figure 1. *Illustration of the model chain of REMO, LMM_{2010}, and the S_{2005} model, which is denoted as the integrated weather-disease model. REMO provides the daily temperature and precipitation values to the LMM_{2010} and the LMM_{2010} supplies EIR_a values to the S_{2005} model.*

Performance Assessment

The LMM_{2010} was calibrated by the usage of data from 34 West African weather stations and malaria observations from field sites in their vicinity. The main result of this calibration was that the LMM_{2010} is able to reproduce EIR_a in the range of field observations. In addition, the S_{2005} model was built by using paired EIR_a and $PR_{<15}$ values, which should ensure the calculation of realistic infection rates of children. It is true that the simulated two-dimensional malaria distribution lacks a quantitative verification, which is provided by this study.

The integrated weather-disease model is compared with an existing modelled malaria map. This provides a comparison between the simulated $PR_{<15}$ values from the S_{2005} model and that of the predicted *Plasmodium falciparum parasite ratio of children between 2 and 10 years* ($PfPR_{2-10}$) from the *Malaria Atlas Project* (MAP; Hay et al. 2009). Note that this comparison is somewhat problematic since: (i) The map of malaria endemicity of MAP was produced for 2007 and the integrated weather-disease model represents average values with regard to 1960-2000. Due to the fact that there is a strong interannual variability, parasite ratios of single years should normally not been compared to climatological values (here: 1960-2000). Hay et al. (2009) found, for example, that the observations from 2007 were substantially lower than the data from the other considered years (1985-2006). In Senegal, the malaria endemicity was, for instance, much higher in the 1960s than after the following drought conditions, when a substantial malaria decline was observed (e.g. Faye et al. 1995). (ii) Hay et

al. (2009) computed the parasite ratio of children between 2 and 10 years, whereas the S_{2005} model is constructed for children under the age of 15 years. Due to the increase of immune individuals with age and the simultaneous decline in the infection rate (e.g. Bekessy et al. 1976), the values from the S_{2005} model should be somewhat lower than that of the MAP model. Despite these two issues the comparison of the two maps enables a further insight in terms of the skill of our integrated weather-disease model.

In order to compare the two maps, the MAP data was aggregated from the 5 km x 5 km latitude-longitude grid to the resolution of the integrated weather-disease model of 0.5°. Differences between the two maps cannot be avoided at the fringe of the malaria area. For example, the malaria gaps (according to an asexual parasite ratio threshold of 0.1%) in the north-east of Somalia included in the MAP data are not captured by the aggregated data (cp. Figure 1b & c). In general, small gaps such as in the East African highlands are reduced by the 0.5° grid.

Results

Simulation of the geographic malaria distribution. The geographic distribution is compared between the integrated weather-disease model and MAP. According to the difference plot of the two maps (Figure 2d), two general statements can be made: (i) Both models show about the same geographic distribution of asexual parasite ratios. (ii) The integrated weather-disease model underestimates the territory of malaria endemicity in the Horn of Africa, especially for the north-eastern part of Somalia (see the crosses in Figure 2d). However, also the MAP model partly predicts gaps for North-Eastern Somalia (Figure 2b), which disappear in the 0.5° latitude-longitude resolution (Figure 2c). Beyond these two major aspects, small differences are found at the fringes of the endemicity area. The distribution of the integrated weather-disease model extends one to two degrees further to the north in various parts of the Sahel (see the dots in Figure 2d). This might be realistic since Hay et al. (2009) found higher parasite ratios for 1985-2006 than for 2007. It should be further noted, that malaria will probably be distributed further north under wet atmospheric conditions such as that before 1970. There seem to be also differences between the two maps for the East African highlands and the Adamawa plateau (see the crosses in Figure 2d). However, this again is an issue of the different grid resolution. The integrated weather-disease model, for instance, simulates a gap in the parasite ratios for the Adamawa plateau (Western Cameroon), which is evident in the full resolution, but vanishes in the 0.5° aggregated resolution (Figure 2a-c).

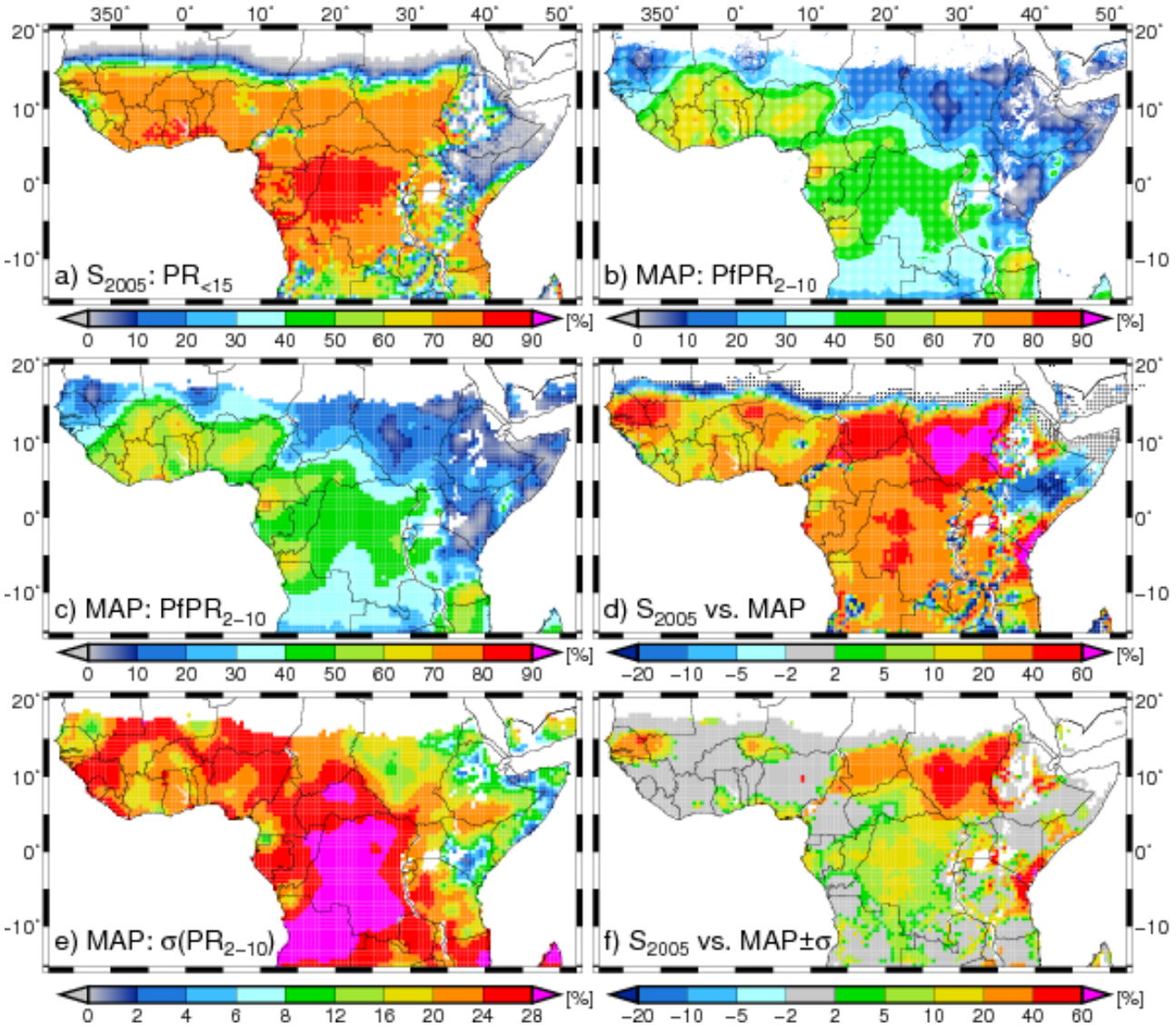


***Figure 2**. Validation of the integrated weather-disease model by the usage of the malaria endemicity map of the MAP (Hay et al. 2009): (a) S_{2005} model simulated $PR_{<15}$ values of 1960-2000. (b) MAP model predicted $PfPR_{2-10}$ values of 2007 on the original 5 km x 5 km latitude-longitude resolution. (c) Same as (b) but for a resolution of 0.5°. (d) Difference between the $PR_{<15}$ values of 1960-2000 from the S_{2005} model and the $PfPR_{2-10}$ values of 2007 from the MAP model. Dots indicate those areas where the S_{2005} model simulates $PR_{<15}$ values, which are higher than or equal to 0.1% and where the MAP model predicts $PfPR_{2-10}$ values lower than 0.1%. Crosses mark grid boxes where the MAP model predicts $PfPR_{2-10}$ values, which are higher than or equal to 0.1% and where the S_{2005} model simulates $PR_{<15}$ values lower than 0.1%. (e) Uncertainty (standard deviation) of the $PfPR_{2-10}$ values ($\sigma(PR_{2-10})$) of 2007 from the MAP model. (f) Same as (d) but including the uncertainty of the $PfPR_{2-10}$ values of 2007. In all maps those areas are excluded which show $PR_{<15}$ or $PfPR_{2-10}$ values below 0.1%.*

Quantitative verification. The parasite ratios between the integrated weather-disease model and MAP are now quantitatively compared using a difference plot. This comparison leads to further three statements: (i) The differences in the values of the parasite ratio disappear in most areas when the uncertainty of the MAP model is considered (Figure 2f). However, the integrated weather-disease model produces for most parts of tropical Africa higher parasite

ratios than the MAP model. (ii) The simulated parasite ratios of parts of Senegal, Chad, Sudan, Ethiopia, and Kenya from the integrated weather-disease model are too high. For these regions, the parasite ratios are 40% higher than predicted for 2007 by the MAP model and this difference is still present when the uncertainty of the MAP model is taken into account (Figure 2f). Either the integrated weather-disease model either fails to simulate lower parasite ratios in these areas or the values of 2007 are not representative for other years. It is most probable, that other factors than weather and climate conditions such as malaria control measures caused these comparable low $PfPR_{2-10}$ values for 2007. Hay et al. (2009) noted that the malaria endemicity was in general stronger in former years (see also Hay et al. 2009, their figure One). Note the present study does not account for other factors than weather and climate such as malaria control measures. (iii) The integrated weather-disease model simulates too low parasite ratios in low endemicity areas in comparison to the MAP model (Figure 2d). For example, in the northern Sahel the parasite ratios of the integrated weather-disease model are up to 20% lower than that of the MAP model (Figure 2d). It must be noted here that the MAP model overestimated the $PfPR_{2-10}$ values of low endemicity areas (Hay et al. 2009) and the difference totally disappears, when the uncertainty of the MAP model is taken into account (Figure 2f).

Conclusions

Parasite ratios from children as simulated by an integrated weather-disease model (REMO-LMM$_{2010}$-S$_{2005}$) were quantitatively compared with analysis data from the Malaria Atlas Project. Given the uncertainty in the MAP model and the constraints when comparing parasite ratios from a single year (here: 2007) with average values of a period (here: 1960-2000), it is valid to state that: The geographic malaria extent of the integrated weather-disease model is comparable to that of the MAP model. Most differences in terms of values of the parasite ratio vanish when the uncertainty of the MAP model is considered. Differences are found for the north-eastern part of Somalia, where the model underestimates the malaria occurrence. The integrated weather-disease model overestimates the malaria prevalence in parts of Senegal, Chad, Sudan, Ethiopia, and Kenya, which is likely attributed to factors such as malaria control.

References

Note that most text phrases of the current description were extracted word-by-word from Ermert et al. (2012), which was published in Environmental Health Perspectives.

Bekessy A, Molineaux L, Storey Y. 1976. Estimation of Incidence and Recovery rates of *Plasmodium falciparum* parasitaemia, from longitudinal data. *Bull World Health Org,* **54:**685-693.

Ermert V, Fink AH, Jones AE, Morse AP. 2011a. A new version of the Liverpool Malaria Model. I. Review of the parameter setting and model structures. *Malar J,* **10:**35; doi:10.1186/1475-2875-10-35.

Ermert V, Fink AH, Jones AE, Morse AP. 2011b. A new version of the Liverpool Malaria Model. II. Calibration and validation for West Africa. *Malar J,* **10:**62; doi: doi:10.1186/1475-2875-10-62.

Ermert V, Fink AH, Morse AP, Paeth H. 2012. The Impact of Regional Climate Change on Malaria Risk due to Greenhouse Forcing and Land-Use Changes in Tropical Africa. *Environ Health Perspect,* **120:**77-84; http://dx.doi.org/10.1289/ehp.1103681.

Faye O, Gaye O, Fontenille D, Hebrard G, Konaté L, Sy N et al. 1995. La sécheresse et la baisse du paludisme dans les Niayes du Sénégal. *Cahiers Santé,* **5:**299-305.

Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, et al. 2009. A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med,* **6:**e1000048; doi:10.1371/journal.pmed.1000048.

Hoshen MB, Morse AP. 2004. A weather-driven model of malaria transmission. *Malar J,* **3:**32; doi:10.1186/1475-2875- 3-32.

Paeth H, Born K, Girmes R, Podzun R, Jacob D. 2009. Regional climate change in tropical and northern Africa due to greenhouse forcing and land use changes. *J Clim,* **22:**114–132.

Smith DL, Dushoff J, Snow RW, Hay SI. 2005. The entomological inoculation rate and *Plasmodium falciparum* infection in African children. *Nature,* **438:**492–495.

[C] VECTRI - VECtor borne disease community model of ICTP, TRIeste.

Overview

VECTRI is a mathematical model for malaria transmission that accounts for the impact of temperature and rainfall variability on the development cycles of the malaria vector in its larval and adult stage, and also of the parasite itself. The majority of the relationships are taken from the literature for the *Anopheles gambiae* vector and the *Plasmodium falciparum* species of the parasite. Temperature affects the sporogonic and gonotrophic cycle development rates, as well as the mortality rates for adult vectors. Water temperature, closely related to air temperature impacts both the growth rate and mortality of larvae.

Rainfall effects on transmission are represented by a simple physically-based model of surface pool hydrology, where low rainfall rates increase available breeding sites that decay through evaporation and infiltration, while intense rain events decrease early stage larvae through flushing. A unique development for a regional scale malaria model is that VECTRI accounts for human population density in the calculation of biting rates. Higher population densities lead to a dilution effect resulting in lower parasite ratios (PR) in urban and peri-urban environments compared to nearby rural locations. In this respect the model is able to reproduce the reduction in Entomological inoculation rates (EIR) and PR with population density that has been widely observed in field observations in Africa.Future population growth could potentially reduce transmission intensity in VECTRI if included.

A schematic of the model approach is shown in the figure below

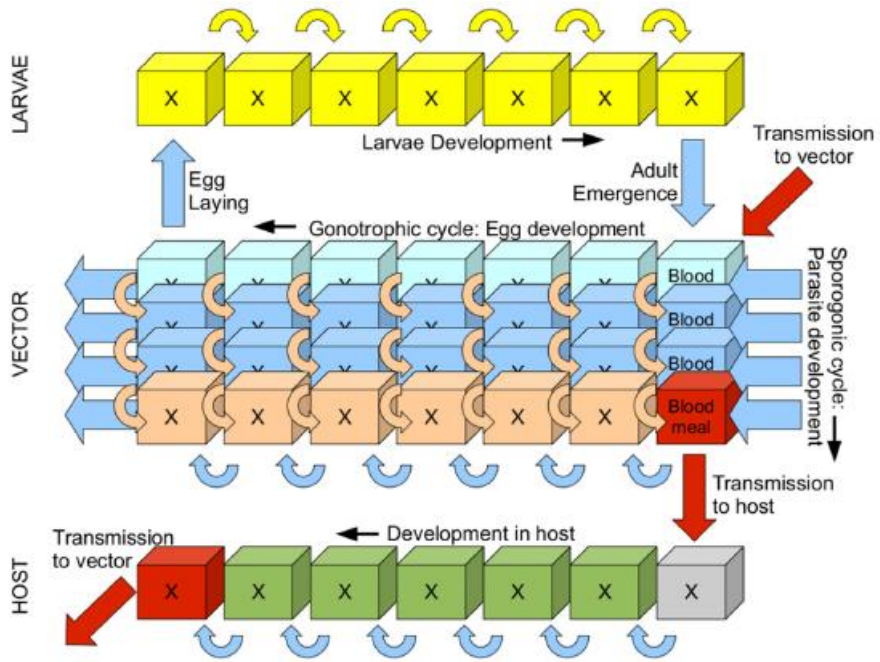


Figure 1 Schematic of the VECTRI model. The top row shows the larvae status divided into a series of bins representing the fractional development state. The 'X' in each bin represents the density of larvae in a specific fractional growth stage. The middle block represents the vector state in two dimensions: the egg development within the female and the infective state. The lower row models the host infective state. The curved arrows represent the progression direction of the larvae, vector and host state, while the straight red arrows mark the parasite transmission pathways between host and vector.

The model is designed for regional to continental scales at high spatial resolutions of up to 5-10km. For full details of the models mathematical framework and its evaluation, [Tompkins

and Ermert 2013].

Model Input dataset and disease observation dataset

The model has been operated and tested using the following types of input climate data for the following purposes:

Input Data	Purpose	Reference
>50 station data for single locations in W. Africa	Intercomparison with LMM2010 and evaluation of EIR and PR against survey data taken from the literature	Tompkins and Ermert (2013)
Present day integrations for West and East Africa domains using temperature downscaled from ERA-interim and rainfall from • TRMM • CMORPH • FEWS	Examine regional validity of the simulations in terms of mean spatial distribution of PR and EIR. Comparison made to the 2010 MAP (Malaria Atlas Project) maps	Tompkins and Ermert (2013)
re-Forecasts for period 1992 to present using ensemble forecasts from monthly EPS and system 4 seasonal forecasting system of ECMWF	Examine potential for an operational forecasting system for malaria early warnings. Reanalysis compared to MAP, while forecasts evaluated again malaria reanalysis in "tier 2" approach. In-country evaluation using case data in Malawi ongoing.	Tompkins and Di Giuseppe (2013) in preparation
Future climate projections using CMIP5 global model output	Future projections of malaria risk – not part of QWeCI project.	Kovats RS, Rocklov J, Caminade C, Tompkins AM, Morse AP, Jesus Colon-Gonzalez F, Stenlund H, Martens P, Lloyd SJ, 2013: Modelling the impact of climate change on malaria: a comparison of global malaria models, PNAS submitted Franziska Piontek et al. 2013: Leaving the world as we know it: Hotspots of global climate change impacts, PNAS submitted

Performance Assessment and results

Two example of a performance assessment are given here, reproduced from Tompkins and Ermert (2013)

a) Station assessment

Note that in the following, the LMM refers to the model version with modified parameter settings as given by Ermert et al. (2011) and referred to as LMM2010. A summary of the results for a selection of stations across the region (Figure below) the control model LMM2010 agrees well with the range of EIR values, which is not surprising, considering these data were used to calibrate the model settings. Although VECTRI was not calibrated with the data, at seven of the eight locations the rural VECTRI EIRa values overlap with the field observations mean- ing that the VECTRI runs produce realistic transmission values for most parts of West Africa. In comparison to the LMM2010, VECTRI shows a much smaller year-to-year variability and some field observations lie outside the range of the model interannual variability. To a certain extent, this is to be expected, since the model is only able to simulate the interannual variability due to climate - other factors such as interventions are neglected.

Instead, the LMM is calibrated to reflect all variability in the observations in its sensitivity to rainfall and temperature, as seen in the earlier equilibrium integrations. That said, VECTRI underestimates the malaria transmission in the northern Sahel. Too small transmission values are simulated by VECTRI in Podor for example. Transmission in urban areas appears to be under-simulated, and in two urban locations in Dakar no transmission at all occurs in the model. Further south, where more rainfall occurs during the monsoon season, VECTRI in general simulates somewhat too high EIRa values in comparison to the rural observations. VECTRI also seems to simulate too high ransmission values in equatorial Africa in Douala, which is subject to high annual rainfall rates. The simulated EIRa values exceeding 500 infectious bites per human per year are rarely observed in Africa with 13 surveys out of a total of 180 reported by [32] registering EIR in this range.

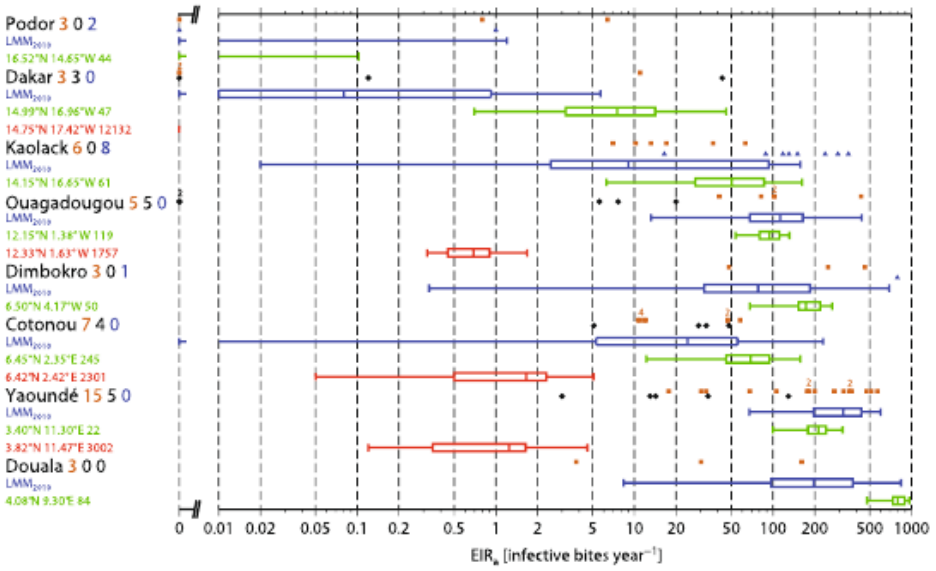


Figure 7 Observations and simulations of *EIR* and *CSPR*. Box-and-whisker plots of the VECTRI and LMM$_{2010}$ simulated annual Entomological Inoculation Rate (*EIR$_a$*; in infectious bites per human per year) for eight West African locations. The box-and-whisker plots of the LMM$_{2010}$ are illustrated in blue. VECTRI results in terms of rural and urban areas are taken green and red, respectively. Also the latitudinal, longitudinal position, and used population density (km^{-2}) is indicated. Climate varies from a semi-arid (Podor) to a tropical climate (Douala). Field *EIR$_a$* observations are furthermore included as brown squares, black dots and blue triangles in rural, urban and irrigated areas, respectively. The number of field observations is entered behind the location names in the following order: 1) rural, 2) urban, and 3) irrigated observations. The number of clustered observations on the logarithmic scale is indicated by a digit above the symbols.

A further detailed analysis is made for Bobo-Dioulasso in southwest Burkina Faso, which was chosen due to a particularly large number of field experiments with which to compare the

models. Malaria is seasonally endemic in this location and a number of field campaigns have sampled EIR and CSPR in both (peri) urban and rural (some with nearby irrigation schemes) environments [97-106].

