

RESEARCH ARTICLE

Uncertainty in malaria simulations in the highlands of Kenya: Relative contributions of model parameter setting, driving climate and initial condition errors

Adrian M. Tompkins^{1*}, Madeleine C. Thomson²

1 Earth System Physics, The Abdus Salam International Centre for Theoretical Physics (ICTP), Strada Costiera 11, Trieste, Italy, **2** International Research Institute for Climate and Society, Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, United States of America

* tompkins@ictp.it



OPEN ACCESS

Citation: Tompkins AM, Thomson MC (2018) Uncertainty in malaria simulations in the highlands of Kenya: Relative contributions of model parameter setting, driving climate and initial condition errors. PLoS ONE 13(9): e0200638. <https://doi.org/10.1371/journal.pone.0200638>

Editor: Ana Paula Arez, Universidade Nova de Lisboa Instituto de Higiene e Medicina Tropical, PORTUGAL

Received: November 20, 2017

Accepted: June 29, 2018

Published: September 26, 2018

Copyright: © 2018 Tompkins, Thomson. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The VECTRI mode source code is located at (<https://gitlab.com/tompkins/vectri>). All model experiment output is available at Pangaea at (<https://doi.pangaea.de/10.1594/PANGAEA.890161>). The health data set is available at (<https://doi.org/10.1371/journal.pone.0024524>). The meteorological data set is available by contacting Kenya Meteorological Department, P. O. Box 30259 - 00100, GPO, Nairobi Phone:254-20-3867880, 254-20-3876957, 254-20-3873682 Fax 254-2-3876955; e-mail: director@meteo.go.ke;

Abstract

In this study, experiments are conducted to gauge the relative importance of model, initial condition, and driving climate uncertainty for simulations of malaria transmission at a high-land plantation in Kericho, Kenya. A genetic algorithm calibrates each of these three factors within their assessed prior uncertainty in turn to see which allows the best fit to a timeseries of confirmed cases. It is shown that for high altitude locations close to the threshold for transmission, the spatial representativeness uncertainty for climate, in particular temperature, dominates the uncertainty due to model parameter settings. Initial condition uncertainty plays little role after the first two years, and is thus important in the early warning system context, but negligible for decadal and climate change investigations. Thus, while reducing uncertainty in the model parameters would improve the quality of the simulations, the uncertainty in the temperature driving data is critical. It is emphasized that this result is a function of the mean climate of the location itself, and it is shown that model uncertainty would be relatively more important at warmer, lower altitude locations.

1 Introduction

Any prediction for the future, whether for days, years or centuries, should be accompanied with an estimate of its uncertainty or confidence [1]. This is valid for short range and seasonal forecasts and long-term projections of climate-sensitive health outcomes. Understanding this uncertainty is critical if forecast information is to contribute to health decision making effectively. Incorporating uncertainty into the decision-making process should help to ensure forecasts are primarily used when confidence is high, thus minimizing potential losses from poor forecasts. Communicating uncertainties effectively is extremely challenging and even in the field of operational weather forecasting, the use of uncertainty information is rarely optimal. In health applications, the necessity to understand the decision context adds an additional layer of complexity [2, 3].

info@meteo.go.ke (<http://www.meteo.go.ke/index.php?q=datarequest>). The authors had no special privileges to the data.

Funding: The author(s) received no specific funding for this work (only travel was supported by NIH).

Competing interests: The authors have declared that no competing interests exist.

In this paper uncertainty is considered in the context of a specific modeling system in which climate information is used to drive a mathematical, dynamical model for disease, in this case to simulate observed clinical malaria cases for a highland tea plantation location in East Africa. In such a model hierarchy, it is useful and usual to separate the uncertainty into a cascade of contributions represented schematically in Fig 1. Broadly speaking, uncertainty can be divided into categories of initial condition, boundary condition and model uncertainty.