Only one integration is conducted for the LMM 2010 since it is unable to account for population density. This is compared to two VECTRI integrations representative of rural (latitude 11.43°N, longitude 4.25°E, human population density H=32.2 km−2) and urban (latitude 11.2°N, longi- tude 4.30°E, H=1040 km−2) field locations. The timeseries of annual EIR and CSPR values from the 3 integrations (Figure 8) show that both models reproduce the correct order of magnitude for these two variables in the rural location. The EIR in the VECTRI model and LMM2010 lie within the spread of measured EIR. The significant variability between the observed EIR values is noted, which could derive from differences in terrain, topography, altitude, vicinity to water bodies, irrigation and land use, interventions and simply experimental sampling error. It emphasizes the uncertainty in individual measurements and the need for ensembles of experiments to gauge this uncertainty. The EIR is generally lower in rural locations with irrigation, possibly as a result of preditor establishment in longer-lived pools but may also be due to irrigation providing farmers with higher incomes permitting further prevention and treatment measures. This emphasizes the difficulty in gauging such influences in the w0 parameter of the VECTRI model.

Intercomparing EIR from the two malaria models, the very high interannual variability of the LMM2010 is obvious, related to its elevated sensitivity to rainfall demonstrated in the earlier constant-input experiments. The LMM2010 EIR value ranges by more than an order of magnitude between the lowest and highest year, while the interannual variability is less than a factor of two for the VECTRI model. It is interesting to note that, despite the basic underlying structure being very similar in the two models, the parametrization choices, in particular the implementation of the surface hydrology in the VECTRI model, results in almost a zero correlation between the two models in their representation of interannual variability (recalling that this is one particular parameter setting for the LMM and that the original LMM uses a different rainfall egg-laying relationship with a positive correlation); indeed, the EIR appear anti-correlated indicating that rainfall variability is determining the interannual variability to a large extent.

The black squares in the figure give EIR values for high population areas in peri-urban Bobo-Dioulasso, which are much smaller than in the rural environments but nevertheless non-zero. The LMM2010 is not designed to simulate these urban cells, however the VECTRI model is seen to reproduce reasonably well the contrast between rural and urban areas. Even though the treatment of the surface hydrology is identical in urban and rural environments - a gross oversimplification - the VECTRI model is able to mimic the drop in EIR which derives merely from the lower ratio of vector to host in urban areas. In contrast with the other West African locations, urban transmission is overestimated in this location. In the second panel, it is seen that while both models are again similar and perform well in reproducing the observed CSPR rates, the VECTRI model does less well in reproducing the distinction between rural and urban environments. Although the CSPR is lower in urban environments, it is still far larger than the observations in VECTRI. An possible implication is that vector lifetimes for urban areas are too long in the model.

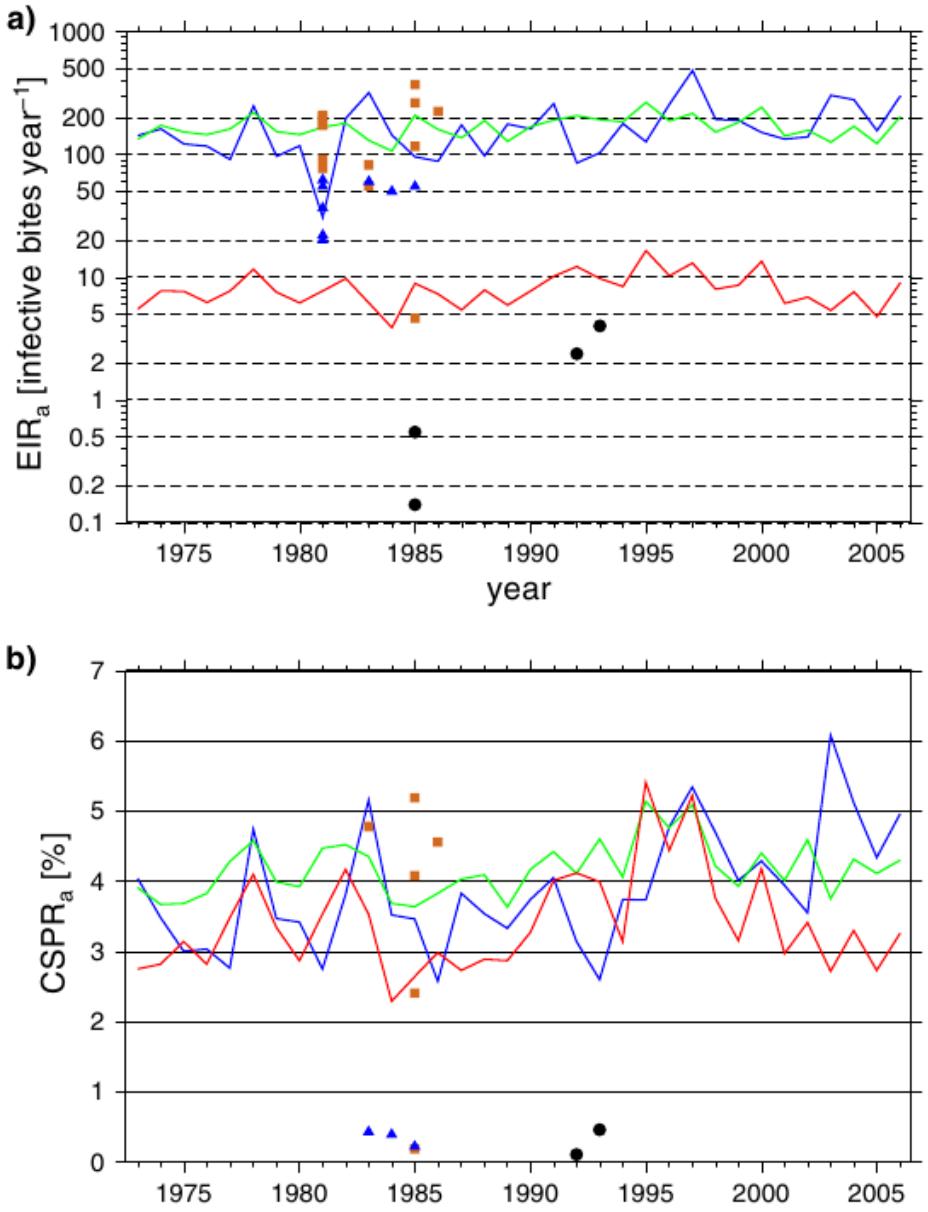


Figure 8 *EIR* and *CSPR* interannual variability in Bobo-Dioulasso. Simulated and observed malaria transmission in the area of Bobo-Dioulasso. (**a**) Annual *EIR* rates (EIR_a; in infectious bites per human per year) and (**b**) annual mean *CSPR* ($CSPR_a$; in %). The blue lines represent LMM_{2010} and the red and green lines VECTRI simulations (red: human population density $H = 1037$ km^{-2}; green: $H = 32.2$ km^{-2}). The brown squares, blue triangles, and black dots stand for observed EIR_a and $CSPR_a$ values in rural, irrigated and urban areas, respectively. See [30] for a definition of rural, irrigated and urban areas and for the sources of the field observations (their Additional file Two; the Word Meteorological Organisation (WMO) station number 65510 of Bobo-Dioulasso is assigned to the data).

b) Regional integrations

The strength of the VECTRI model is its ability to run on a regional scale at a relatively high horizontal resolution. The VECTRI integration made for Eastern Africa is com- pared to the MAP malaria analysis (Figure below). As this is a single deterministic integration of the malaria model it is simply compared to the mean PR of MAP. This simple comparison of mean PR distributions indicates that the model is able to reproduce the general patterns observed in the MAP analysis, with high rates in central and northeastern Uganda and Western Kenya, but dropping over the higher terrain, with the central Kenya, central Tanzania and most of Rwanda largely malaria-free (EIRa < 0.01, PR < 1%), as is the southwestern-most tip of Uganda.

The VECTRI model also reproduces the malaria zones in the warmer and more humid coastal regions in Kenya and Tanzania, but with values of PR along the Kenyan coast and northern Tanzania in VECTRI (30-70%) appear to greatly exceed those in the MAP analysis (0-30%). This is a region in which the survey data incorporated into the MAP analysis is dense, although still identified as relatively uncertain. A large part of the discrepancy between VECTRI and MAP is likely due to the increasing interventions including widespread distribution of insecticide-treated nets (ITNs) that have occurred over the past decade that have greatly decreased parasite ratios and hospital admissions. For example, the survey of 30 villages in Malindi, Kilifi, and Kwale Districts carried out in the late 1990s before ITN distribution started [119] (included in the MAP analysis) reported a parasite prevalence ranging from 38 to 83%, with a mean slightly exceeding 60% in each district, in close agreement with the VECTRI model. This highlights the importance of the future incorporations of interventions into VECTRI if it is to be applied to the seasonal forecasting task.

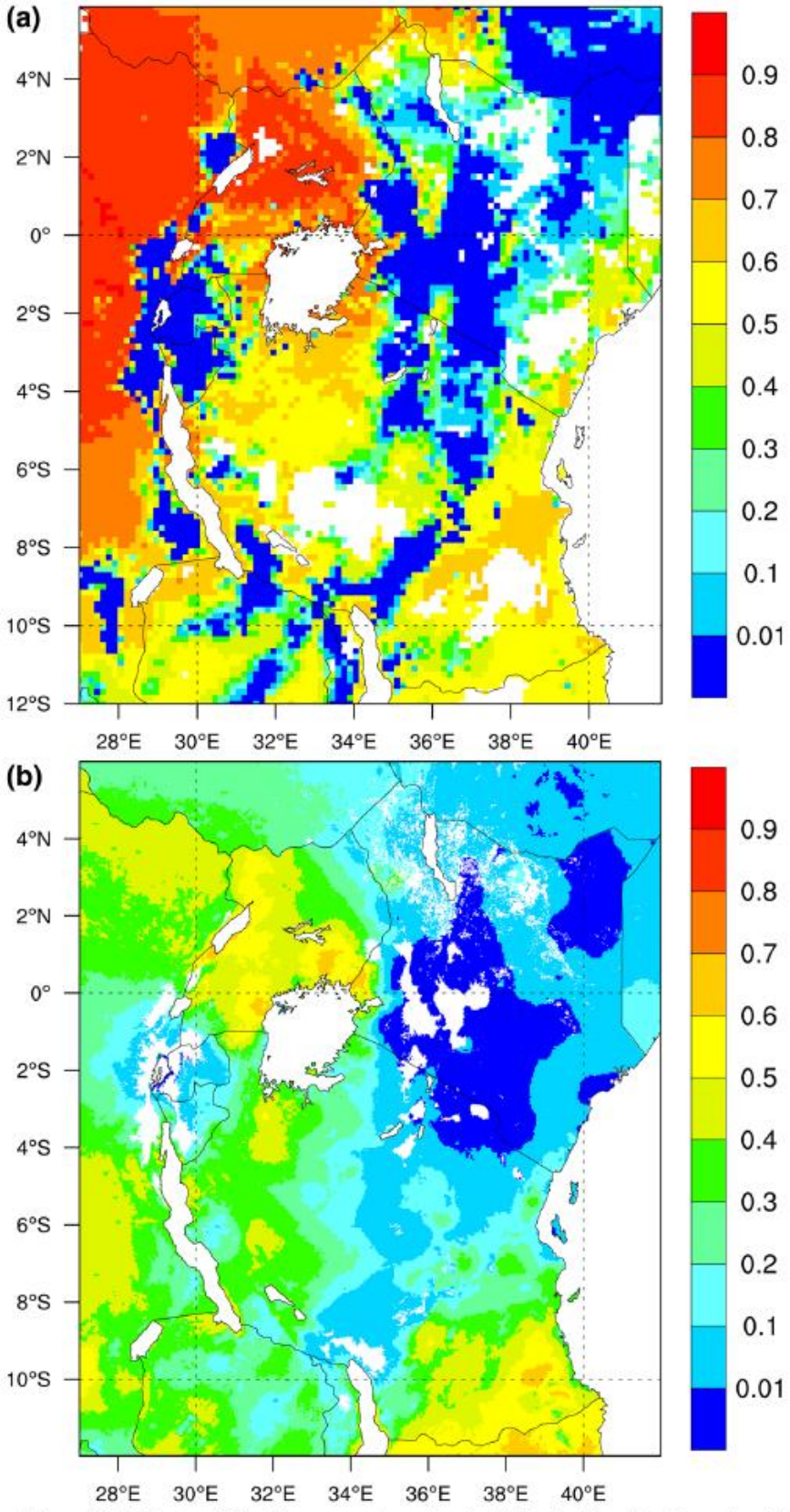


Figure 11 Regional simulation of *PR* in Eastern Africa. Mean asexual parasite ratio (*PR*; fraction) from (**a**) a 10 year-long VECTRI integration for East Africa and (**b**) MAP *PR* analysis. See text for details of model integration and data source.

Conclusions

A new dynamical community malaria model is publicly available that accounts for climate and population density to simulate malaria transmission on a regional scale. The model structure facilitates future development to incorporate migration, immunity and interventions. **It has been evaluated in a number of settings and forms the basis of the monthly to seasonal forecasting system that has been developed within the QWeCI project with partner ECMWF.**

While already representing a useful tool to aid research into the understanding of malaria transmission, the VECTRI is undergoing further development to incorporate further aspects important for malaria transmission. These developments include: a flexible incorporation of various models for host immunity, a stochastic ensemble framework to account for uncertainty in model parameters and observations, improved surface hydrology that accounts for terrain slope, soil type and land use characteristics, a dynamical host model that includes migration, population demographics and urbanisation, explicit incorporation of commonly employed interventions, and lastly multiple vector types and vector dispersion.

References

A. Tompkins and V. Ermert. 2013. A regional-scale high resolution dynamical malaria model that accounts for population density, climate and surface hydrology. *Malaria Journal,* **12:**65; doi:10.1186/1475-2875-12-65

Mbogo CM, Mwangangi JM, Nzovu J, Gu W, Yan G, Gunter JT, Swalm C, Keating J, Regens JL, Shililu JI, Githure JI, Beier JC. **2003**. Spatial and temporal heterogeneity of Anopheles mosquitoes and Plasmodium falciparum transmission along the Kenyan coast. *Am J Trop Med Hyg*, **68**:734–742.

[D] Mixed malaria model for Malawi

Overview

A spatio-temporal statistical model framework (Lowe et al., 2011, Lowe et al., 2012) has been extended to simulate malaria risk, as a function of climate and socio-economic drivers, in one of the QWeCI pilot sites: Malawi. Malaria transmission is influenced by variations in meteorological conditions, which impact the biology of the mosquito and the availability of breeding sites, but also socio-economic conditions such as levels of urbanisation, poverty and education, which influence human vulnerability and vector habitat. The many potential drivers of malaria, both extrinsic, such as climate, and intrinsic, such as population immunity are often difficult to disentangle. This presents a challenge for modelling of malaria risk in space and time.

Model Input dataset and disease observation dataset

Using an age-stratified spatio-temporal dataset of malaria cases from July 2004 – June 2011, a spatio-temporal modelling framework has been developed to explore variations in malaria risk in the 28 districts of Malawi. District level data is tested in the model to account for confounding factors, including the proportion of the population living in urban areas; residing in traditional housing; with no toilet facilities; who do not attend school, etc, the number of health facilities per population and yearly estimates of insecticide-treated mosquito net distribution.

Climatic and topographic variations are included by using an interpolation method to relate gridded products (e.g. CPC FEWS-Net rainfall estimates based on satellite and rain gauge data and ERA-Interim Reanalysis temperature data) to administrative districts. Figure 1 shows the relationships between malaria standardised morbidity ratios (SMR, the ratio of observed to expected malaria cases), climate variables are the most significant time lags (precipitation 1 month lag, temperature 3 month lag) and the relationship between precipitation and temperature. The solid line shows the linear model fit and the dashed curve shows the local polynomial regression fit.

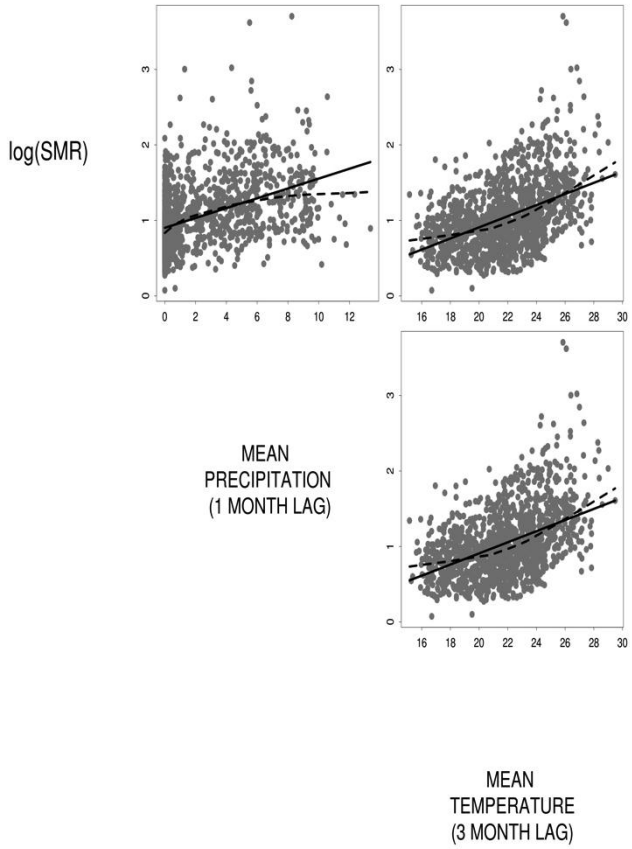


Figure 1*: Scatter plots to show the relation between malaria standardised morbidity ratios (SMR) and precipitation (1 month lag) and temperature (3 month lag) for the 28 districts of Malawi, July 2004-June 2011.*

Performance Assessment and Results

Using the exploratory variable outlined above, a generalised linear model framework was used to test and select spatial and temporal variables, factors, interactions and polynomial terms. Stepwise model selection was performed using Akaike Information Criterion (AIC). Categorical variables of importance were age group (over and under five years), region (north, central, south), zone (lowland, lake shore, highland and combinations), annual cycle. Important climate information was temperature (lag 3 months), precipitation (lagged 1 month with a quadratic association, see Fig. 1), interaction between temperature and precipitation (see Fig, 1). Non-climate information included altitude, longitude and latitude (quadratic relationships), demographic information: urbanisation, population density, housing conditions: one room for sleeping, no toilet facilitates, health facilities per population and education level.

In order to account for the unobserved confounding factors that influence malaria, which are not accounted for using measured covariates, a generalised linear mixed model (GLMM) is adopted, which includes structured and unstructured spatial and temporal random effects. A hierarchical Bayesian framework, using Markov chain Monte Carlo (MCMC) simulation, was used for model fitting and prediction. Using a stepwise model selection procedure, several explanatory variables were identified to have significant associations to malaria including climatic, cartographic, and socio-economic data. Once unobserved confounding factors and spatial correlation were considered in a Bayesian framework via spatial random effects (see

Figure 2), a final model emerged with statistically significant predictor variables limited to age group, an autocorrelated annual cycle, the number of health facilities per inhabitant, average precipitation (quadratic relation) and average temperature during the three months previous to the month of interest (see Figure 3).

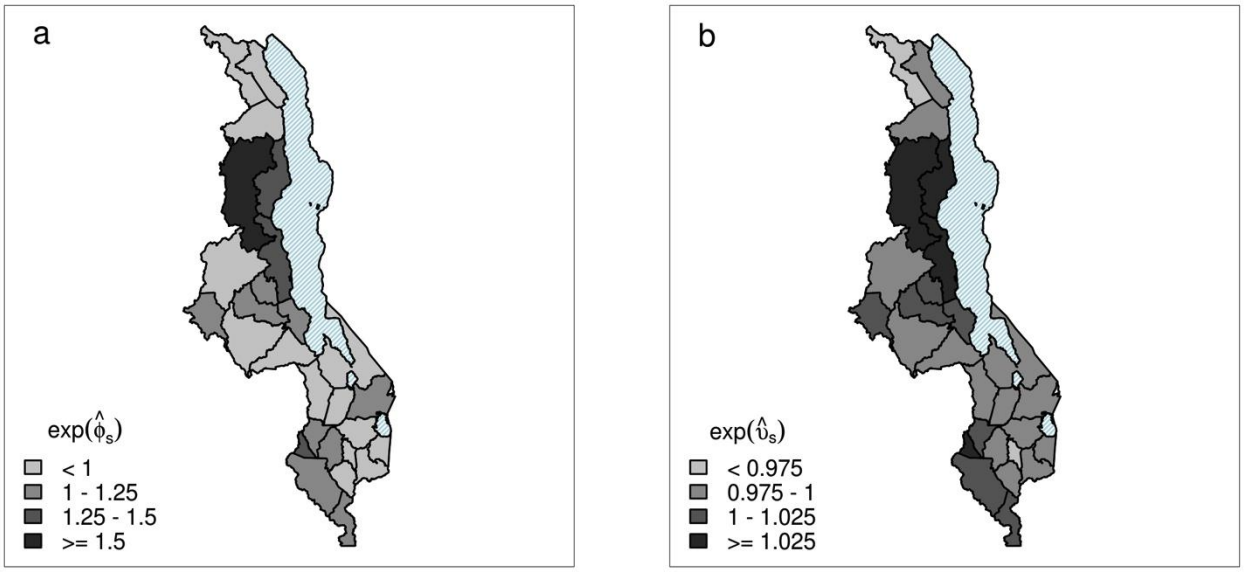


Figure 2. *Spatial distribution of multiplicative contribution of posterior mean spatially (a) unstructured (to account for heterogeneity between districts) and (b) structured (to account for correlation/clustering between districts) random effects.*

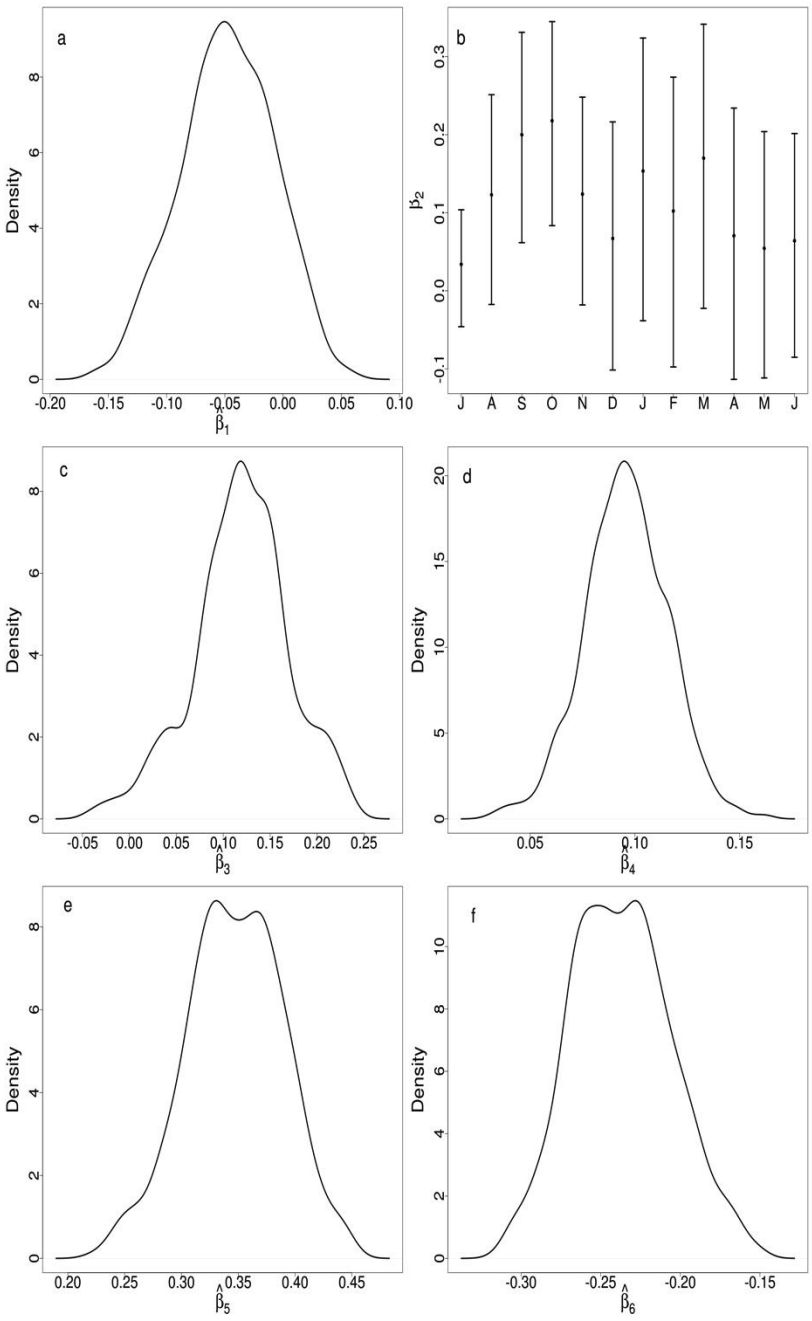


Figure 3. *Kernel density estimates for the marginal posterior distributions for the significant parameters estimates associated with the categorical variables (a) age group and (b) calendar month and (c) health facilities per inhabitant, (d) temperature, (e) precipitation and (f) precipitation squared.*

Figure 4 shows a summary of the model fit, divided into five years and over age group (top panel) and under five years age group (lower panel). The scatter plots and time series show the relation between model fit and observed malaria SMR for the whole of Malawi for the 84 month time period. Observed and fitted values appear to agree quite well. The annual cycle in malaria is well captured. Although the model does not well represent inter-annual variation in malaria, the temporally varying climate information does explain some of this variability. Over all, the model explains 41% of the variation in malaria risk. The maps show the root mean

squared error (RMSE) of the difference between model fit and observed SMR over the time period in each of the 28 districts, highlighting districts where the model performs less well. The feasibility of using dynamical disease model output to drive statistical models is in progress.

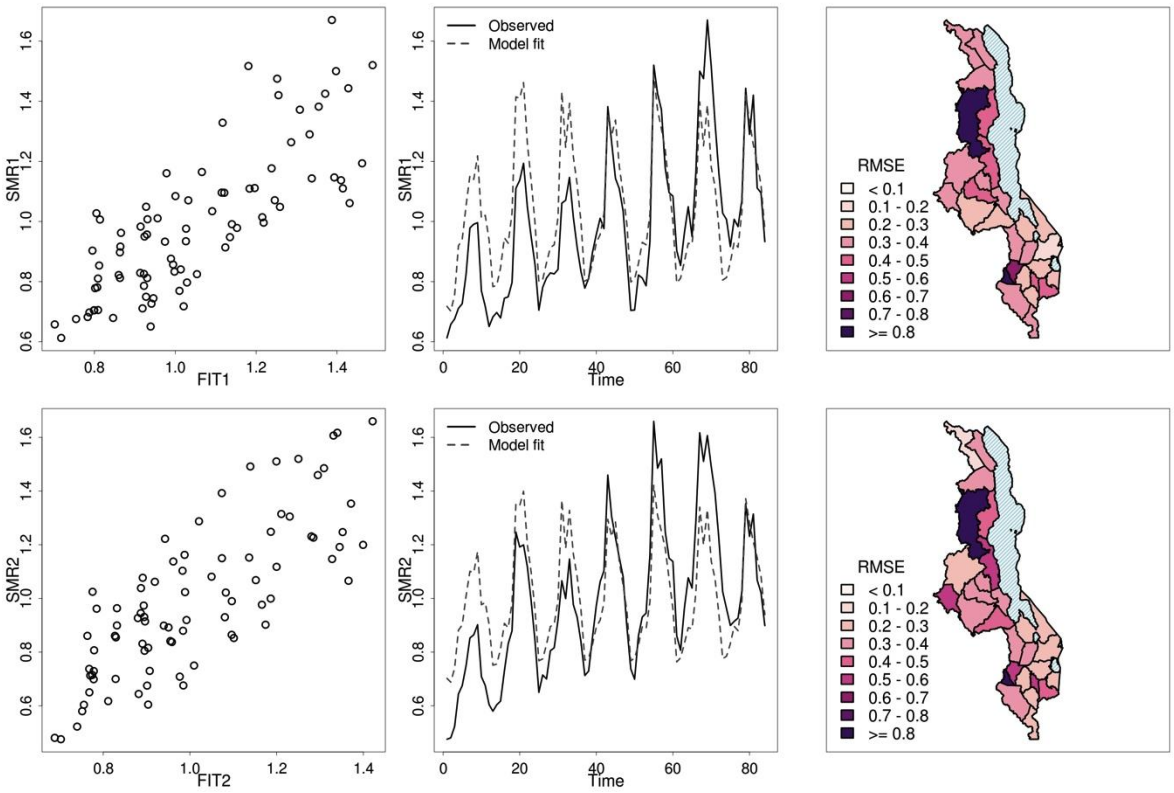


Figure 4*: Space averaged scatter plot (left) and time series (centre) of observed and model fit malaria SMR in Malawi for the 84 month time period. Map (right) of the root mean squared error (RMSE) between observed and model fit malaria SMR for each district in Malawi over the 84 month time period. Top panel: five years and over age group, lower panel: over 5 years age group.*

We plan to drive this statistical malaria model for Malawi using seasonal to decadal forecasts, in order to assess the ability of a forecasting system to predict variations in malaria seasons, from one year to the next, with a lead time of several months. This will require a thorough skill assessment of seasonal climate forecasts for this region of Africa.

Conclusions

When modelling malaria risk in Malawi it is important to account for spatial and temporal heterogeneity and correlation between districts. Once observed and unobserved confounding factors are allowed for, precipitation and temperature in the months prior to the malaria season of interest are found to significantly determine spatial and temporal variations of malaria incidence. Climate information was found to improve the estimation of malaria relative risk in in 41% of the districts in Malawi.

References

Lowe R., Chirombo J. Tompkins A. M. Relative importance of climatic and socio-economic determinants of malaria in Malawi, Malaria Journal (to be submitted).

Lowe, R., Bailey, T. C., Stephenson, D. B., Jupp, T. E., Graham, R. J., Barcellos, C. and Carvalho, M. S. 2013. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Statistics in Medicine* **32**: 864–883. doi: 10.1002/sim.5549.

Lowe R., Bailey T. C., Stephenson D. B., Graham R. J., Coelho C. A. S., Carvalho M. S., Barcellos C., 2011. Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Computers & Geosciences* **37**: 371-381.

[E] Statistical modelling of Rift Valley fever vectors abundance in a Sahelian area (Barkédji in Senegal)

Overview

A spatio-temporal model (Statistical Bayesian model) was developed in order to investigate the impact of climatic factors on Rift Valley fever (RVF) vector abundance and to predict the period as well as the areas with the highest vector productivity. Datasets on RVF vectors and environmental factors were generated as part of a longitudinal study conducted in 2005 every fortnight at 79 sites including temporary ponds, barren, shrubby savannah, wooded savannah, steppes, and villages at different distances (ranging between 0 and 600 m) from the main ponds of the Barkedji area (northern Senegal).

Our findings showed the importance of environmental conditions in predicting mosquito abundance. Relative humidity was positively correlated with the main RVF vectors abundance. Maximum temperature and rainfall were associated with the number of collected mosquitoes. The highest vector densities were observed around ground pools and neighboring sites. Improvements to the model have been carried out as part of this pilot study.

Model Input dataset and disease observation dataset

Maximum and minimum temperature, relative humidity, Normalized Difference Vegetation Index (NDVI), and cumulative rainfall were used to fit the Bayesian model. NDVI was derived from the Moderate Resolution Imaging Spectro-radiometer (MODIS) satellite (National Aeronautic and Space Administration, NASA). NDVI data was based on a 16-day average at a spatial resolution of 250 m. Temperature, rainfall and relative humidity were provided by the ground level weather stations managed by CSE (Centre de Suivi Ecologique) in the area. Cumulative rainfall 15-20 days prior to mosquito collection was chosen to take into account the cumulative effect of rainfall on ponds and consequently on vector productivity and abundance. Six landscape classes were defined using SPOT 5 satellite images. The method is based on a combination of two different studies (FAO, 1997 and Anon, 1956). Mosquitoes were collected bi-weekly using CDC light traps in 79 sites including all landscape classes.

Performance Assessment

The Rift Valley Fever vector model performance is assessed using the following metric:

- Comparison of the modelled versus observed vector abundance for all landscape classes.
- Generation of forecast maps during the vector abundance peak in order to provide information about the areas at risk

Results

Comparison between modelled and observed data in all landscape classes

For *Culex poicilipes*, except in the barren and in the steppe categories, the best model simulates the spatial variability of observed data in the other land cover classes (Figure 1.1). For *Aedes vexans*, the model reproduces the observed seasonality distribution for the different landscapes. All observed data were included in the prediction interval (Figure 1.2).

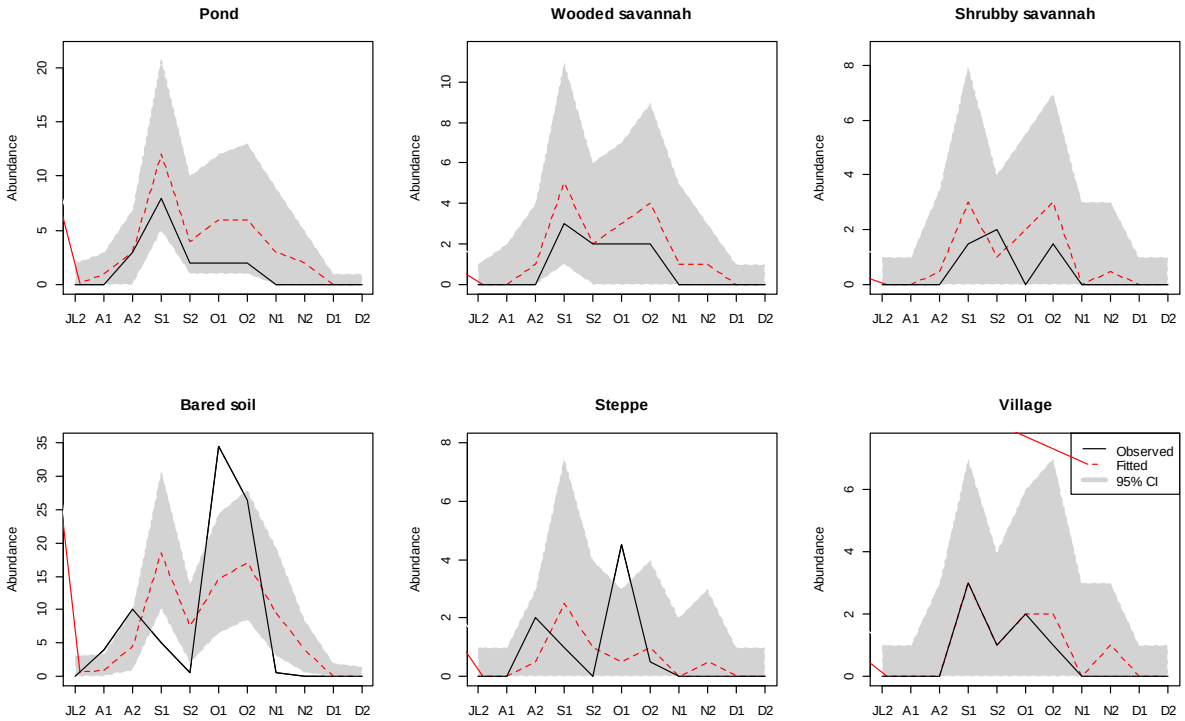


Figure 1.1: Observed (solid line) and predicted (dashed; shaded area=95%prediction interval) abundances for *Culex poicilipes* for the different landscapes.

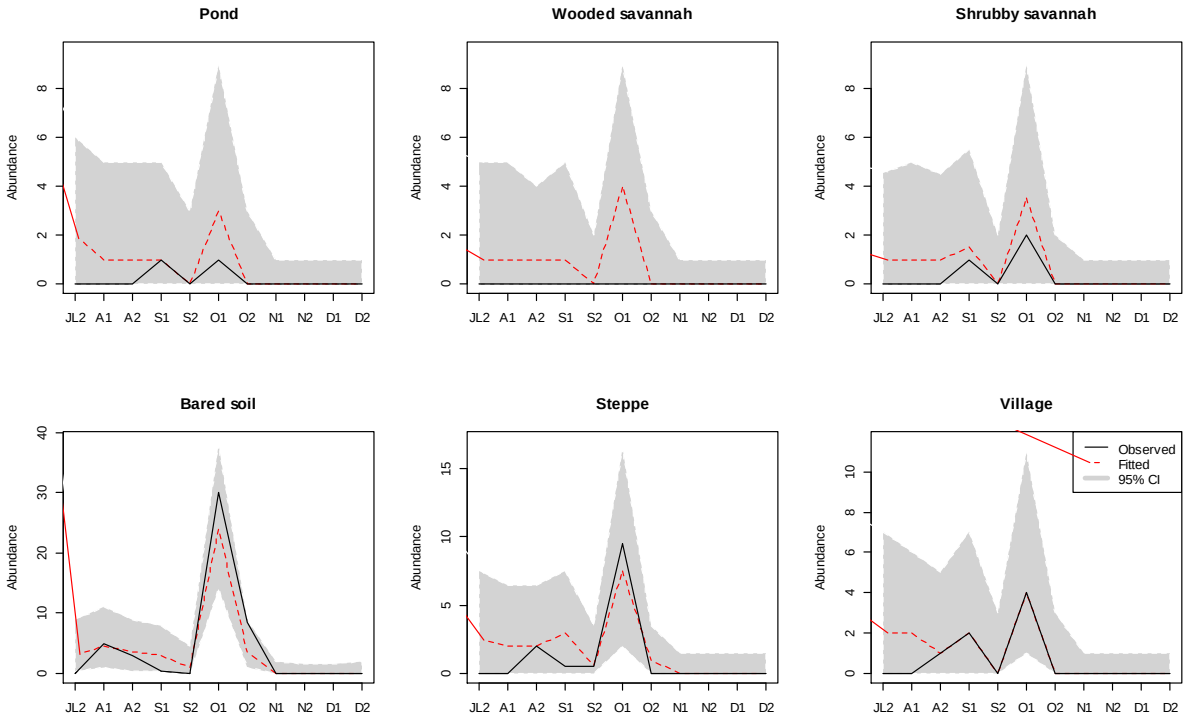


Figure 1.2: Observed data (solid line) and predicted (dashed; shaded area=95%prediction interval) for *Aedes vexans* for the different landscapes.

Comparison between modelled and observed data during the vector abundance peak

Forecast maps were generated during the peak in vector abundance. In September and October, maps predicted that the highest number of *Culex poicilipes* will be found near the Kangaledji and Niakha ponds (Figure 2.1). For *Aedes vexans* vector the highest abundance was simulated around the Niakha pond and in the northern part of the study area (Figure 2.2).

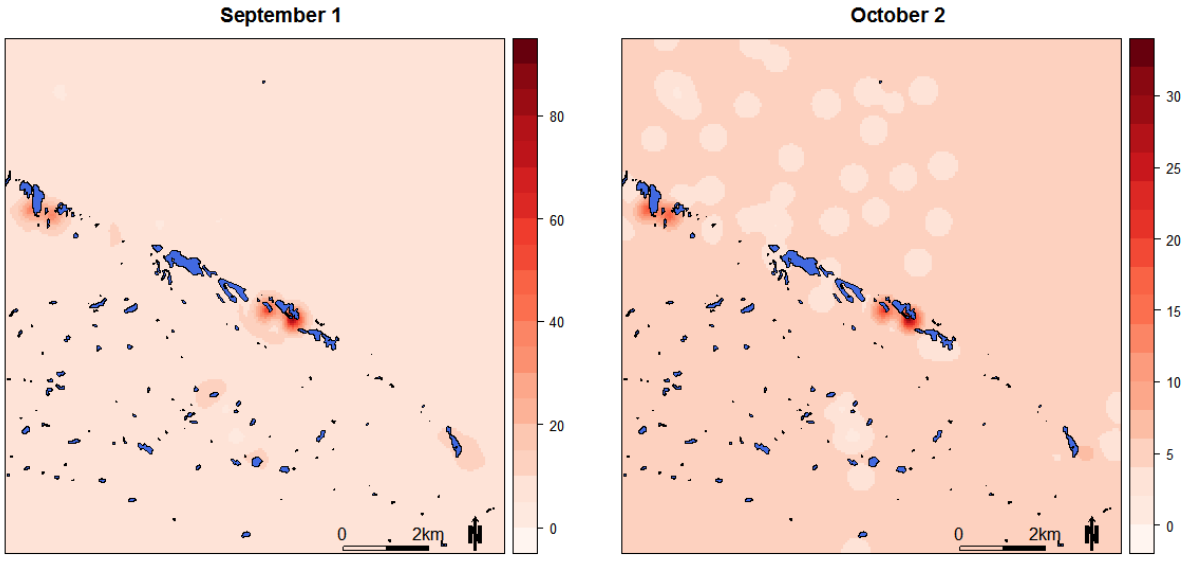


Figure 2.1: predicted maps during peak abundance for *Culex poicilipes*.

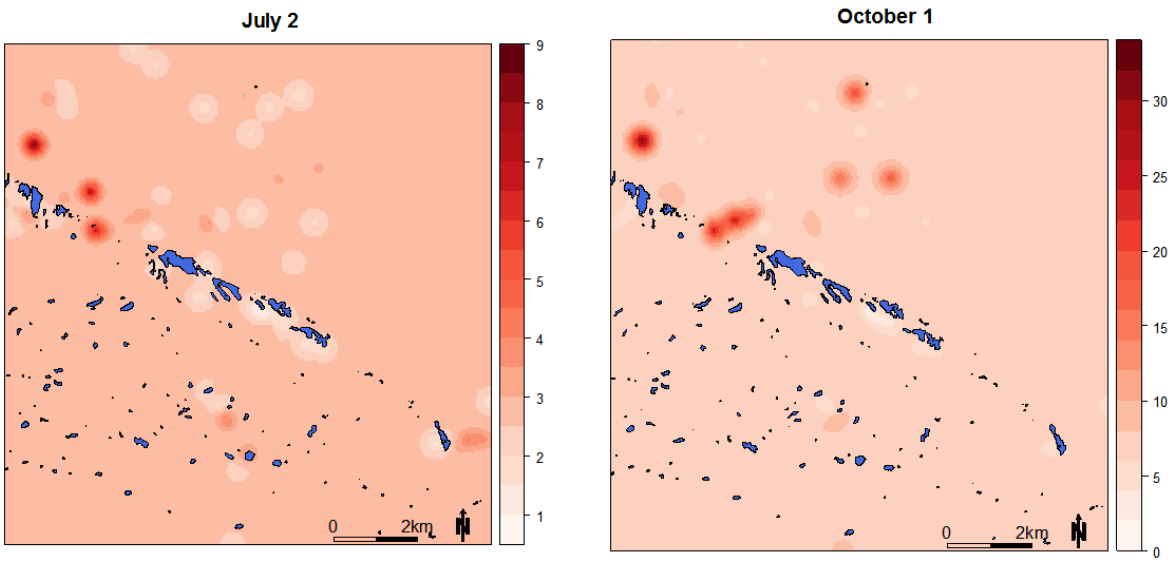


Figure 2.2: predicted maps during peak abundance for *Aedes vexans*

Conclusions

These results indicate that climatic factors affect the abundance of both RVF vectors. The modelling approach indicated that cumulative rainfall, minimum temperature were negatively associated with *Culex poicilipes* abundance while NDVI, minimum temperature and relative humidity were positively associated with it. For *Aedes vexans*, only rainfall was positively associated with the vector abundance (peaking in phase with the rainy season e.g. boreal summer). These results can be used to improve the survey and control of Rift Valley Fever vectors in the Barkedji area in northern Senegal.

References

FAO. 1997. Africover, classification de l'occupation du sol.80p. Rome. Grenier, A.

Anon. 1956. Phytogéographie, Yangambi. Réunion des spécialistes du CSE en matière de Phytogéographie, 28 juillet au 8 août 1956, Yangambi, CSA/CCTA, Londres, 22, 35 pp.

[F] Pilot Integrated multi-agent system for RVF in Senegal

Overview

The main objective was to implement a multi-agent pilot system based on different environmental parameters and RVF disease components (vector-host-pathogen) for the Barkedji area in Senegal.

We developed a data management platform to centralize the environmental and epidemiological field/laboratory data collected by the different QWeCI project partners in Senegal.

Based on the published literature, we setup a decision system with the same data to integrate analyses tools and data mining with cartographic views.

UML formalization

UML (Unified Modeling Language) is used to formalize the different model components. UML is an object oriented standardized programming language. It includes a set of graphic notation techniques to create visual models of object-oriented software-intensive systems and is commonly used in the field of software engineering.

In strength depends on its ability to "clearly and accurately model the structure and behavior of a system irrespective of any method or programming language" (Muller and Gardner, 2000).

MAS implementation

The approach we used is based on the implementation of a multi-agent system. The MAS principle is to study at an aggregate level several sub-systems known at the local level (Daude, 2005).

The selected approach aims at identifying the agents; specifying the environmental conditions through which these agents evolve and finally transcribing the different algorithms governing their behavior and interactions. We used different idealized scenarios to improve our understanding of the relationship between RVF disease outbreaks and environmental factors for the Barkedji area.

GAMA platform

GAMA (Taillandier et al, 2000) was used to run and interpret the integrated model system. GAMA is a simulation platform, which aims at providing field experts, modellers, and computer scientists with a complete modelling and simulation development environment for building spatially explicit agent-based simulations. It is being developed by several French and Vietnamese research teams under the umbrella of the IRD/UPMC International Research Unit since 2007. GAMA is easy to use and proposes:

- Spatial analysis tools: a GIS (geographic information system) interface which allows studying the Barkedji area

- Configuration tools: parameterization of the number of camps, the initialization of population of vectors and hosts …
- Models output tools: Estimate vectors and hosts population dynamics, output risk maps and simulations…

Model Input dataset and disease observation dataset

Spatial data

The study was conducted with field and environmental data collected from the rural Barkedji area which is included in the district with the same name (Common Dahra-Linguère Department Linguère Louga). We selected the Barkedji area because of the large RVF outbreaks that took place historically over the region. The Barkedji area is characterized by a large livestock density and by a large number of temporary ponds. For the sake of generalization, we chose to integrate a broader spatial dimension which starts from the administrative division of Senegal. All actions are logged handled according to the following temporal hierarchy: Year -> period (dry or rainy season) -> month -> day. All fixed objects were geo-referenced.

Climatic, hydrological and water quality data

One of the fundamental subjects of our study is the pond dynamic which is related to several environmental factors:

- The rainfall stations and automatic weather stations were installed around water bodies: measures such as wind speed, temperature, rainfall… were automatically generated and collected periodically.
- Limnimetric measurements: they can measure the water level of ponds.
- Water quality parameters: they measure the pH, temperature, suspended solids from pond water, turbidity of the ponds…

Entomological and virological data

During fieldwork, mosquitoes were captured using standard and CO2 bait traps surrounding large livestock population (mutton and chicken). This was followed by individual identification of the different mosquito species to determine their age, sex… In addition, blood samples were collected on sentinel herds. Virology parameters in mosquitoes were assessed on the basis of ground sample. Virology parameters in animals were also tested in order to check for RVF circulation (using IGG, IgM antibody standard tests).