Any dynamical model requires knowledge of the initial state. In climate and weather applications, data is collected each day via the global telecommunication system (GTS) for assimilation, however no near real-time data on entomological or epidemiological conditions exists to assess the initial state for malaria transmission simulations. Recent developments in digital health information management and surveillance systems across the continent offer some hope of improved health data assimilation systems [4], but to date have not been used for this purpose and are difficult to access outside health ministries. Monitoring of vector density directly is presently not possible on a large-scale in an operational context. A pilot forecasting system estimated entomological conditions using climate information to drive a malaria model to circumvent this difficulty [5]. Even if not available in real-time, surface hydrology data may also be used to evaluate the modeling system output. However, large hydrological, entomological and epidemiological data uncertainties will hinder the evaluation of alternative modeling systems. Research experiments and surveys of vector biting rates or malaria prevalence, as well as health ministry records of clinical cases are also subject to errors and uncertainties.

The climate information and other environmental and socio-economic settings that essentially provide the boundary conditions for the disease model may temporally vary over a range of timescales. The climate information may be provided by a weather forecast model, a climate model, or directly from observations, depending on the context of the simulation. In this paper, we employ observational data, which will be subject to measurement errors. These are likely to be dwarfed by the representativeness uncertainty, whereby the highly heterogeneous environment is summarized with a point value at the station location (or by an area-averaged satellite measurement or model grid output if these are alternatively used). A variation in altitude of 150 meters across a health clinic catchment is adequate for a variation in temperature of 1 K, which can have a significant impact on transmission rates, especially at mean

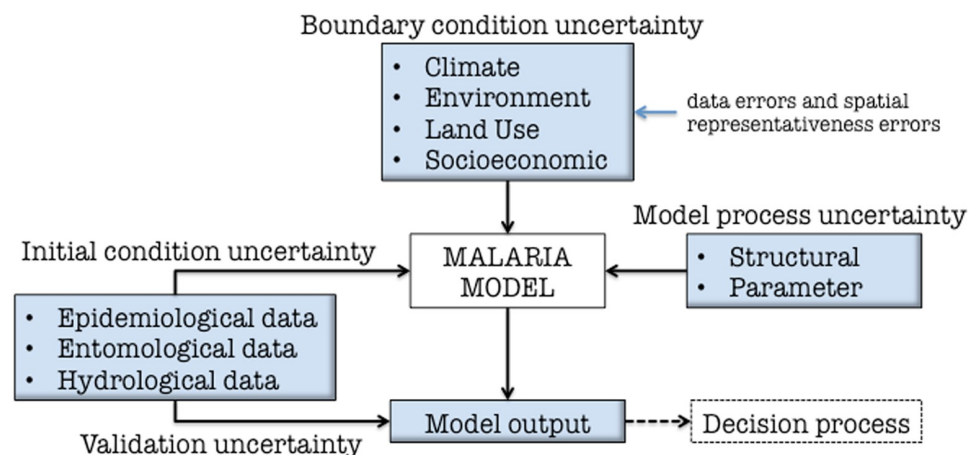


Fig 1. Schematic of the potential sources of uncertainty when using a weather-sensitive disease model to simulate observed health outcomes.

<https://doi.org/10.1371/journal.pone.0200638.g001>

temperatures approaching the lower or upper temperature bounds for transmission [6]. Obviously, random errors and biases in a model-produced climate dataset may be far larger.

Another source of uncertainties pertains to the dynamical disease model itself, which can suffer from structural uncertainty, that is, errors in the model formulation through missing or poorly represented processes. Given a specific model structure, there is also uncertainty with regard to the setting of model parameters. These parameters may relate to entomological processes, such as the degree day length of the sporogonic cycle or the temperature sensitivity of vector mortality, for example, or other aspects relating to the environmental and socio-economic setting.

In weather prediction and climate projection there is a similar cascade of uncertainty, with errors in the initial and boundary conditions in addition to the model system itself [7]. In this field, uncertainty is systematically accounted for by employing (preferably large) ensembles of simulations [7–9]. These often consist of multiple modeling systems (different models and/or parameter settings/perturbations of a single model) subject to multiple initial and boundary conditions [10–13], although recent work points out that ensemble sizes are often inadequate to correctly separate initial condition and model structural uncertainty [14]. In a study using large climate model ensembles, the parameter perturbation approach involved defining a prior estimate of parameter uncertainty from expert opinion or field experiments, and then applying a certain model performance criterion to reject some parameter combinations a posteriori [9]. The use of ensembles is sometimes referred to as a forward modeling approach, incorporating the cascade of uncertainty at each stage of modeling [8, 15, 16]. There are, however, limitations to this and alternative formal statistical approaches to assessing uncertainty [17].