Results

A first simulation has been developed using the GAMA platform (Figure 1) to allow us to investigate the impact of different rainfall scenarios on RVF dynamics in the Barkedji area.

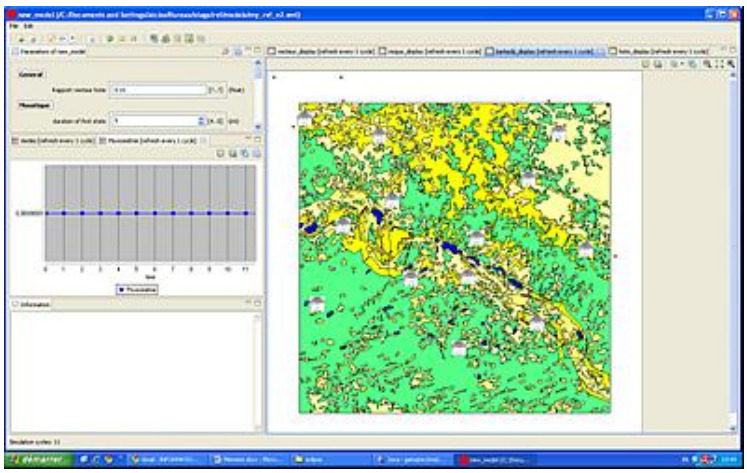


Figure 2 *: Screen simulation (initialization phase)*

The first simulation results for Barkedji corroborate the RVF emergence mechanism identified for the Ferlo region (Ba et al, 2005; Ndione et al, 2008) e.g. unusual large precipitation occurring during the end of the rainy season (Sep-Oct) drives risk in RVF outbreaks. The simulation data essentially came from different scientific areas such as climatology, hydrology, entomology, veterinary science...

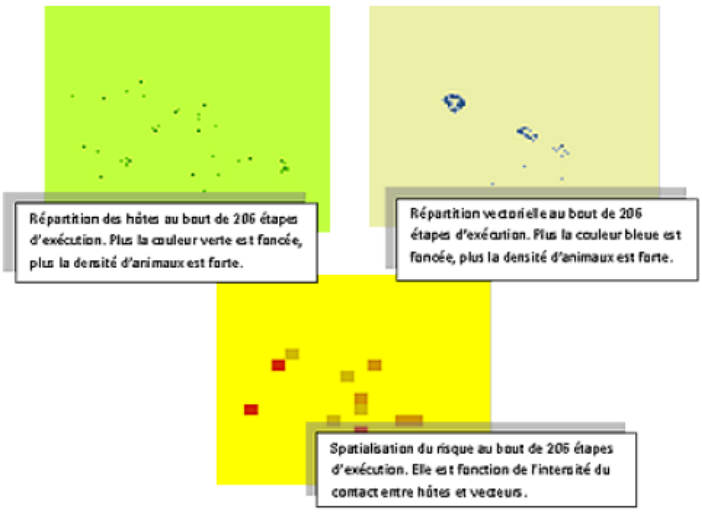


Figure 3 *: Output Screens of the simulation*

Scenario 1: Reference scenario

The 2012 rainy season remains the reference scenario. Large intra-seasonal rainfall variability was observed over the study area. Large precipitation was observed from week 20 to week 27 (figure 3). Then a long pause in rainfall (21 days) was observed between the weeks 28 and 32. This seasonal rainfall distribution had significant impacts on the vector dynamics as shown on figure 4.

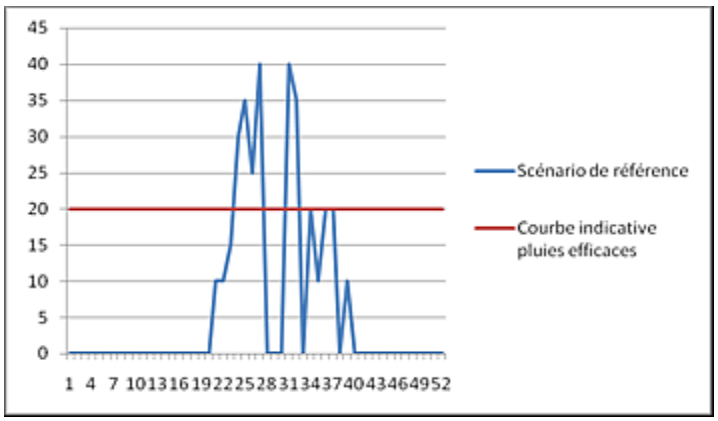


Figure 4*: Reference scenario rainfall curve*

Number of simulated Culex Poicilipes and Aedes Vexans (related to Scenario 1)

A quasi extinction of Aedes Vexans population following the end of the rainy season (day 220) can be observed (red line on Figure 4). At the same time, the large hatching of Culex Poicipiles is shown on Figure 4 (white line) corresponding to the Aedes Vexans extinction phase. In an average situation, the Aedes emergence only occurs during the first half of the rainy season. On the other hand, following the long rainfall pause (that lasted 21 days) that particular year, a large rainy event has favored deposition formerly sufficiently dry Aedes Vexans eggs in temporary ponds, to finally cause uncommon Aedes Vexans emergence in October (Weeks 32 and 33).

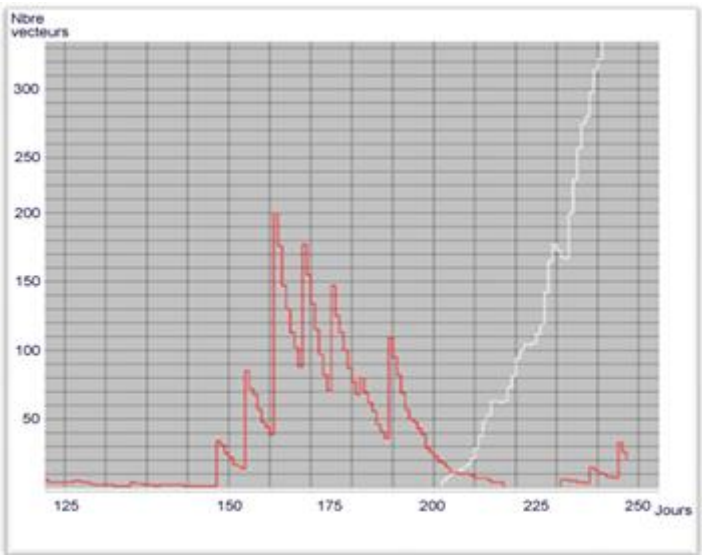


Figure 5*: Evolution of the RVF vectors based on the reference scenario*

Those unusual late rainfall events (occurring at the end of rainy season e.g. Oct) allow the Aedes eggs to hatch, while in principle they should have hatched at the beginning of the following rainy season. This also maintains the Culex population at a relatively high level.
The RVF virus is found again in a double favoring environment with possibilities of dispersion (or initialization of the epidemiological cycle e.g. Aedes role) and of intensification (Culex amplification role). A strong virus circulation was consequently simulated in herds (Figure 5) when the hatching of both vectors occurred (day 235-250).

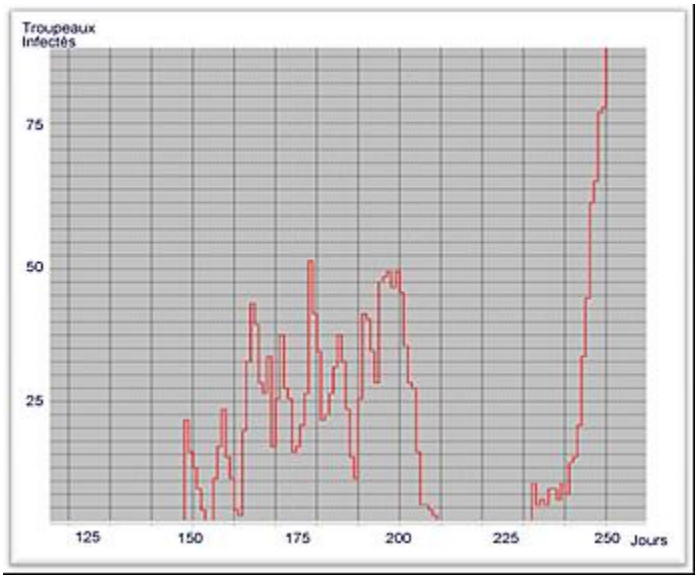


Figure 6*: Reference scenario Infected herd evolution*

Scenario 2: intense and regular rains

Scenario description: We selected a particular season when the rains were intense (clearly higher than the reference threshold of efficient rains e.g. 20mm) and regular (Figure 6).

Number of simulated Culex Poicilipes and Aedes Vexans (related to Scenario 2)

The first efficient rain has caused a large emergence of the Aedes Vexans population; the abundance of Aedes vectors then varies following the rainfall intra-seasonal variability before falling dramatically at the end of the rainy season (Figure 7).

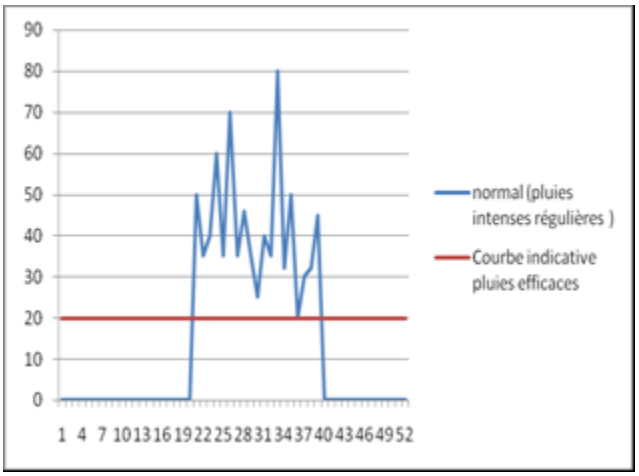


Figure 7*: Scenario 2 rainfall curve*

The emergence of culex Poicilipes occurs early in the season (day 200), but their presence is only observed when Aedes Vexans population are almost non-existent.

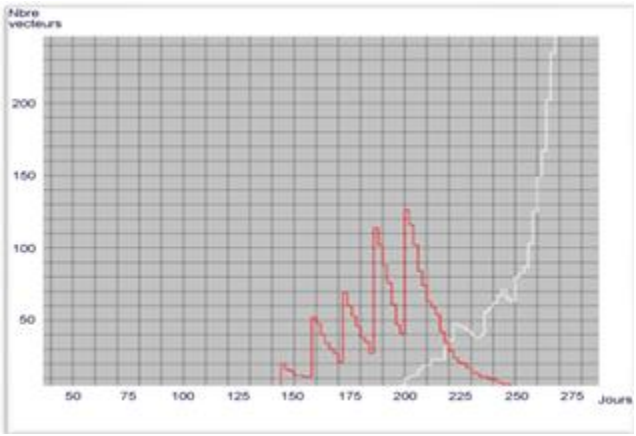


Figure 8*: Scenario 2 Vectors Evolution*

Low infected herd peak (less than 2% of infected herds; not shown):

- The first peak coincides with the massive hatching of Aedes Vexans. As there is almost no Culex Poicilipes emerging during that period, Aedes Vexans are the only vectors able to transmit RVF. However, their abundance and aggressiveness are not sufficient to allow significant virus transmission.
- The second peak coincides with the hatching of Culex eggs without the presence of Aedes. Despite large vector densities, the virus transmission remains relatively low.

Conclusions

We highlighted the relationships between rainfall, RVF vectors and infected herds (mainly livestock). We showed that an unusual rainfall event occurring at the end of the rainy season drives an unusual hatching of Aedes Vexans (that should have hatched during the following rainy season). Large abundance of Aedes in phase with the emergence of Culex mosquitoes during that period causes amplification in virus transmission, leading to an increased risk of RVF outbreaks over the region.

References

A. Cisse, A. Bah , A. Drogoul , A.T. Cisse , J.A. Ndione, C.M.F. Kebe, 2012. Un modèle à base d'agents sur la transmission et la diffusion de la fièvre de la Vallée du Rift à Barkédji (Ferlo, Sénégal), *Studia informatica universalis*,**12**: 77-97.

A.T. Cisse, a, A. Cisse, A. Bah, J. A. Ndione, C.M.F. Kebe, **2011**. *An agent based modeling method: Agent from "UP". An example on the Rift valley fever at Barkedji (Senegal);* Advance in Artificial Intelligence, Vol. 1, 685-690, 1-2 October 2011, Chengdu, PRC, ISBN 978-1-61275-994-4; Electronically (to be) available at http://www.ier-institute.org.

A. Cissé, A. Drogoul, P. Taillandier, A. T. Cissé, A. Bah, J.A. Ndione, C. M. F. Kébé, I. Dia, A. Th. Gaye, M.Sangharé, 2010 : Un modèle à base d'agents pour étudier l'émergence et la transmission de la RVF (La fièvre de la vallée du Rift), Commisco 2010, Paris, Bondy, France

A. T. Cissé, A. Bah, Z. Guessoum, 2011. Conception d'un méta modèle agent sur les maladies à vecteurs, Doctoriales PDI 2011, Paris, Bondy, France.

Ndione J.-A., Besancenot J.-P., Lacaux J.-P. Et Sabatier P., 2003: Environnement et épidémiologie de la fièvre de la vallée du Rift (FVR) dans le bassin inférieur du fleuve Sénégal. *Environnement, Risques et Santé*, **2(3)**: 176-182.

Ndione J.-A., Bicout D., Mondet B., Lancelot R., Sabatier P., Lacaux J.-P, Ndiaye M. , Diop C., 2005: Conditions environnementales associées à l'émergence de la fièvre de la vallée du Rift dans le delta du fleuve Sénégal en 1987. *Environnement, Risques et Santé*, **4(2)**: S5-S10.

Muller P., Gaertner N., 2000. Modélisation objet avec UML. 2eme edition. (Paris : Eyrolles), 2000.

Daude E., 2006. A monte carlo approach to diffusion : une étude " historique " revisitée par la modélisation multi-agents. Modélisation et simulation multi-agents applications pour les Sciences de l´ Homme et de la Société, pages 385–409, 2006.

Linthicum KJ., Davies FG., Bailey CL., and Kairo A. 1983. Mosquito species succession in a dambo in an east African forest. *Mosquito News*, **43**: 464–470.

Taillandier, P., Drogoul, A., Vo, D.A. and Amouroux, E. 2012. GAMA: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. In 'The 13th International Conference on Principles and Practices in Multi-Agent Systems (PRIMA)', India, Volume 7057/2012, pp. 242-258.

[G] RVF mixed statistical Model for Kenya

Overview

Rift Valley fever is a mosquito-borne viral zoonosis that mainly affects sheep, goats, cattle and camels. It has been recognised as an important disease problem internationally; its emergence and spread is influenced by climate, land use patterns and trade. We analysed historical data on the disease outbreaks in Kenya obtained from the Department of Veterinary Services over the period 1951 – 2010 to identify predictors for outbreak. The analysis was conducted at the spatial scale of the division (n = 505) and the monthly infection status of each division was modeled as a binary response variable. Explanatory variables investigated included: monthly precipitation, normalized difference vegetation index (NDVI), elevation, land use/land cover, livelihood zone, soil type, livestock and human population densities and the number of times a division has had an outbreak before. Multiple regression models were used but the spatial multiple membership model supported by a Bayesian Markov Chain Monte Carlo (MCMC) framework provided the best fitting model. Descriptive analyses indicate that a total of 91 divisions in 42 districts (of the original 69 districts in place by 1999) reported RVF outbreaks at least once over the period. The mean interval between outbreaks was determined to be 43 months. Factors that were positively associated with RVF occurrence included increased precipitation, NDVI, low altitude and soil types: vertisols, solonertz and luvisols. The model generated has been used to generate a risk map that can be used to design risk-based surveillance. Its utility is demonstrated by its ability to identify new/potential hotspots (e.g. the northwestern Kenya) where outbreaks had not been reported.

Model Input dataset and disease observation dataset

RVF outbreak data

Annual records on RVF outbreaks for the period 1912 to 2007 were obtained from the Department of Veterinary Services (DVS) and Centers for Disease Control (CDC), Kenya. These records identified the province, district and division affected and the year when each area reported an outbreak. These data have been published by Murithi et al. (2010) and they have also been used by CDC to generate an RVF risk map for Kenya (Ministry of Livestock Development [MoLD], 2010). In these records, RVF outbreak is defined as a higher than normal occurrence of still births, abortions of perinatal mortality and hemorrhagic syndrome in livestock. However, some of the cases were confirmed using antigen and antibody enzyme-linked immunosorbent assay (ELISA) and real-time reverse transcription polymerase chain reaction (RT-PCR) tests for the purposes of official declarations.

For this analysis, data cleaning and refinement, guided by the records archived at the Department of Veterinary Services (DVS), were done to change the time component to monthly rather than the annual scale reflected in the original data.

Spatial data

Table 1 shows a list of variables used as predictors and the source of each data type. Data on livelihood zones were obtained from Famine Early Warning Systems Network (FEWS NET,

http://www.fews.net/Pages/default.aspx). FEWS NET has classified geographical areas into homogeneous units where people share similar livelihoods including options for obtaining food, income and market opportunities to inform food security analyses. Maps produced from this livelihood profiling are also used for monitoring food security risks and identifying intervention needs. Kenya has 19 livelihood zones (Grillo and Holt, 2009). For the purpose of descriptive analyses, these levels were collapsed into 6 more general livelihood categories: high potential, medium potential, marginal, pastoral, agro-pastoral and riverine.

We used global land cover data assembled by Food and Agriculture Organization of the United Nations (FAO) from the Global Land Cover (GLC2000) analysis that had been done by an international partnership of research groups coordinated by European Commission's Joint Research Centre (Di Gregorio and Jansen, 2000). GLC2000 data comprised daily observations collected over a 14-month period from VEGETATION's sensor on board SPOT 4. The GLC 2000 analyses led to the production of detailed regional and cover maps that were then aggregated by FAO using a simpler thematic global legend. These data are gridded at 1 km spatial resolution. The distribution of land cover types (at 1 km^2) was generated by division and the most dominant land cover type was assigned to corresponding divisions. Overall, the country has 11 land cover types; the most common types are various herbaceous domains, cultivated and managed areas. These were collapsed into 6 land use types including artificial, cultivated, herbaceous cover, tree cover, mosaic and water, for ease of use in the analysis.

Gridded ERA-Interim reanalysis precipitation data and minimum and maximum temperature estimates for the period January 1979 to December 2010 were downloaded from European Centre for Medium-Range Weather Forecasts (ECMWF) (Dee et al., 2011). These data have been optimised (global best estimates) to fit both short-range forecasts (from a model) and observed data. Normalised Difference Vegetation Index (NDVI) data for the period 1999 to 2009 were obtained from SPOT VEGETATION (http://free.vgt.vito.be/). NDVI, a measure of amount and vigour of vegetation on land surface, is derived from radiometric sensor measures of reflectance for both red and near infrared bands on two separate channels or images. NDVI estimates are derived by subtracting red band measures from the near-infrared and dividing the difference by the sum of the two measures. These values range between -0.1 and 1.0; negative values indicate clouds and water, positive values near zero indicate bare soil and higher values indicate dense vegetation. NDVI extracts are available on 10 day-intervals at a spatial resolution of 1 km. For this study, minimum, maximum and average values for each division were extracted.

The digital elevation data were originally generated by NASA Shuttle Radar Topographic Mission (SRTM) based on Digital Elevation Models (DEM) (http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1). The data used in this analysis were obtained from CGIAR-CSI STRM at a resolution of 90 m. The CSI STRM data have been upgraded using new interpolation algorithms for completeness and to enhance wider application for geospatial analyses.

Data on soil types were extracted from the Harmonised World Soil Database (HWSD) developed by FAO and the International Institute for Applied Systems Analysis (IIASA) (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009). The data has a resolution of 1 km and over 15000 different soil mapping units are recognised in the database. The database contains information of the soil units, soil properties and other parameters such as organic carbon, pH, water storage capacity, soil depth, etc.

Human population data for the years 1960, 1970, 1980, 1990 and 1999 were obtained from the Kenya National Bureau of Statistics.

Table 1: *Predictors used in the RVF statistical model*

Variable	Description
Livelihood zones	Livelihood practices (2006), FEWS NET
Land cover	Global land cover data (GLC2000), FAO
Precipitation	Monthly minimum, maximum and average for the period: 1979 – 2010, ECMWF
NDVI	Monthly average, minimum, maximum values from: 1999 – 2010, SPOT VEGETATION
Human population	Human and household census for 1960, 1970, 1980, 1990, 1999, Kenya National Bureau of Statistics
Elevation	CSI SRTM
Soil types	FAO's Harmonized World Soil Database (HWSD), 2008, FAO

Statistical analyses

A dummy variable was used to indicate whether, in a given month, a division had an outbreak (1) or not (0). A value of 1 was assigned only when a division reported the disease for the first time within a period of 6 months. Descriptive analyses were conducted to determine the distribution of outbreaks. Crude associations between RVF outbreaks and most of the predictors given in Table 1 were analysed using various types of statistical tests. A Chi-square test was used to determine the association between soil texture (broad classification of soil types into clay, loamy, sandy, and very clayey). The effect of elevation was assessed via a graphical analysis that allowed for the establishment of a threshold height under which a majority of the outbreaks were found.

Univariable analysis

Univariable Logistic regression models were also used to assess the association between precipitation and NDVI and RVF outbreaks. Precipitation variable was formulated in various ways before fitting in the model, i.e.:

- monthly values,
- 1- and 2-month lagged values,
- 2- and 3-month running cumulative values,
- 2- and 3-month running mean values.

For NDVI, minimum, mean and maximum values were fitted. Forms of precipitation and NDVI variables that gave the largest log likelihood estimates were used in the subsequent analyses.

Elevation and the number of times a division had experienced an outbreak were fitted as a continuous variable to assess the linearity assumptions. Both of these variables were categorized since this assumption was not met.

Multivariable analysis

Multivariable models fitted to the data included:

(i) Mixed effects logistic regression model with livelihood zone as a random variable
(ii) Mixed effects logistic model utilizing Bayesian MCMC estimation methods,
(iii) Spatial Multiple-Membership (MM) model fitted under the MCMC framework to account for spatial autocorrelation at the division level.

Random variables were included in the models to account for spatial clustering at the livelihood zone level. MCMC model used non-informative priors and up to 50,000 iterations were executed to allow for a good mixing of the chains.

Results

Descriptive analyses

Spatio-temporal distribution of RVF cases

A total of 505 divisions (from the 1999 human and housing census) were used in the analysis. Twenty percent (n = 102) of these had at least one outbreak of RVF over the period 1979 – 2010. Mean inter-epidemic period was estimated to be 5.4 years, with a 95% confidence interval of 4.4 – 6.4 years. Temporal and spatial distributions of the outbreaks per month for the period January 1979 to December 2010 are outlined in Figures 1 and 2, respectively. Figure 1 suggests that the number of divisions affected by RVF outbreaks have been increasing over time, particularly in 1997/98 and 2006/2007 outbreaks.

Univariate analyses

Soil texture

Table 2 gives the results of a Chi-square test used to determine the association between soil texture and RVF occurrence. The analysis uses 502 divisions that had reliable information on the dominant soil texture and the outcome represents the number of divisions that had had at least one infection over the study period. This analysis shows RVF incidence was significantly higher in divisions where clay soils were the dominant soil texture.

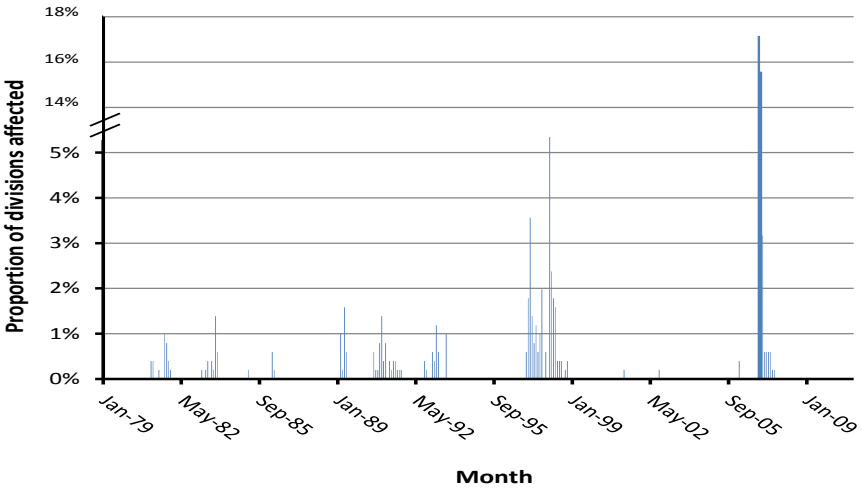


Figure 1: *Temporal distribution of RVF outbreaks per division-month over the period January 1979 to December 2010*

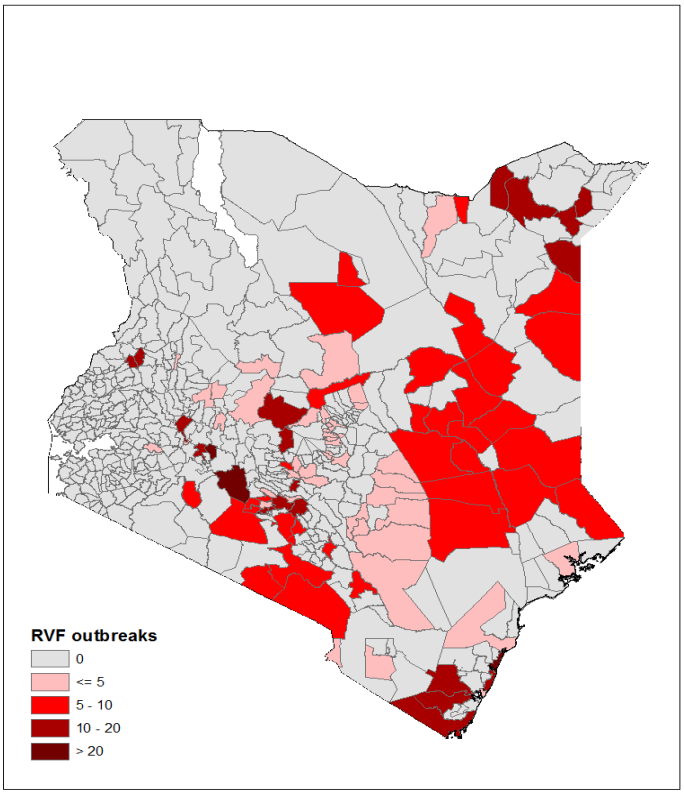


Figure 2: *Spatial distribution of RVF outbreaks per division-month over the period January 1979 to December 2010*

***Table 2**: Association between the soil texture and RVF occurrence based on historical data on RVF (January 1979 – December 2010)*

Soil texture	Number of divisions with Soil texture	Number outbreaks by soil texture	%	Chi (P)[1]
Clay	345	80	23.2	5.6 (0.02)
Loamy	70	11	15.7	1.1 (0.30)
Sandy	25	2	8.0	2.5 (0.12)
Very clayey	53	8	15.1	0.1 (0.32)

[1]Chi square is estimated for each soil texture by comparing the frequency of outbreaks in divisions with and without a given soil

Land cover and livelihood zones

The distribution of the number of divisions by land cover and livelihood variables is outlined in Table 3. The table also gives the distribution of the number of divisions that have had at least one outbreak at various levels of land cover and livelihood variables. A high proportion (30.1%) of divisions with herbaceous land covers as the dominant vegetation have experienced RVF outbreak at least once. Similarly, a high proportion (31.5%) of divisions that practice pastoralism as the dominant livelihood activity have had RVF outbreak at least once.

Table 3*: Distribution of the number of divisions that have had at least one RVF infection by land cover and land use variables based on historical data on RVF epizootics*

Land cover				Livelihood zones			
Levels	cases	n	%	Levels	Case	n	%
Artificial	3	6	50.0	High potential	30	174	17.2
Cultivated	7	98	7.1	Marginal	20	111	18.0
Herbaceous	62	206	30.1	Medium potential	12	57	21.1
Mosaic	1	55	1.8	Pastoral	29	92	31.5
Tree cover	22	111	19.8	Agro-pastoral	8	41	19.5
Water	0	2	0	Riverine	0	5	0

Elevation

A graphical analysis of the relationship between elevation and incidence of RVF is given in Figure 3. This graph uses monthly incidences of RVF over the period. It shows that most areas that have had the disease fall below approximately 2300 m above sea level. The variable was then categorized into three levels: 0 – 1000 m, >1000 - $\leq$2000 and >2000 m.

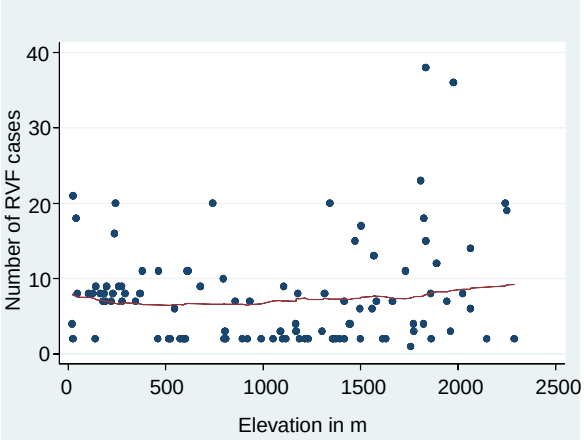


Figure 3*: Scatter plot showing the relationship between elevation (in metres) and RVF cases in Kenya based on historical data on RVF outbreaks for the period Jan 1979 – Dec 2010.*

Precipitation and NDVI

Log likelihood estimates generated from crude Logit models used to analyse forms of the precipitation variable were:

- -3102.69 for monthly values,
- -3050.08 for 1-month lagged estimates,
- -3033.71 for 2-month lagged values,
- -3061.06 for 2-month running cumulative values, and
- -3018.20 for 3-month running cumulative values.

This suggests that using cumulative values of precipitation for the recent 3 months gives a better fitting model than the other forms of the variable. Means and variances from precipitation estimates for the last 2 and 3 month had similar log likelihood estimates as their respective running cumulative values. A similar analysis using forms of NDVI variables (minimum, mean and maximum) suggests that maximum NDVI values predicts RVF outbreaks better than minimum and mean values.

Multivariate analyses

Three models fitted to the data are illustrated in Table 4. All the models, identified soil type (vertisols, solonertz, and luvisols), elevation, precipitation, NDVI and the number of previous outbreaks as significant predictors. Livelihood zone was also significant as a random effects variable.

SMM was the best fitting model indicated by the low DIC. The results indicate:

- Low altitude has a higher risk of RVF than high altitude
- Areas that have had repeated infections have higher chances of experiencing new outbreaks
- Cumulative precipitation (over recent 3 months) and maximum NDVI are positively associated with outbreaks. For each of these variables, squared derivatives- precipitation^2 and NDVI^2 - were included since the linear forms failed to meet the linearity assumption.

Table 4: *Regression models fitted to the RVF historical data from Kenya*

Variable	Levels	Mixed logit model		MCMC model		SMM model	
		β	SE(β)	β	SE(β)	β	SE(β)
Fixed effects							
Soil type	Vertisols	0.59	0.27	0.66	0.31	0.84	0.52
	Solonertz	1.32	0.35	1.19	0.44	1.54	0.60
	Luvisols	0.72	0.38	1.21	0.58	0.86	0.63
	Others	0.00	-	0.00	-	0.00	-
Elevation	0 -1000	0.97	0.39	2.74	0.38	1.51	0.57
	>1000-≤2000	0.00	-	0.00	-	0.00	-
	>2000	-2.08	0.44	-1.99	0.48	-2.13	0.62
Precipitation		0.99	0.09	1.07	0.08	1.02	0.12
Precipitation^2		-0.04	0.01	-0.04	0.01	-0.04	0.01
NDVI (max)		-11.88	4.96	-10.89	1.79	-10.93	1.66
NDVI^2		16.17	3.95	15.75	1.31	16.81	1.52
Case_no	<2	0.00	-	0.00	-	0.00	-
	2 - 5	2.57	0.22	3.62	0.34	3.85	0.62
	>5	2.62	0.23	3.59	0.38	4.09	0.71
Constant		-11.65	1.60	-14.63	0.75	-14.67	1.21
Random effects							
Livelihood		1.84	0.48	3.34	0.97	4.12	2.67
AIC/DIC		1749.9		1648.9		1611.6	

The risk map generated from predicted risk of RVF is given in Figure 4. Compared to the mapped surveillance data illustrated in Figure 1, this output indicates new/potential RVF hotspots in the northwestern Kenya where RVF outbreaks have not been reported before.

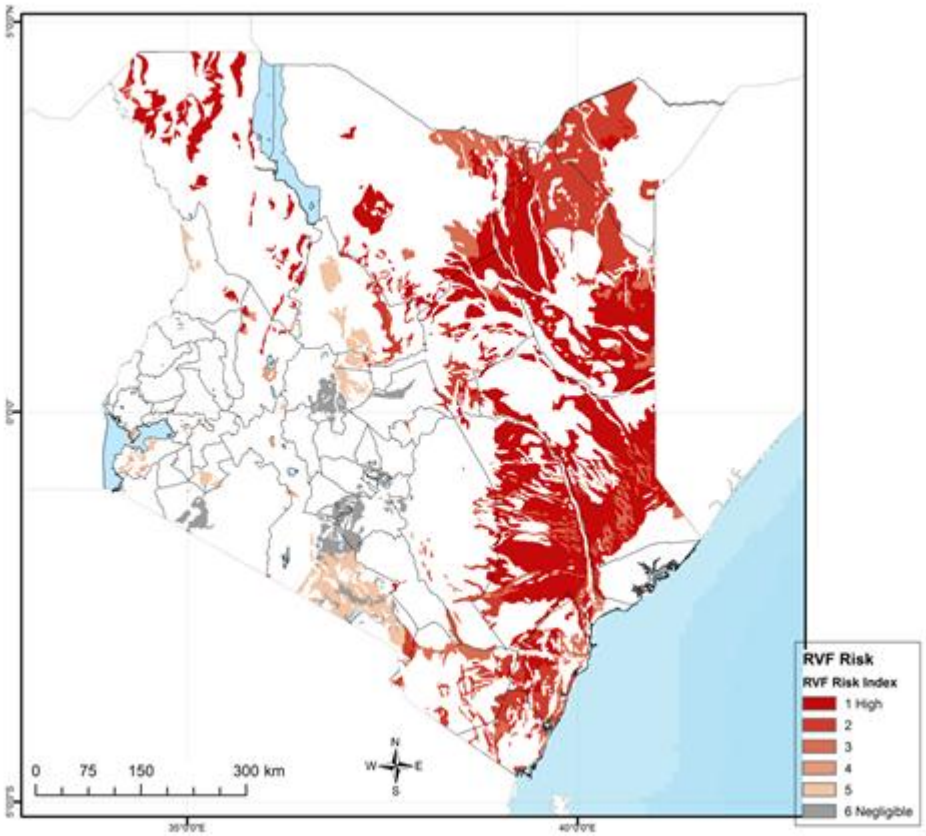


Figure 4: *Predicted risk of RVF based on the mixed logistic regression model*

Conclusions

This analysis has generated fundamental information on risk factors for RVF outbreaks and enabled the development of a risk map that can be used to design risk-based surveillance. We used surveillance data obtained from the DVS, Kenya which were collected in the course of RVF epizootics for there haven't been any structured surveys to determine the distribution of the disease nationally. Surveillance data have inherent biases related with spatial distribution of cases given that areas that report outbreaks diligently would get overrepresented in the surveillance databases. In addition, confirmatory laboratory diagnosis based on PCR/ELISA tests are often done for initial/primary outbreaks but as the epizootic unfolds, surveillance would be diagnosed based on syndromes /case definitions that match with the confirmed outbreaks.

Multiple models were used to compare their efficiencies in accounting for spatial autocorrelations. Spatial autocorrelation is an important factor to consider in this analysis because most of the areas that report RVF epizootics practice pastoralism where livestock move across community boundaries, potentially introducing the virus to contiguous areas that they visit. The mixed effect logistic regression model controls for spatial autocorrelation within a livelihood zone by using livelihood zone ID as a random effects variable. The Bayesian MCMC model uses structure though with a Bayesian estimation routine. The SMM model however uses an advance model that takes into consideration the neighbourhood structure of the divisions. Judging from the low DIC, this model fits the data better.

RVF outbreaks occur when multiple events converge in a given space/time. The relationship between RVF outbreaks and physical or climatic factors – precipitation, NDVI and soil types, as highlighted earlier, have been described earlier (Hightower et al., 2012; Anyamba et al., 2009). This work has revealed additional information such as the type of soils associated with outbreaks. Clayey soils such as vertisols, luvisols and solonetz have low water percolation rates and most often occur in lower altitudes, leading to the development and persistence of floods whenever persistent precipitation occurs. Linthicum et al. (1983) have shown that floods, whether local or extensive, increases the risk of RVF. They further shows that the persistence of floods for 10 – 15 days is necessary for the emergence of flood water *Aedes* mosquitoes and the persistence of floods for a further 4 – 6 weeks and their colonization by seconday mosquito vectors such as Culex spp, *Mansonia* spp and others can allow amplification of the virus to epidemic levels provided susceptible hosts such as sheep, goats and cattle are available. In arid and semi-arid areas, vegetation responds to precipitation changes and so it is expected that there would be a strong correlation between precipitation, NDVI and the disease outbreaks. Good vegetation cover also acts as good breeding sites for the vectors.

Further analyses are being do to better understand RVF transmission dynamics from a multidisciplinary perspective. This involves the collation of socio-economic data such as poverty levels, risk practices and levels of investment in disease control efforts e.g. vaccination. These analyses would involve the use of ecological niche models to capture future patterns of risk depending on climate, land use and livelihood changes.

This model will also be used to map RVF risk over West Africa in a near future.

References

Anyamba, A., Chretien, J-P., Small, J., Tucker, C.J., Formenty, P.B., Richardson, J.H., Britch, S.C., Schnabel, D.C., Erickson, R.L., Linthicum, K.J., 2009. Prediction of a Rift Valley fever outbreak. *Proceedings of the National Academy of Sciences of the United States of America*, **106**: 955-959.

Dee D, Uppala S, Simmons A, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda M, Balsamo G, Bauer P, et al., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**: 553 – 597.

Di Gregorio, A., Jansen, L.J.M., 2000. Environment and Natural Resources Service, GCP/RAF/287/ITA Africover - East Africa Project and Soil Resources, Management and Conservation Service. 179 pages, FAO, Rome.

FAO, 2000. Landcover classification systems (LCCS): classification concepts and user manual. Natural Resource Management and Environmental Department, FAO.

FAO/IIASA/ISRIC/ISS-CAS/JRC, 2008. Harmonized World Soil Database (version 1.0). FAO, Rome.

Grillo, J., Holt, J., 2009. Application of livelihood zone maps and profiles for food security analysis and early warning. Guidance for Famine Early Systems Network (FEWS NET) representatives and partners. http://www.fews.net/docs/Publications/Guidance_Application%20of%20Livelihood%20Zone%20Maps%20and%20Profiles_final_en.pdf (Accessed: 2/2/2013)

Hightower, A., Kinkade, C., Nguku, P.M., Anyangu, A., Mutonga, D., Omolo, J., Njenga, M.K., Feikin, D.R., Schnabel, D., Ombok, M., Breiman, R.F., 2012. Relationship of Climate, Geography, and Geology to the Incidence of Rift Valley Fever in Kenya during the 2006-2007 Outbreak. *The American Journal of Tropical Medicine and Hygiene*, **86**:373-380.

Linthicum, K. J., Davies, F. G., Bailey, C. L., Kairo, A., 1983. Mosquito species succession in a dambo in an East African forest. *Mosquito News*, **43**:464 – 470.

Soumare, P. O., Freire, C. C., Faye, O., Diallo, M., de Olieveira, J.V., Zanotto, P.M. and Sall, A. A., 2012. Phylogeography of rift valley fever virus in Africa reveals multiple introductions in Senegal and Mauritania, *PLoS One*, **7**, e35216

[H] Malaria and climate seasonality over Limpopo, South Africa

Overview

Malaria is the most nagging parasitic infection affecting humans, accounting for an estimated 300–500 million cases of malaria worldwide of which ninety percent of this occurs each year in sub-Saharan Africa [1]. Recent resurgence of Malaria in the East African highlands involves multiple factors, ranging from climate and land use change, to drug resistance, variable disease control efforts, and other socio-demographic factors [2, 3]. Malaria epidemics have long been reported to occur among vulnerable populations where immunity is often non-existent or poorly developed. It is estimated that epidemic malaria causes between 12% and 25% of estimated annual worldwide malaria deaths, including up to 50% of the estimated annual malaria mortality in persons less than 15 years of age [4]. Malaria is an extremely climate-sensitive tropical disease [5]. Whereas rainfall anomalies are widely considered to be a major driver of inter-annual variability of malaria incidence in the semi-arid areas of Africa [4, 6], recently recorded warming trend in East African highlands corresponded with concomitant increases in malaria incidences [3, 5]. The same study also reported that the biological response of mosquito populations to warming can be more than an order of magnitude larger than the measured change in temperature. This finding thus shows the importance of nonlinear and threshold responses of malaria (a biological system) to the effect of regional temperature change.

A mere 0.5°C increase in temperature trend can translate into a 30–100% increase in mosquito abundance, thus indicating a ‘‘biological amplification’’ of temperature effects [3]. Climatic determinants are considered particularly important, since both the disease agent (*Plasmodium*) and vector (Anopheles mosquitoes) are strongly affected by climate. Temperature, rainfall and humidity have been associated with the dynamics of malaria vector populations and, therefore, with spread of the disease [7]. Temperature determines parasite and vector development; rainfall provides conducive conditions for mosquito breeding sites; humidity and temperature together affects mosquito survival [8]. It is now believed that malaria epidemics caused by meteorological factors can be predicted from climatic indicators and climate forecasts. However, further efforts are needed to assess Malaria response to climate change in view of the known climatic determinants. Edi et al. [9] have compared climate suitability maps for malaria in the topographically diverse country of Zimbabwe and found that the projected warming from global climate models would make the country’s entire highland area climatologically more favourable to malaria by the year 2050.

Malaria is a notifiable disease in South Africa. By using both passive and active surveillance, the case reporting system aims to capture every infection rather than clinical cases only [10]. Malaria transmission is distinctly seasonal, with transmission limited to the warm and rainy summer months [11]. Historically, and without intensive control programs, case notifications generally increase from November onwards, peak in late-summer to autumn (March–May) and decline by the end of June. As a result, the average seasonal pattern in malaria incidence follows the periodicity in rainfall and temperature with a 3–4-month lag [11].

Large epidemics of malaria elsewhere have been associated with climate and temperature anomalies, such as in Colombia, the Indian subcontinent, and Uganda; and recently, it is shown that, in Botswana, indices of El Niño-related climate variability can serve as the basis of malaria risk prediction and early warning [4, 8, 12, 13 and 5]. In South Africa, there have been

reports of significant reduction in Malaria trends over recent years [14] as a result of increased DDT spraying. It is also known that changes in temperature and rainfall results in changes in habitat. This could result in shifting, expanding, or contracting known mosquito boundaries, and thus result in changes in Malaria transmission. However, despite being armed with this knowledge, it still remains very poorly understood what the influence of recent climate changes has on Malaria prevalence in South Africa. There are also indirect effects of climate change on Malaria transmission emanating from population movements due to extreme events such as droughts. Such population movement may be from non-infested to infested areas, or it could result in transferring the parasite to a population with no immunity. It is important to understand current drivers of malaria, so as to put in place effective measures toward its control.

In South Africa, Malaria is endemic in the low-altitude northern and eastern areas along the border with Mozambique and Zimbabwe. Transmission is prevalent mainly in KwaZulu-Natal, Limpopo, and Mpumalanga provinces (10, 15 and 16). All the three provinces have recorded reduction of Malaria cases as a result of various social, policy and economic factors [15, 16]. Although climate is known to influence Malaria significantly, it was not considered in their analysis [15, 16]. Landman et al. [17] analysed meteorological station data from 1960 to 2003 and found that 15/26 stations in central and eastern parts of the country had experienced the largest positive trends in mean temperature during April. Also, the months of largest minimum trend appears to fall in the late spring and early summer (September to December). Results further show that cool days are generally on the decline, thus re-affirming that temperatures have generally risen all over South Africa. The records of minimum temperatures (Tn) also revealed that warmer nights have increased while cooler nights have decreased. It is known that after the Mozambican floods in 2000, there was a huge increase in Malaria cases. Floods and higher rainfall in areas such as Mozambique have also played their part by creating new, or larger, breeding sites for mosquitoes which carry the malaria parasite [18]. Due to the fact that climate change by itself will increase vulnerability [19, 20], target planning is necessitated, and the need to obtain clear pictures by carefully considering all factors cannot be overemphasized. This study complements existing studies on social, policy and economic factors as determinants of malaria transmission, by analyzing the importance of climate variability and change on malaria transmission in South Africa.

This paper aims to assess the temporal and spatial pattern of malaria in Limpopo province in relation to current climate variables. This will form the basis for the development of an early warning system, based on seasonal climate forecast. For improved control in epidemic regions, the World Health Organisation (WHO) advocates the development of integrated malaria early warning systems, based on vulnerability assessment, seasonal climate forecasts, weather and environmental monitoring, and case surveillance [21]. In order to achieve the above aims, this study assesses a 10 year spatio-temporal spread of malaria in Limpopo province (at municipality and district levels) in relation to current climate variables. It superimposes rainfall and temperature (minimum and maximum) of the closed stations, and assesses the correlation between climate variables and Malaria cases. We present a 10 year spatial spread of malaria in Limpopo province at district level, superimpose rainfall and temperature (minimum and maximum) of the closed stations, and assess any correlation between the climate variables and malaria cases using Pearson Correlation. Ordinary Least Squares (OLS) Method is applied to test whether variation in rainfall and temperature can help explain malaria cases, and Pearson correlation used to test correlation of the variables. Time series data are tested for stationarity using ADF and KPSS, and Autoregressive

Distributed Lag Model (ARDL) – bounds test method is applied to establish the existence of short-run and long-run relationship.

In essence, this study attempts to analyses spatial and temporal change in malaria incidences at district level; determines the influence of rainfall and temperature on malaria transmission; ascertains the strength of the influence and; finally, test existence(or non-existence) of any short or long-run relationship between climate variable and malaria cases in Limpopo Province, South Africa.

Model Input dataset and disease observation dataset

This study uses combination of methods: spatial-temporal, correlation anaysis and econometric approaches. Ordinary Least Squares (OLS) is used to analyse variation in rainfall and temperature in explaining malaria cases and Pearson Correlation tests the strengths of the influence. Time series data are tested for stationarity using Augmented Dickey Fuller Test (ADF) and Kwiatkowki-Phillips-Schmidt-Shin (KPSS), and Autoregressive Distributed Lag Model (ARDL) – bounds test method is applied to establish existence of short-run and long-run relationship.

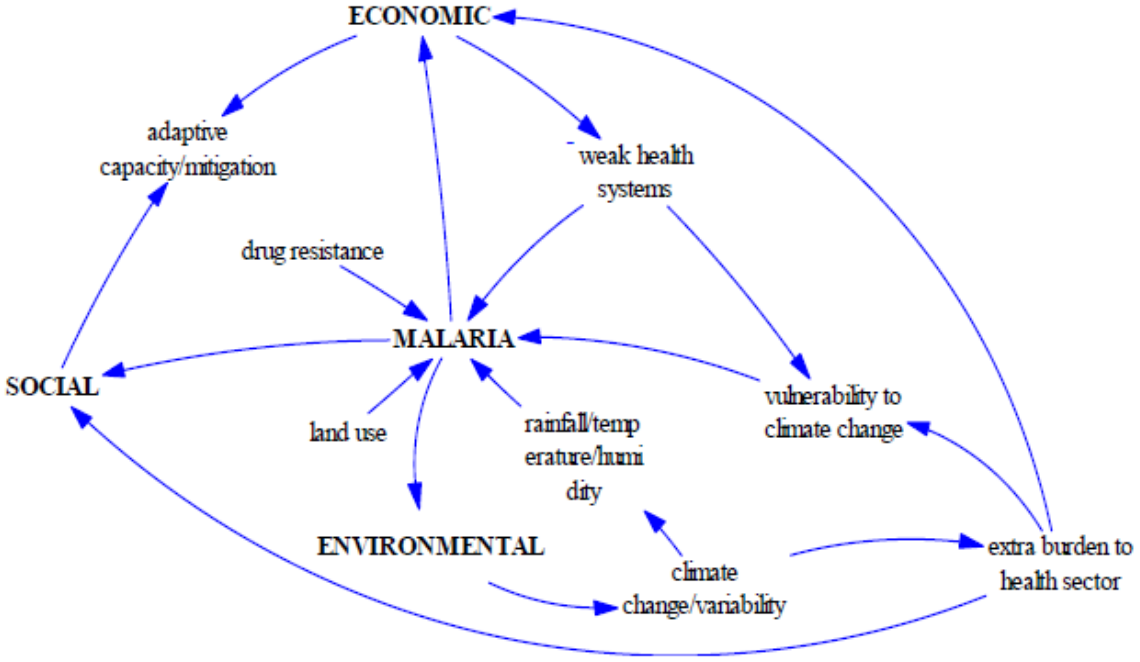
Data and Sources

The study uses time series of average rainfall and temperature, as well as malaria cases spanning the period from 1998 to 2007. Climate data was obtained from the South Africa Weather Services. Daily station data of precipitation and temperature (minimum and maximum) from 1998 to 2007 was used to construct climate disease envelopes at municipal and district levels, while malaria data were obtained from the South African Department of Health and from the Malaria control Centre in Tzaneen (Limpopo Province). This data contains different variables, including malaria cases captured through passive and active surveillance systems. Demographic, geographic and social-economic data were also captured, these include: sex, age, date of which blood sample was taken, district (current and of origin), local municipality, and the name of the health facility. The data also contain information on deaths. Details of the methods on how this data was collected can be obtained from [15]. For the purposes of this paper however, municipal and district malaria cases were extracted for analysis.

Conceptual Framework

The conceptual framework for this study advances a multiple-factor explanation for malaria, ranging from climate and land use change, to drug resistance, variable disease control efforts, and other socio-demographic factors. Figure 1 below illustrates a simplified non-detailed interrelationship.

Figure 1: Malaria-Environment Nexus



Malaria is influenced by social, economic and/or environmental factors [22]. Of importance to this study is the influence of environmental factors in terms of variation in rainfall and temperature. Empirical studies have reported rainfall [4, 23 and 24] and temperature [25, 26] as the main climate factors that influence malaria transmission; however, other studies have included climate variables such as humidity and vegetation [27]. Variation in rainfall and temperature influences mosquito population. By shortening its life cycle, its population is increased [28, 29]. High temperature shortens the development time of vector-borne pathogens; and combined with favourable climate conditions, the population of carrier-mosquitoes increases. In both theory and in literature, variation in rainfall and temperature will lead to the increase or decrease in malaria cases observed. Growth of Anopheles vector is accelerated under conditions of increased temperature with optimal larval development at 28^0C and optimal adult development between 28^0C and 32^0C [30, 31]. Very high rainfall and temperatures negatively affect mosquito development, but a moderate climate will provide a condusive environment for the vectors to grow. Significant warming trend therefore amplify mosquito population dynamics, so as to contribute alongside drug resistance, and land-use patterns, to the increased incidence of malaria [3, 32, 33, 34]. Apart from the influence of climate in malaria transmission, socio-economic factors such as population and migration also play significant roles [35]. Rise in malaria disease increases pressure on the health sector. Moreover, a combination of mutating malaria parasites (and subsequent resistance to drugs), resource constrains, and weak health systems, implies low adaptive capacity [36].

This paper uses correlation, spatial, and econometric approaches to analyse the relationship (and strength of the relationship) between malaria and climate variables (i.e. temperature and rainfall).

Spatial and temporal change of district Malaria incidences in Limpopo Province

Using GIS, the spatial distribution of Malaria at municipality and district levels were mapped. Changes in the distribution were obtained using GIS spatial analytical techniques. Malaria records for the various municipalities were spatially weighted and aggregated at district level.

Weighted points at the centroid of each district were then interpolated using the Inverse Distance weighting (IDW) model [37, 38 and 39]. By assigning values to locations based on the surrounding weighted values, the model outputs continuous surfaces which represent spatial distribution of malaria cases at a power of 2 and a search radius of 20km. The model assumes that the mapped variable (malaria cases) decreases in influence with distance from its weighted location [40].

Given the large seasonality in climate variables; temperature (T) and total precipitation (P) and malaria cases (M), a linear relationship between T & M, P & M and T & P can be derived from the Pearson Correlation coefficients ($\gamma_{T,M}$; $\gamma_{P,M}$ and $\gamma_{T,P}$) as reported in [41]. According [42, 43], the linear relationship between say, T & M with the influence of P removed can be determined from the partial correlation given by

$$\gamma_{TM} = \frac{\gamma_{TM} - \gamma_{PM} \times \gamma_{TP}}{\sqrt{1-\gamma_{PM}^{\ 2}} \times \sqrt{1-\gamma_{TP}^{\ 2}}} \quad \text{..... (1)}$$

Here, $\gamma_{T,M}$ is computed as a simple correlation between the residuals from a regression of M on T and a regression of P on T. In this correlation analysis, three season lengths i.e., number of months j={1, 3, 12} are considered over a 14 month malaria cycle (resulting in 42 season groupings). The correlation analysis is repeated for all the 42 seasons specified by a malaria ending month (in this case September is considered as the ending malaria month).

Econometric approach

The study applies ordinary least squares estimation technique, and tests for stationarity and cointegration. As a standard procedure when working with time series data, rainfall, temperature and malaria data are tested for stationarity, to avoid spurious results common to running regressions with non-stationary data. A stationary process (unit root) is a stochastic process whose joint probability distribution does not change when shifted in time or space. Consequently, parameters such as the mean and variance, if they exist, also do not change over time or position. Stationarity is used as a tool in time series analysis, where the raw data are often transformed to become stationary. Stationarity test for data is a fundamental requirement if the results of relationship between variables are to be considered reliable in time series analysis [44, 45]. In the mathematical sciences, a stationary process is a stochastic process whose joint probability distribution does not change when shifted in time or space. We test the null hypothesis of unit root (stationary) in our variables. If the null hypothesis is not rejected, this implies that the variable(s) are stationary.

Model specification

We follow the same methodology applied by [27] who investigated the relationship between climatic parameters: rainfall, temperature, humidity, sea surface temperature (SST), El Niño-Southern Oscillation (ENSO), and the normalized difference vegetation index (NDVI), to examine malaria cases in a district in Bangladesh.

However we simplify the model to include only two variables: rainfall and temperature. Our objective is to determine whether there is a long term relationship between rainfall, temperature, and malaria in Limpopo Province, South Africa. We used simple Ordinary Least Squares (OLS) method and Cointegration analysis to investigate this long term relationship.

$$LOG(Y) = \alpha + \beta_2 LOGRAIN + \beta_1 LOGTEMP + \varepsilon \quad \text{..... (2)}$$

Where

$LOG(Y)$: Change in Malaria cases
$LOGTEMP$: Change in average temperature
$LOGRAIN$: Change in average rainfall
$\beta_1 and \beta_2$ are the coefficients
ε : is the error term

The error term, ε accounts for all other variables that may explain malaria cases. Thus, while the model could have been significantly improved by accommodating more variables influencing malaria, for a long term relationship two variables can suffice.

We test the following hypothesis:

$$Null - Hypothesis : H_0 = \beta_1 = \beta_2 = 0$$
$$Alternative - Hypothesis : H_1 = \beta_1 \neq \beta_2 \neq 0$$

The null hypothesis states that the coefficients of the regression mode would jointly equal to zero, while the alternative hypothesis states that they would not jointly equal to zero. We expect the coefficients of the regression to exhibit a significant divergence from zero because; the notion that malaria is affected by temperature and rainfall is a scientific axiom.

From the hypothesis (H_o) of no co-integration, the asymptotic distribution of the obtained *F-statistic* is nonstandard, regardless of the degree of integration of the variables. This depends on: whether variables included in the ARDL model are I (0) or I (1); the number of regressors; whether the ARDL model contains an intercept and/or a trend and; the sample size. Two sets of critical *F-values*, representing the lower bound and the upper bound, have been provided by [46] for large samples. Narayan [47] presents the critical *F-values* for sample size ranging 30–80. If the computed *F-statistic* for a chosen level of significance lies outside the critical bounds, a conclusive decision can be made regarding the cointegration of the regressors. If the statistic is higher than the upper bound, the null hypothesis of no cointegration can be rejected and the next step is to estimate the ARDL ECM where the short-run and long-run elasticities may be determined [46, 47 and 48].

The Autoregressive Distributed Lag (ARDL) Model Specification

The study used Autoregressive Distributed Lag (ARDL) cointegration framework to examine the short term and long term characteristics of rainfall and temperature, as it relates to Malaria in Limpopo Province, South Africa. The ARDL specification takes the following form:

$$\Delta \ln mala_t = \gamma + \sum_{i=0}^{m} \alpha_i \Delta \ln rain_{t-i} + \sum_{i=0}^{n} \delta_i \Delta \ln temp_{t-i} + \sum_{i=1}^{k} \varpi_i \Delta \ln mala_{t-i} +$$
$$\beta_1 \ln rain_{t-1} + \beta_2 \ln temp_{t-1} + \beta_3 \ln mala_{t-1} + \varepsilon_t \quad \text{..... (3)}$$

Δ means first difference

ARDL estimation proceeds in two steps. The first step is to estimate the above equation by OLS in order to establish the existence of a long run relationship. Once cointegration is confirmed, the second step is to estimate the long run coefficients and the short run coefficients using the respective ARDL and Error Correction Methods (ECMs). We estimate the unrestricted model and progressively reduce it by eliminating the statistically insignificant coefficients, and reformulating the lag structure where appropriate, to achieve orthogonality

in terms of levels and differences. The lag of the error term, Ecm_{t-1}, indicates the speed of adjustment of the model towards equilibrium when there are shocks in the system.

This procedure is applied irrespective of whether the regressors are purely I (0), purely I (1) or mutually cointegrated [49]. The unrestricted ECM minimizes the possibility of estimating spurious relations, whilst retaining the long run information, suitable for economic interpretation [50]. The investigation of the long run relationship using ARDL approach involves the estimation of equation 3, through an Unrestricted Error Correction Model (UECM). Since specification assumes that the disturbances are serially uncorrelated, the choice of appropriate lag order is important [48]. To obtain the appropriate lag length, a parsimonious UECM is estimated by introducing a lag length of three for the differenced variables and then, variables which are non-significant are dropped. A battery of diagnosis test can then be used to check the performance of the UECM [51, 52 in 48].

The ARDL Bounds Test

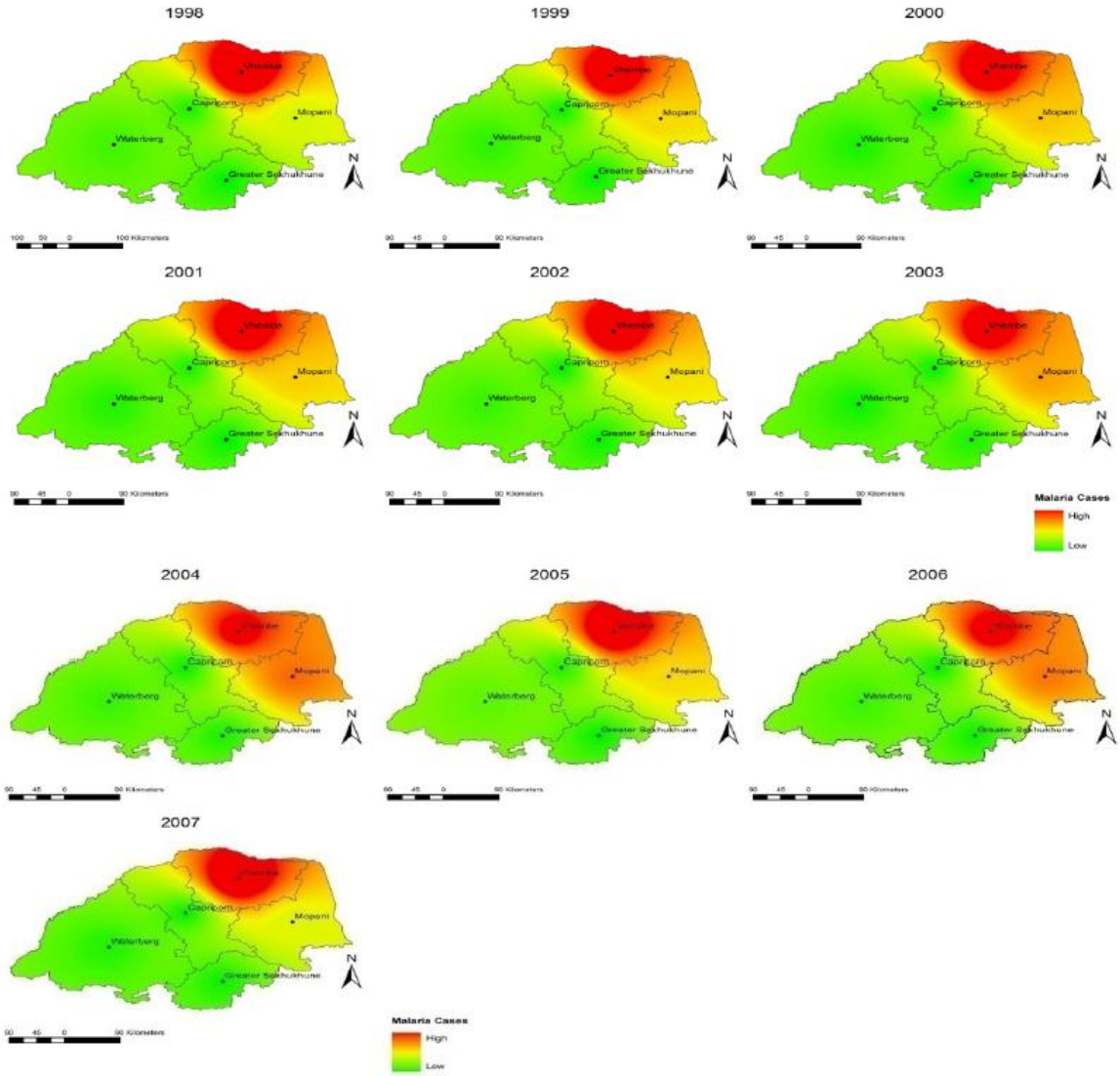
Testing the presence (or absence) of cointegration in variables implies a test of the existence of long run relationship. We use the Wald test (Bounds Test) to test the long-run relationship. Computed and critical bounds of the F-Statistic are provided by [49]. The F-statistics should lie outside the bounds for a long-run relationship to exist, but for short run, the coefficient of the error correction model (ECM) should be negative and statistically significant.

Results

In this section, we report spatial, correlation, time series, and short and long-run results in sequence:

Spatial distribution of Malaria in Limpopo Province

Figure 2: 10 years municipal and district spatial distribution of Malaria in Limpopo



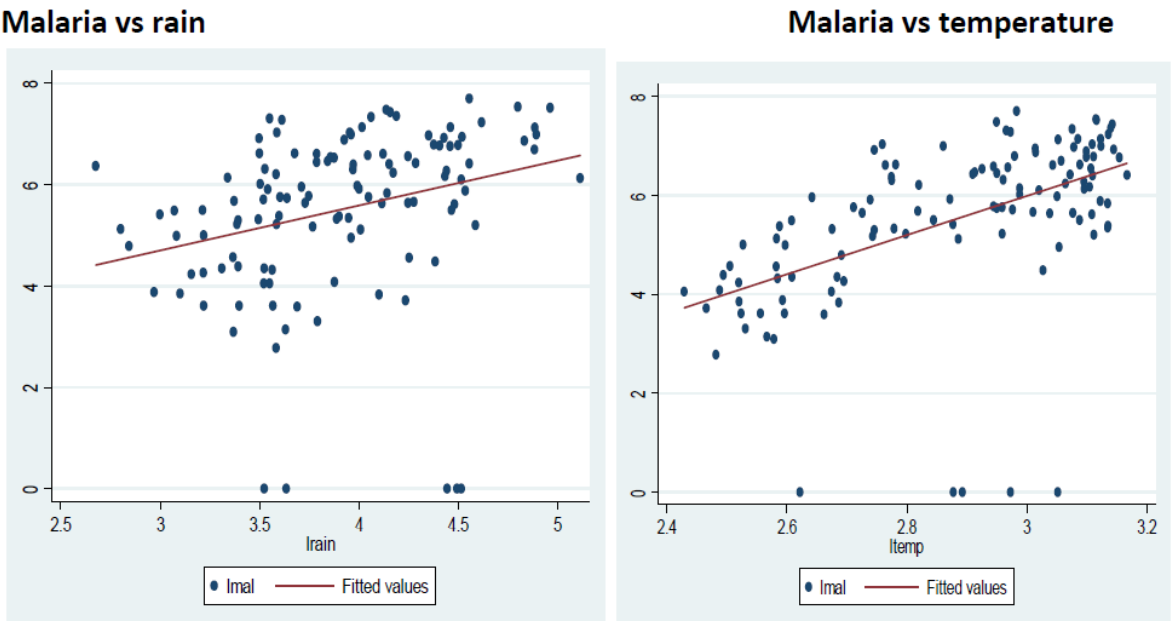
A visual analysis of district malaria cases in Limpopo province show that there has been a reduction in total cases from 1998 to 2007. However not all districts show the same change. Throughout the period of analysis, Vhembe district consistently shows more Malaria cases. Very few cases were reported in Capricorn, Waterberg and Greater Sekhukhune throughout the period of analysis. In Mopani district on the other hand, Malaria cases appear to be erratic (i.e. increasing and decreasing with no marked trend) as shown in the maps. Overall trend shows that, whereas there were fewer cases in 1998, this was followed by a somewhat increase from 1999 to 2006. More Malaria cases were recorded in 2004 and 2006, and were followed by a reduction in 2007.

Relationship between malaria cases with rainfall and temperature

It is sometimes recommended that *feel* techniques are first applied, to check correlation between the variables. This helps to roughly determine the expected signs of the regression model (positive or negative) "a priori". Here, the two independent variables (rainfall and temperature) are plotted against the independent variable (malaria). The closer and higher

the number of observations to the fitted values of malaria cases -the straight line, the higher and positive the correlation. It is negative otherwise. Figure 3 summarises the results of this exercise.

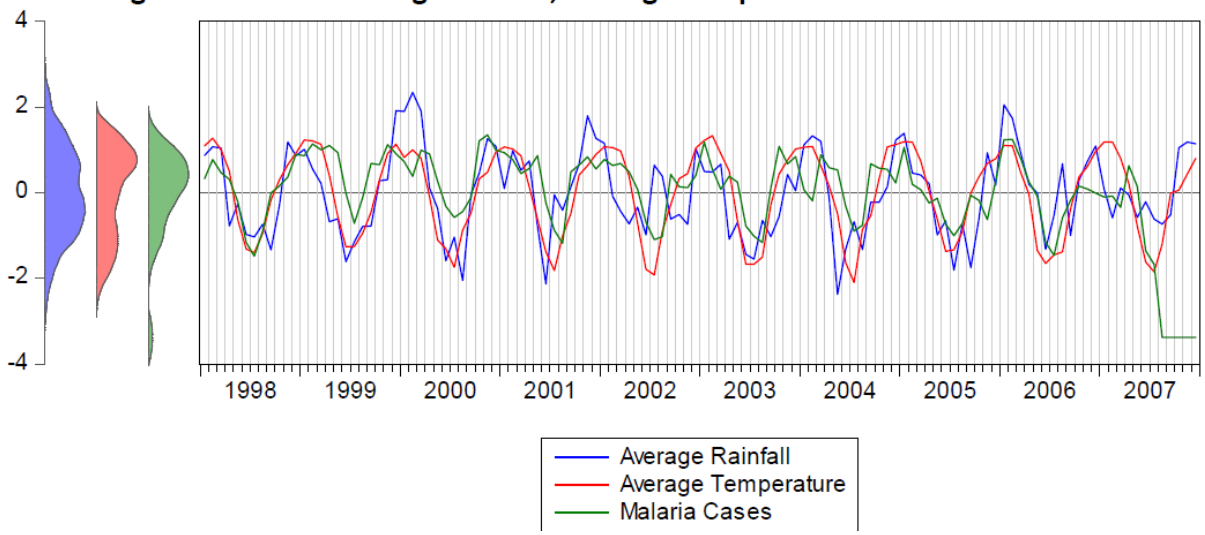
Figure 3: Correlation of rainfall and temperature with Malaria– A graphical outlook



From figure 3 above, both panels indicate a positive correlation with malaria. However, more observations are scattered away from the fitted line in the first panel (for rainfall) than in the second panel (for temperature). In this case, temperature shows a higher correlation than rainfall, as it relates to malaria cases. Since positive relationships are observed, we expect the regression coefficients to also match the above plots and thus carry a positive sign.

Figure 4 illustrates the trend relationship between average rainfall and average temperature as it relates to malaria.

Figure 4: Plot for Average Rainfall, Average Temperature and Malaria Cases



From the figure, it is clear that rainfall and temperature are important in explaining malaria in Limpopo. A casual look shows that very high rainfall does not increase malaria cases significantly (e.g. 1999, 2001 and 2005); increase in temperature is however consistent with increase in malaria cases. From a statistical viewpoint, the actual influence is further confirmed by the correlation coefficient as summarised in the table below.

Cross Correlation Matrix

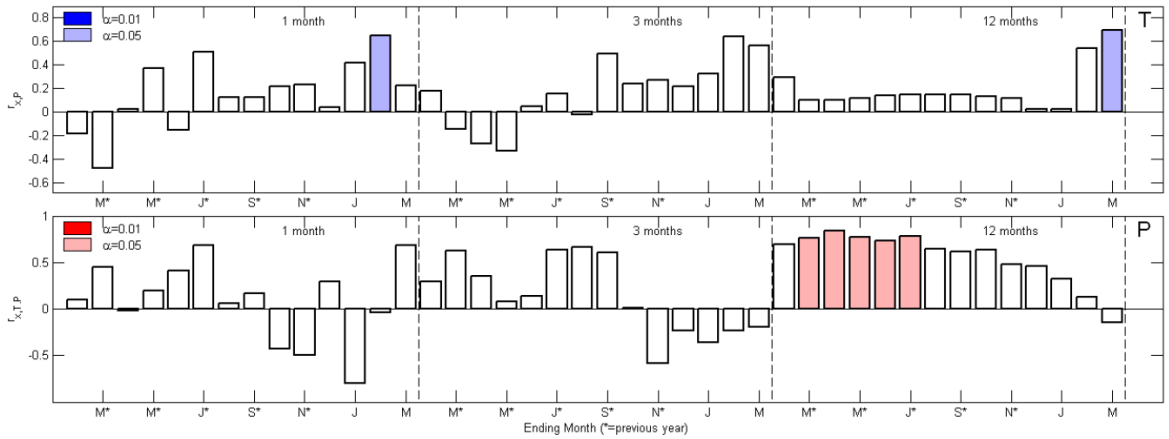
pwcorr lmal	**lrainltemp, star(5)**		
	Lmal	lrain	ltemp
lmal	1.0000		
lrain	0.2810*	1.0000	
ltemp	0.5212*	0.6656*	1.0000
*Indicates significance at 10%			

Correlation coefficient for temperature is 0.5212 and 0.2810 for rainfall. Rainfall and temperature are thus positively correlated with malaria. Temperature however shows stronger influence as compared to rainfall in relation to malaria.

Seasonal grouping

The occurrence of malaria cases in Limpopo province has been reported to be highly dependent on seasons (12). In order to assess the seasonal climate signal in malaria cases, temperature and precipitation have been partially correlated across each of 42 seasons and are depicted in Figure 5.

Figure 5: Correlations and partial correlations of malaria index with seasonalized climate variable at 0.05/0.01 significance (α).

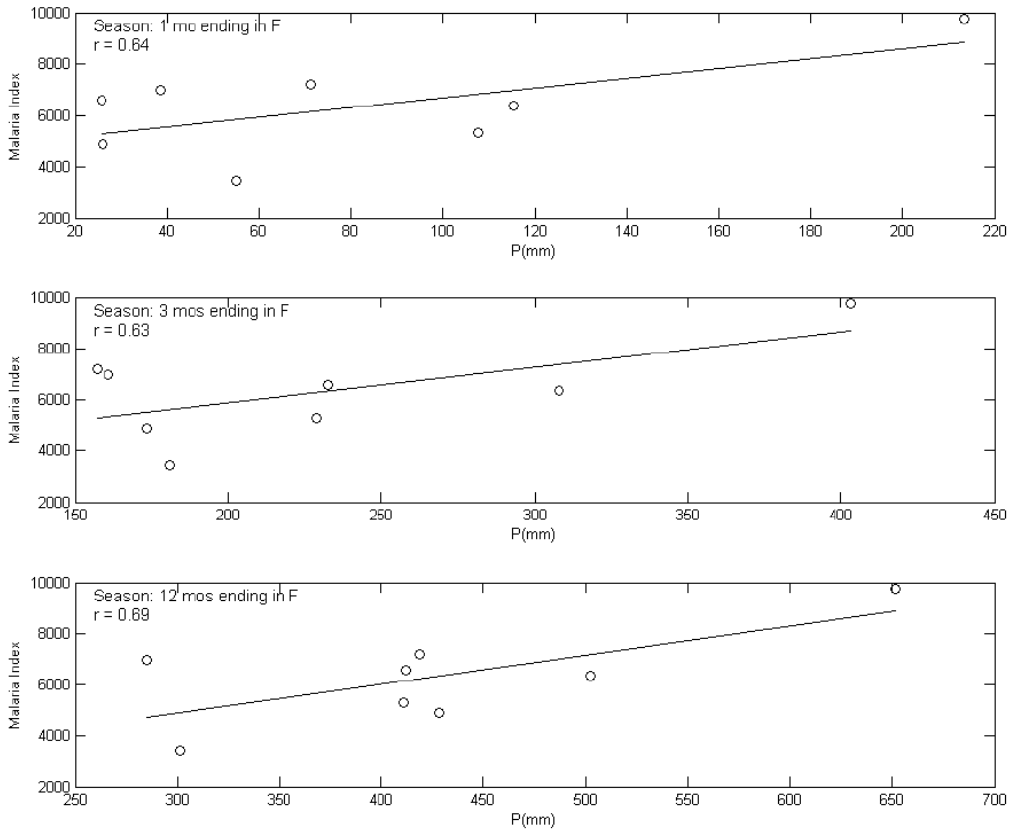


Top panel: Simple correlations with maximum temperature (T: primary climate variable). Bottom panel: Partial correlations (the influence of primary climate variable is removed) of malaria index with secondly climate variable (P: monthly total precipitation).

As shown in Figure 5, bars are proportional to simple correlations. Confidence interval is

approximate with no adjustment for serial correlation and non-normality. The approximate 95% confidence interval is set at $1.96/N^{0.5}$, where N is the sample size. Correlation of P and T is negative for all months and significant for more than six months. The negative correlation carries over to multi-month seasons as it is largely same-sign in individual months and this could be attributed to land-surface energy balance relationships, with variant allotment of energy to latent and sensible heat.

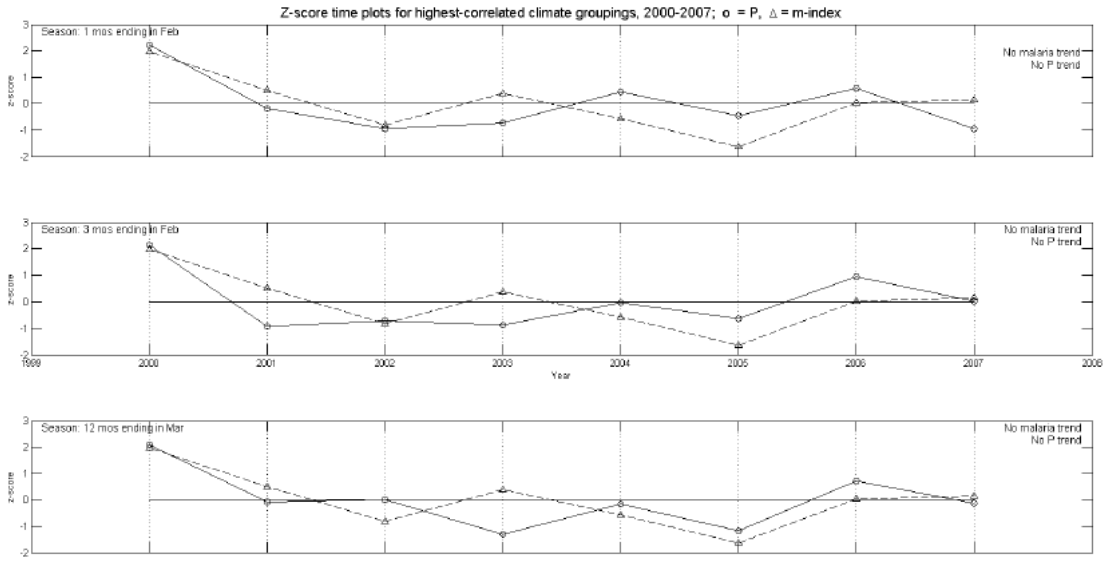
Figure 6: Scatter plots of malaria cases on the seasonalized primary climate variable (T) for the best seasons



Note: for figure 6, these are seasons with the highest correlation for each of the three specified season-lengths

Figure 6 depicts a fairly linear relationship between the primary climate model and the malaria cases during the best seasons (corresponding to the largest absolute correlations in the Top panel of Figure 6). The corresponding time series of the malaria cases and maximum temperature are depicted in Figure 7.

Figure 7: Time plots of malaria index (z-scores) and seasonalized maximum temperature (z-scores) for the highest correlated seasons.

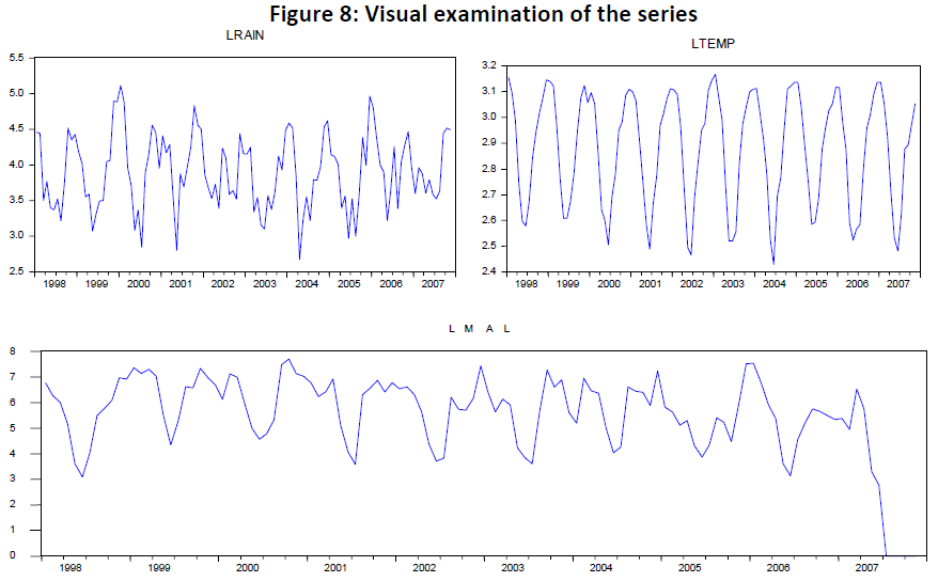


The time plots in Figure 4 enable the identification of years in which the agreement in malaria cases (hereafter malaria index) and maximum temperature is good or bad. Furthermore, a trend in one series and not in another might suggest that there could be some variable other than the primary climate variable for that season on the malaria index.

Stationarity (Unit Root Tests) Results

Before the tests for stationarity is done, informal test are used to test stationarity of the variables through plots. Informal method involves 'eye-balling' the graphs of the series plots. A stationarity series reverts back to its mean over time. Figure 9 below, summarises the graphical outlook of rainfall, temperature and malaria over time.

Figure 8: Visual examination of the series



Temperature and malaria seems to exhibit mean reversion (stationary), while rainfall is inconclusive. This can only be validated using the formal approaches below. To validate this, we formally carried out two different tests for stationarity, that is, Augmented Dickey-Fuller

(1981) and the Kwiatkowski, Phillips, Schmidst and Shin (1992) tests. The ADF test has the null of unit root, whereas the KPSS tests the null of stationarity. The null hypothesis is the existence of a unit root (stationary).

Table I: Unit root test results

Test	Log of Mala		Log of Rain		Log of Temp	
	Levels	First difference	Levels	First difference	Levels	First difference
ADF_{μ} ADF_{τ} $KPSS_{\mu}$ $KPSS_{\tau}$	 -4.283*** 0.620		-7.926*** 0.033***		-2.252 0.021***	-11.029***
Conclusion	Stationary at levels: I(0)		Stationary at levels: I(0)		Non-Stationary	Stationary at First Difference: I(1)
*, ** and *** means significance at 10%, 5% and 1% respectively. Source: Computed						

Results above indicate that temperature and malaria follow an autoregressive process with a unit root. The null hypothesis of a unit root could not be rejected for the variables, expressed in levels. Rainfall, on the other hand, is integrated for order 1 (non-stationary).

Short-run and long-run relationship between Malaria, RF and Temperature

Given that we have a combination of I (0) and I (1) variables, we opt to use the Unrestricted Error Correction Model (UECM), and employ the ARDL, or rather, the General-to-Specific (GETS) modelling procedure. Following similar procedure by [51, 52], UECM results are summarised in table III (from estimation of equation 3).

Table II: Unrestricted Error correction model

Variables	***Coefficient***	***Standard Error***
Constant	-3.158603	2.156372
D(LMALA(-2))	-0.473095	0.123357***
D(LRAIN(-1))	0.745233	0.248330***
D(LTEMP(-1))	4.343676	1.129335***
LMALA(-1)	0.249101	0.104620**
LRAIN(-1)	-0.499685	0.300813*
*, ** and *** means significance at 10%, 5% and 1% respectively. Source: Computed		

Rampsey RESET = 2.271595 (0.1350):

- *implying that Ramsey's RESET; Null hypothesis: Model has no omitted variable is not* rejected

White's test = 1.2668 (0.3869)

- *implying that White's test Null hypothesis of homoscedasticity is not rejected*

Breusch-Godfrey LM test = 0.868 (0.423)

- *implying that Breusch-Godfrey LM test: null hypothesis of No serial correlation is not rejected*

The R-Squared for the UECM is 50 percent, which indicates, a relatively good fit and satisfactory in this case. It means 50% of variation in malaria is explained by rainfall and temperature. This could have improved significantly with more independent variables. Durbin Watson statistic is 1.96, suggesting that the regression does not suffer from auto-correlation. The model passes the Ramsey RESET test of omitted variables; White's test of homoscedasticity and; Breusch-Godfrey LM test of serial correlation (refer to table III). Test of skewness and kurtosis of residuals (Jarque-Bera test) confirm that the residuals are normally distributed.

Bounds test (cointegration) results are presented in table IV.

Table III: Cointegration Properties

		Critical bounds (5%)	
Dependent variable	F-stat	Bottom	Top
d(lmala)	8.29	3.23	4.35
k=3 Source: Computed, critical bounds are obtained from [47]			

The F-statistic is outside the critical bounds (8.29 lies outside 4.35_{top} and 3.23_{bottom}). We therefore reject the null hypothesis of no cointegration at 5% significance level, and conclude that there exist a long-run relationship between malaria and the climate variables (rainfall and temperature).

The long-run relationship is reported in table V, while the short-run results are reported in table VI.

Table IV: Long run

Variable	Coefficient	Standard Error
C	-6.155823	0.0006***
LRAIN	-0.373873	0.2648
LTEMP	4.557185	0.0000***
*** means significance at 1% respectively.		

Table V: Short run

Variable	Coefficient	Standard Error
C	-0.080311	0.2668
D(LMALA(-2))	-0.231066	0.0047***
D(LMALA(-3))	-0.205359	0.0120**
D(LRAIN)	-0.263281	0.1509
D(LTEMP(-1))	4.784184	0.0000***
Ecm_{t-1}	0.005002	0.9783
** And *** means significance at 5% and 1% respectively.		

Cointegration results (based on ARDL error-correction model) show a long-run relationship between malaria and the climate variables. The long-run relationship between malaria and rainfall is -0.373873 (0.2648) and for temperature is 4.557185 (0.0000); while the short-run relationship between malaria and rainfall is -0.263281 (0.1509) and for temperature is 4.784184 (0.0000). In both instances, long and short-run temperature maintains a very high level of significance; while rainfall is insignificant in both long and short-run. This leads us to conclude that temperature is the main driver of malaria in Limpopo province. This result contrasts with [16] who found that rainfall was the only climatic variable significantly associated with monthly malaria cases, with no significant association between temperature and relative humidity for a neighbouring province (Mpumalanga).

We find coefficient of adjustment Ecm_{t-1} to be 0.005002 (0.9783), positive and statistically insignificant instead of negative as expected. The implication of this is that, should there be a sudden rise or fall in malaria transmission (i.e. a shock in the system), there will be no possibility of the system adjusting back to equilibrium. This means that every shock will imply beginning malaria interventions at a higher/lower equilibrium than before. Although the R-Squared performed fairly well, given the number of the explanatory variables (46%), the F-Statistic is highly significant at 1% F-statistic 7.81 (0.000).

Conclusions

We report GIS results of 5 districts (Capricorn, Greater Sekhukhune, Mopani, Waterberg and Vhembe) in Limpopo Province, based on the availability of clinical malaria data. Vhembe district consistently shows more Malaria cases, while very few cases were reported in Capricorn, Waterberg and Greater Sekhukhune throughout the period of analysis. In Mopani district on the other hand, Malaria cases appear to be erratic (i.e. increasing and decreasing with no marked trend). Many reasons can be put forwards to explain these spatial differences. Socio-economic reasons could range from access to health care; site of dwelling; to type of housing. Other possible reasons are patterns of migration in the area; Malaria control programs and; even climate change. Although determining the reason behind the differences in spatial distribution of Malaria cases is important, it is even more important to know which areas are more burdened by Malaria, so that control measure are targeted.

Positive correlation between malaria and climate variables (rainfall and temperature), has been reported elsewhere. Rainfall: Tibet [53], Sri Lanka [54]; rainfall and temperature [11, 23]; rainfall, temperature, humidity and vegetation cover [27]. The strength of the effect

seems to flow from humidity to temperature and rainfall. In this study, although rainfall and temperature are positively correlated with malaria, temperature shows stronger influence as compared to rainfall in relation to malaria. We find the temperature correlation coefficient to be 0.5212 as compared to rainfall correlation coefficient of 0.2810. This result is consistent with [53], who found the correlation coefficient for Tibet to be 0.518 and 0.348 for temperature and rainfall respectively, concluding that temperature had a greater influence on malaria and malaria responded quickly when temperature was varying.

Correlation of minimum temperature and malaria cases is positive and significant between May and June of the preceding year. Summing over months increases the correlation (albeit being non-significant) between maximum temperature and malaria cases. The 3-month sum exhibits a non-significant positive between May and September of the preceding year, but negative in the months between June and May of the current year. A 12-month sum has positive simple and non-significant correlations overall. The partial correlation of maximum temperature is only significant over a 12-month sum, ending in June of the previous year. The partial correlations vary (there is no consistent sign influence of precipitation on malaria index for a consecutive block of months) across summations over 1, 3, and 12-months; increasing only during the 12-month summation.

The OLS results suggest that the coefficient of rainfall and temperature are both positive and statistically significant as expected. This implies that increasing rainfall and temperature leads to an increase in malaria transmission. Specifically, it means that, holding all other variables constant, a 1% increase in rainfall will lead to 0.74% increase in malaria cases in Limpopo Province. Consequently, a 1% increase in temperature will translate to 4.34% increase in malaria cases. Here again, results confirm the high influence of temperature on malaria transmission in Limpopo province, South Africa. Regardless of temperature's greater influence, this result indicates that, increases in temperature and rainfall would create the conditions for malaria vectors to thrive in areas immediately surrounding their current distribution limits [55]. Thus, small changes in temperature and precipitation will boost the population of disease carrying mosquito, and result in increased malaria epidemics [56, 57]. In Ghana, a positive correlation was found to exist between malaria and climate elements [56]**.** Increases in temperature generally accelerate vector life cycles and also decreases the incubation period of the parasite [53, 58]. R-Squared of 50%, interpreted as the variation in malaria is explained by rainfall and temperature which corroborates the results found in KwaZulu Natal province, South Africa as 49.7% [11].

The coefficient of adjustment is statistically insignificant, instead of negative as expec**t**ed. This implies that, given any sudden rise or fall in malaria transmission, there will be no possibility of the system adjusting back to equilibrium. Thus, we will always begin at higher (or lower) malaria levels after an increase (or fall), and malaria interventions at a higher/lower equilibrium than before in the province.

It is interesting to find that malaria in Limpopo is found to be driven by temperature, while in Mpumalanga (the neighbouring province), malaria is driven by rainfall [16].

This paper has provided useful basis upon which other studies can work. We recommend a probe to ascertain the thresholds of temperature (and rainfall) under which malaria cases are probable. In addition, there is need to ascertain the reason why malaria in Limpopo is found to be driven by temperature, but by rainfall in the neighbouring province. We find that

malaria pressure varies in different districts. Thus, in terms of combating malaria, the requisite public health response to climate-malaria interaction, should be part of the key functions that exist within public health systems in Limpopo Province. We recommend the development and enhancement of early warning systems for malaria at district level. For improved control in regions where malaria is endemic, the World Health Organisation (WHO) advocates the development of integrated malaria early warning systems, based on vulnerability assessment, seasonal climate forecasts, weather and environmental monitoring and case surveillance. There is therefore need to strengthen collaboration, partnership and response integration with other principle sectors, e.g. the social sector and the meteorological sector in this regard in Limpopo Province. The requisite public health response to climate-malaria should be part of the key functions of the public health systems in Limpopo Province, in terms of combating malaria.

The study only considered rainfall and temperature as the main drivers to malaria transmission. While is recognizes the importance of looking at mosquito characteristic in the province; this is limited by the lack of data on Entomological Inoculation Rates (EIR) and Human Biting Rate (HBR) of mosquitoes.

References

1. Reiter P: **Global warming and malaria: knowing the horse before hitching the cart.** *Malaria Journal,* 2008, **7** (Suppl 1):S3 doi:10.1186/1475-2875-7-S1-S3
2. Patz JA, Hulme M, Rosenzweig C, Mitchell TD, Goldberg RA, Githeko AK, Lele S, McMichael AJ, Le Sueur D: **Climate change: Regional warming and malaria resurgence.** *Nature* 2002, **420:**627-628. Discussion 628.
3. Pascual M., Ahumada JA, Chaves LF, Rodo´ X and Bouma M: **Malaria resurgence in the East African highlands: Temperature trends revisited.** *Proc. Natl. Acad. Sci.* 2006, *USA* **103,** 5829–5834.
4. Thomson MC, Mason SJ., Phindela T, Connor SJ: **Use of Rainfall and Sea Surface Temperature Monitoring for Malaria Early Warning in Botswana**: *Am. J. Trop. Med. Hyg.,* 73(1), 2005, pp. 214–221
5. Patz JA, Olson SH: **Malaria risk and temperature: Influences from global climate change and local land use practices.** *Ann Trop Med Parasitol.* 2006 100(5-6):535-49.
6. Connor S, Thomson M, Molyneux D, 1999. **Forecasting and prevention of epidemic malaria: new perspectives on an old problem**. *Parassitologia 41:* 439–448.
7. Alemu A**,** Abebe G**,** Tsegaye W**,** Golassa Lemu**: Climatic variables and malaria transmission dynamics in Jimma town, South West Ethiopia.** *Parasites & Vectors* 2011, **4**:30
8. Poveda G, Rojas W, Quinones ML, Velez ID, Mantilla RI, Ruiz D, Zuuaga JS, Rua GL: **Coupling between annual and ENSO timescales in the malaria-climate association in Colombia.** *Environ. Health Perspect.* 2001, **109:**489–493.
9. Ebi KL, Hartman J, Chan N, McConnell KJ, Schlesinger M, Weyant J. **Climate suitability for stable malaria transmission in Zimbabwe under different climate change scenarios.** *Clim. Change* 2005 **73,** 375–393.
10. Sharp BL, Ngxongo S, Botha MJ, Ridl FC, Le Sueur D: **An analysis of 10 years of retrospective malaria data from the KwaZulu areas of Natal**. *South African Journal of Science* 1988, 84,102–106.
11. Craig MH, Kleinschmidt I, Nawn JB, Sueur DL, Sharp BL: **Exploring 30 years of malaria case data in KwaZulu-Natal, South Africa: Part I. The impact of climatic factors**; *Tropical Medicine and International Health* 2004 Vol 9 no 12 1247–1257
12. Bouma MJ. van der Kaay HJ: **The El Niño/Southern Oscillation and the historic malaria epidemics on the Indian subcontinent and Sri Lanka: an early warning system for future epidemics?** *Trop. Med. Int. Health* 1996 1, 86–96.
13. Lindblade KA, Walker ED, Onapa AW, Katungu J, Wilson ML: **Highland malaria in Uganda: prospective analysis of an epidemic associated with El Nino** *Trans. R. Soc. Trop. Med. Hyg*. 1999, **93,** 480–487.
14. Blumberg L, Frean J: **Malaria control in South-Africa – challenges and successes**. SAMJ 2007, 97:1193-1197.
15. Gerritsen AAM. Kruger P, Schim van der Loeff MF, Grobusch MP): **Malaria incidence in Limpopo Province, South Africa, 1998–2007.** *Malaria Journal* 2008, **7**:162 doi:10.1186/1475-2875-7-162
16. Ngomane L, de Jager C: **Changes in malaria morbidity and mortality in Mpumalanga Province, South Africa (2001- 2009): A retrospective study**. *Malaria Journal* 2012, 11:19

17. Landman WA. Goddard L: **Predicting southern African summer rainfall using a combination of MOS and perfect prognosis.** *Geophys. Res. Lett.* 2005. 32, L15809, doi: 10.1029/2005GL022910.
18. Kondo H, Seo N, Yasuda T, Hasizume M, Koido Y, Ninomiya N, Yamamoto Y: **Post-flood infectious diseases in Mozambique.** *PrehospDisast Med.* 2002; 17(3):126–133.
19. Bohle HG, Downing TE, Watts MJ: **Climate change and social vulnerability: Toward a sociology and geography of food insecurity. Global Environmental Change**. 1994, 4(1) 37–48.
20. van Lieshout M, Kovats RS, Livermore MTJ, Martens P: **Climate change and malaria: Analysis of the SRES climate and socio-economic scenarios.** *Global Environmental Change,* 2004. 14(1) 87-99
21. WHO. **Malaria Early Warning Systems, Concepts, Indicators and Partners. A framework for Field Research in Africa.** 1–-84, Report no. WHO/CDS/RBM/ 2001.32 (WHO/Roll Back Malaria/Technical Support Network for Prevention and Control of Malaria, Geneva, 2001)
http://mosquito.who.int/ cmc_upload/0/000/014/807/mews2.pdfl
22. Haines A, McMichael AJ, Epstein PR : **Environment and health: 2. Global climate change and health.** *CMAJ* 2000. 163 (6)
23. Githeko AK, Ndegwa W: **Predicting Malaria Epidemics in the Kenyan Highlands Using Climate Data: A Tool for Decision Makers.** *Global Change & Human Health* 2001, 2, (1), 54-63, DOI: 10.1023/A:1011943131643
24. Nkomo JC, Nyong AO, Kulindwa, K: **The Impacts of Climate Change on Africa.** *Final Draft Submitted to: The Stern Review on the Economics of Climate Change. July 2006*
25. Ngomane L, de Jager C: **Changes in malaria morbidity and mortality in Mpumalanga Province, South Africa (2001- 2009): A retrospective study**. *Malaria Journal* 2012, 11:19
26. Paaijmans KP, Blanford S, Bell AS, Blanford JI, Read AF. Thomas MB: **Influence of climate on malaria transmission depends on daily temperature variation.** PNAS 2010 107 (34) 15135–15139
27. Haque U, Hashizume M, Glass GE, Dewan AM, Overgaard HJ, Yamamoto T: **The Role of Climate Variability in the Spread of Malaria in Bangladeshi Highlands.** *PLoS ONE* 2010 5(12): e14341. doi:10.1371/journal.pone.0014341
28. Patz JA, Olson SH: **Malaria risk and temperature: Influences from global climate change and local land use practices** *Ann Trop Med Parasitol.* 2006 Jul-Sep; 100(5-6):535-49.
29. Patz JA, Campbell-Lendrum D, Holloway T, Foley JA: **Impact of regional climate change on human health.** *Nature* 2005 438, 310-317 | doi: 10.1038/nature04188
30. Naqvi ZR: **Using remote sensing to assess potential impacts of hurricanes on mosquito habitat formation: Investigating the mechanisms for interrelationship between climate and the incidence of vector-borne diseases.** Theses/Dissertations-Environmental Studies. Submitted to the Graduate Faculty of Baylor University. December 2009
31. Atul K, Nettleman M: **Global warming and infectious disease** *Archives of Medical Research* 2005, 36: 689-696.

32. Harrus S, Baneth G: **Drivers for the emergence and reemergence of vector-borne protozoa land bacterial diseases.** *International Journal of Parasitology,* 2005. 35(11-12):1309-1318.
33. IOM 2008: **Vector-borne diseases: Understanding the environmental, human health and ecological connections**. Washington, DC: The National Academies Press.
34. Relman DA, Hamburg MA, Choffnes ER, Mack A: **Global Climate Change and Extreme Weather Events: Understanding the Contributions to Infectious Disease Emergence: Workshop Summary**. ISBN: 0-309-12403-4, 304 pages, 6 x 9, (2008)
35. van Lieshout M, Kovats RS, Livermore MTJ, Martens P: **Climate change and malaria: Analysis of the SRES climate and socio-economic scenarios.** *Global Environmental Change,* 2004. 14(1) 87-99
36. Kovats RS, Haines A: **Global climate change and health: recent findings and future steps.** *Canadian Medical Association Journal*. 2005; 172(4):501–502.
37. Hanafi-Bojd AA, Vatandoost A, Oshaghi MA, Charrahy Z, Haghdoost AA, Zamani, G, Abedi F, Sedaghat MM, Soltani M, Shahi M, Raeisi A: **Spatial analysis and mapping of malaria risk in an endemic area, south of Iran: A GIS based decision making for planning of control.** *Acta Tropical,* 2012, 122(1), pp. 132-137.
38. Jorgensen P, Nambanya S, Gopinath D, Hongvanthong B, Luangphengsouk K, Bell D, Phompida S, Phetsouvanh R: **High heterogeneity in *Plasmodium falciparum* risk illustrates the need for detailed mapping to guide resource allocation: a new malaria risk map of the Lao People's Democratic Republic**. *Malaria Journal,* 2010, 9(59), doi:10.1186/1475-2875-9-59.
39. Messina JP, Taylor SM, Meshnick SR, Linke**,** AM, Tshefu AK, Atua B, Mwandagalirwa K**,** Emch M: **Population, behavioural and environmental drivers of malaria prevalence in the Democratic Republic of Congo.** *Malaria Journal* **2011***,* 10(161), doi:10.1186/1475-2875-10-161. [Accessed 20 May 2012]. Available at: <http://www.malariajournal.com/content/10/1/161>.
40. Baltas E: **Spatial distribution of climatic indices in northern Greece**. *Meteorological Applications* 2007*, (*14) 69-78.
41. Wilks DS: **Statistical methods in the atmospheric sciences**. International Geophysics Series. 1995; 59: 469.
42. Mardia K, Kent J, Bibby J: **Multivariate analysis*. Academic Press, London***. 1979; 518.
43. Panofsky HA, Brier GW: **Some applications of statistics to Meteorology**. The Pennsylvania State University, University Park, Pennsylvania. 1968; 224.
44. Gupta R and Komen K: **Time Aggregation and the Contradictions with Causal Relationships: Can Economic Theory Come to the Rescue?** *Studies in Economics and Econometrics Journal (S.E.E) 2009.* 33 (2009) 13 – 24 ISSN: 03796205
45. Komen DK, Kapunda, SM: **Macroeconomic determinants of poverty reduction in the era of globalisation in Kenya: Policy implications.** *African Journal of Economic policy* 2006. 13 (2). December 2006. ISSN 1116-4875.
46. Pesaran MH, Shin Y (1999): **An autoregressive distributed lag modelling approach to cointegration analysis**. In: Strøm S, editor. Econometrics and economic theory in the twentieth century: the Ragnar Frisch Centennial Symposium.Cambridge: Cambridge University Press.
47. Narayan PK: **Reformulating critical values for the bounds F-statistics approach to cointegration: An application to the tourism demand model for Fiji**. Department of Economics 2004 Discussion Papers no.02/04. Monash University, Melbourne, Australia.

48. Sultan R: S**hort-run and long-run elasticities of gasoline demand in Mauritius: an ARDL bounds test approach.** *Journal of Emerging Trends in Economics and Management Sciences* 2010: 1 (2): 90-95
49. Pesaran MH, Shin Y, Smith RJ: **Bounds testing approaches to the analysis of level relationships.** *Journal of Applied Econometrics 2001, 16: 289-326*
50. Greenidge K, Holder C, Mayers S: **Estimating the Size of the informal economy in Barbados**. *Business, finance and emerging economies 2009 4(1)*
51. Akinboade OA, Ziramba E, Kumo WL: **The demand for gasoline in South Africa: An empirical analysis using co-integration techniques.** *Energy Economics* 2008, 30: 3222–3229
52. Hendry DF, Pagan A, Sargan JD: **Dynamic specification**. In: Griliches, Z. Intrilligator, M. (Eds.), Handbook of Econometrics, 1984: vol. 2. North Holland, Amsterdam.
53. Huang F**,** Zhou S**,** Zhang S, Wang H, Tang L: **Temporal correlation analysis between malaria and meteorological factors in Motuo County, Tibet**. *Malaria Journal* 2011, **10**:54 doi:10.1186/1475-2875-10-54
54. Briët OJT, Vounatsou P: **Gunawardena DM,** Galappaththy GNL, Amerasinghe PH: **Temporal correlation between malaria and rainfall in Sri Lanka.** *Malaria Journal*. 2008; 7: 77.
55. Epstein PR, Diaz HF, Elias SA, Grabherr G, Graham NE, Martens WJM, Mosley-Thompson E, Susskind J: **Biological and physical signs of climate change: focus on mosquito-borne diseases.** *Bulletin of the American Meteorological Society* 1997*,* **78,** 409-417.
56. Nkomo JC, Nyong AO, Kulindwa, K: **The Impacts of Climate Change on Africa.** Final Draft Submitted to: The Stern Review on the Economics of Climate Change. **July 2006**
57. Lindsa**y** SW, Martens WJM: **Malaria in the African highlands: past, present and future.** *Bulletin of the World Health Organization*, 1998 **76,** 33-45.
58. Kovats RS, Martens P 2000: Human **health. In: Assessment of Potential Effects and Adaptations for Climate Change in Europe: The Europe ACACIA Project** [Parry, M.L. (ed.)]. Jackson Environment Institute, University of East Anglia, Norwich, United Kingdom, pp. 227-242.

[I] Development of a stochastic differential equation model capable of simulating malaria dynamics in Senegal[1]

Overview

A dynamical stochastic differential equation (SDE) model of malaria has been constructed and validated against observed malaria data for Senegal. Due to the lack of other reliable long-term time series of malaria incidence in any of the two other countries in QWeCi, namely Malawi and Ghana, a colloboration was established with partners external to the project that might provide the quality data for malaria needed to construct such a dynamical model. This as allowed the model to be validated against an extraordinary dataset existent for Senegal, probably the most reliable record of its sort in West Africa.

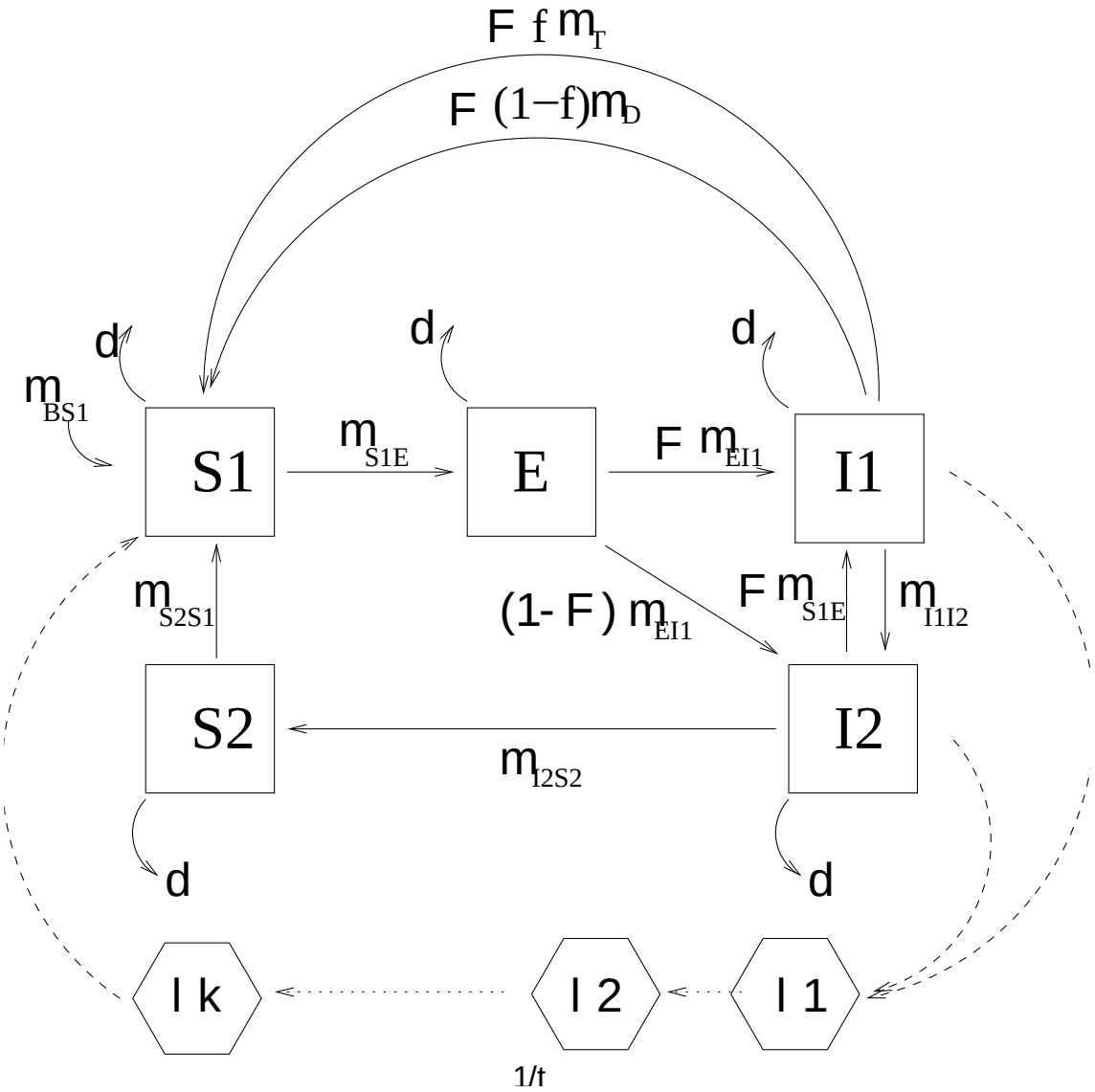


Figure 1. *Schematic flow diagram of the preliminary ODE system, constructed within the framework of the VIC3 malaria model for Dielmo (Senegal).*

[1] Work in progress – only a partial report has been included here.

The malaria model structure is illustrated in Figure 1. According to the characteristics of this location in Senegal, we divided humans into five distinct classes: S1, susceptible to infection; E, exposed (i.e., carrying a latent infection but not yet infectious); I1, infected and infectious; I2, possessing an asymptomatic infection which is expected to be less infectious than I1; S2, recovered and having some load of parasites (not completely cleared), not enough to be detected but allow them to infect mosquitoes (expected to be less infectious than I2). Mosquito-parasite classes are λ (force of infection at previous time t-s) and λκ (force of infection at time t). The possibility of transition between class X and Y is denoted by a solid arrow, with the corresponding rate written as μXY. The dotted arrows represent interactions between the human and mosquito stages of the parasite. The model is formalized by a set of 14 equations described in full detail in Laneri, Rodó et al (submitted). The approach used to model vector dynamics was through a delayed equation for the force of infection μS1E(t) at time t when an infected mosquito bites a human, taking into account that the transmission from human to mosquito occurred at a previous time s. Therefore, we modelled the duration of parasite life cycle inside mosquito and vector survival as a gamma distribution.

The preliminary ODE system constructed for the Dielmo dataset, was then adjusted in the framework of VIC3 environment at IC3 and the MIF POMP iterative approach used to estimate and cross-validate the parameters of the model.

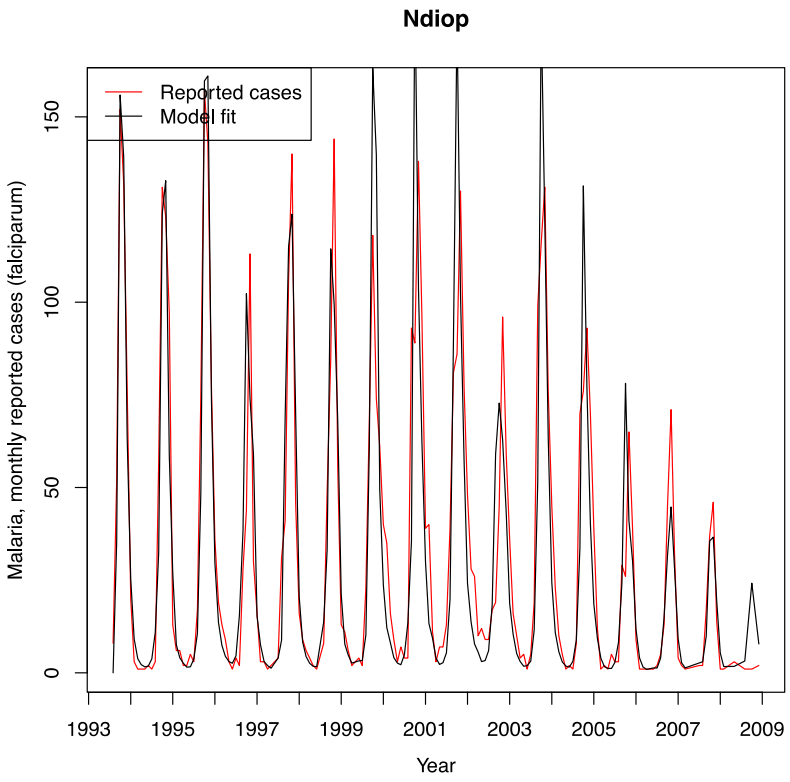
Model validation

The model is capable of reasonably capturing both the seasonal dynamics in this malaria endemic regime and the different shifts in mean malaria states after the changes produced by the drug treatment periods. The model behaves properly and successfully reproduces the population dynamics of malaria in two pilot villages (Dielmo and Ndiop*, Figure 2*). When rainfall (R) was incorporated as an extrinsic driver the resulting models notably gained in skill with regard to simulations without R as a covariate. Similarly, the different drug treatments implemented in the region were properly integrated in the simulations, in their corresponding intervals. Seasonal cycles of malaria in the two villages were successfully reproduced also, with the incorporation of temperature anomalies. Runs were also successfully validated for recent out-of-fit data.

Future work

Next work to be conducted will center on trying to understand the systematic bias in the variability at both maxima and minima malaria cases, as well as trying to perform skilful predictions with VIC3 at seasonal to decadal timescales. These simulations will be then compared with persistence projections arising from observations themselves. We will also consider the potential addition of other extrinsic drivers in the search for longer lead times for prediction. The model will be tested for other regions in Africa, where malaria conditions are similar to those in the fringe semi-desert areas of Senegal.

a)



b)

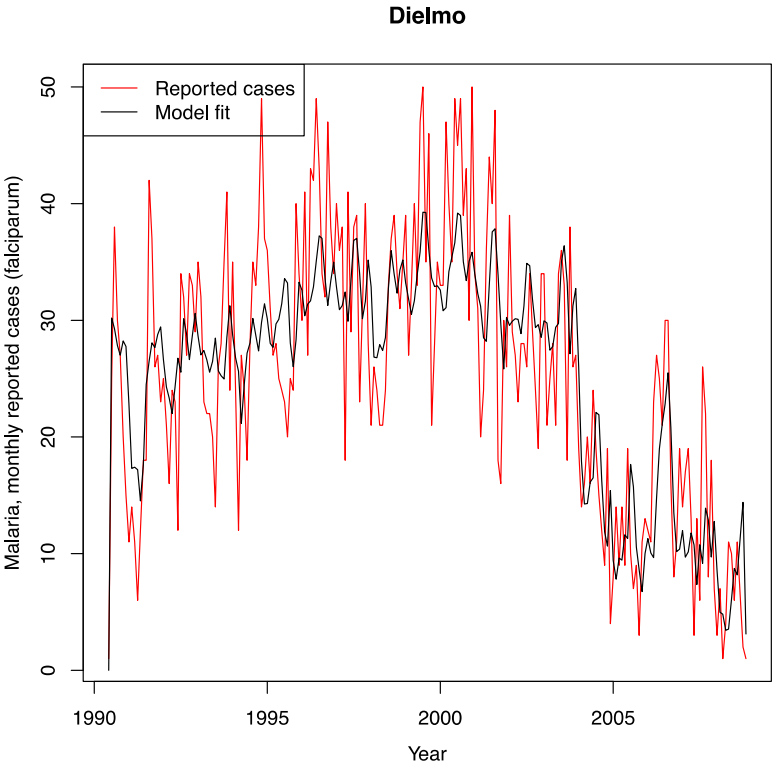


Figure 2. *Resulting fittings from VIC3 to malaria time series in a) Ndiop and b) Dielmo, with rainfall as a covariate. Drug treatment intervals were also integrated in the simulations.*

General Conclusion

This report summarizes the various modelling efforts carried out by the various QWeCI partners to model malaria and RVF risk for different regions in Africa. During the QWeCI project new models have been developed and validated and the modelling effort and the on-going collaborations built within QWeCI will continue beyond the scope of the project. Amongst the significant modelling contributions we can cite:

- [A] Improved parameter settings for the standard version of the LMM and validation with observations and other malaria endemicity estimates
- [B] New parameter settings of the LMM (LMM2010), simulations driven by Regional Climate Change projections to estimate the impact of climate on malaria transmission in the future
- [C] The development and validation of the VECTRI malaria model
- [D] The development and validation of a statistical malaria model for Malawi which includes the effects of climate and socio-economic parameters (poverty...)
- [E] Statistical forecasting model of RVF vectors for northern Senegal that might be used as an early warning system
- [F] A pilot integrated multi-agent system for RVF in Senegal. This corroborates findings from other WP/ published studies
- [G] A mixed RVF statistical model for Kenya that should be tested over West Africa in a near future
- [H] Investigation of the impact of climate on malaria in the Limpopo province of South Africa before developing an early warning system prototype. Prototype of a high spatial resolution malaria model using a climate envelope based approach and satellite estimates (not shown in the current report but currently in progress).
- [I] Development of a stochastic differential equation model capable of simulating malaria dynamics in Senegal (work in progress; partial report included here).