In health research, large multi-model, multi-simulation ensemble methods for accounting for uncertainty are relatively rare, or if employed, often sample a single aspect of uncertainty such as that associated with the weather and climate driving information [5, 18, 19]. As an example, a recent investigation into the climate change impact on malaria in West Africa uses a single integration of two climate models to drive a single implementation of the malaria transmission model under a single greenhouse gas emission scenario [20]. Thus most aspects of climate initial condition, boundary conditions and model structural uncertainty are neglected, limiting the relevance of the study to inform health policy.

One bottleneck to considering model structural uncertainty in health applications in a multi-model ensemble framework is the disparate nature of the health modeling systems themselves. Relative to global climate models, which solve approximately the same underlying equation set and represent a similar set of physical processes using parameterization schemes, disease models may be structurally much more diverse. For example, models for malaria range from statistical, to compartmental SEIR models that focus on the disease in the human host, to more comprehensive dynamical models that include the vector-parasite biology. Indeed, these diverse models may not even predict a common set of health indicators. Taking malaria models as an example, some models may be simple R_0 models, or statistical threshold models for climate suitability, contrasted with mathematical Susceptible-Exposed-Infectious (SEI) compartmental models or full vector-parasite life cycle process dynamical models that may instead simulate infectious biting rates or clinical cases expected per 1000 population. Such disparities confound the attempt to construct multi-model approaches to assess model structural uncertainty, although some recent multi-model assessments have been made for malaria [21, 22]. For the climate projection problem, a multi-climate model, multi-health model cascade has been examined revealing similar responses from two (similarly structured) dynamical health models, but large disagreements between the statistical and dynamical approaches [23, 24]. Uncertainty for malaria application models is identified as far exceeding that of other application sectors such as hydrological or agriculture [25]. The small size of the health and climate

model ensembles implies that the relative contributions of climate and health model (structural) uncertainty to overall uncertainty in health projections remains an open question. Identifying structural uncertainty is facilitated for models of a given structure, for example compartmental models, with a recent study showing limited impact of model structural uncertainty of 37 different compartmental models of ebola on intervention decision making [26].

Relative to structural uncertainty, the uncertainty associated with model parameter settings has received considerable attention [27–31]. Bayesian techniques have been employed for model parameter setting, which render an estimation of parameter uncertainty post priori from a prior estimation and a measure of the model fit to data [30, 31]. This reflects similar developments in other application fields such as hydrology, where Bayesian selection or alternative Kalman filtering approaches have sometimes been used to construct probability density functions of outcome and posteriori parameter uncertainty [32–36]. A series of predictive studies of a range of health outcomes used a Bayes approach to produce probabilistic predictions [2, 37, 38]. Regarding dynamical models of malaria, model parameter sensitivity experiments in low and high transmission settings determined mosquito biting rate to be key for R_0 in the OpenMalaria model [39–41]. The methodology examined the output Jacobians (the matrix of all first order function derivatives), a method that has also been employed for atmospheric models [42–44], but has the drawback of being dependent on the basic state employed. Parallel to the climate modeling approach, some studies employ ensembles of health models integrations with stochastic parameter perturbations [45–47]. These studies showed limited perturbations of a single model parameter could lead to a very large spread in the simulated transmission intensity in highland setting where temperature was close to the 18°C threshold.

These studies have demonstrated that, apart from model structure uncertainty, parameter uncertainty in health application models could potentially be very significant. Even simple R_0 models contain parameters concerning transmission probabilities and biting rates that are uncertain. Complex process-based distributed models that account for climate in the disease-vector life cycles contain an extended range of life-cycle and transmission-related parameters that are set using data from a very limited number of field or laboratory studies [48–50].

When considering the cascade of uncertainty, it is useful to understand where the greatest contribution to uncertainty originates. For example, if climate observations are used to drive a model of malaria transmission, do the instrumental and representativeness errors described above dominate the structural and parameter errors of the malaria model, and how do these relate to initial condition uncertainty? Answering this question is important as it allows the prioritization of error reduction.

A number of studies have attempted to answer this question using an additive approach with statistical models of malaria or dengue [51–53]. These fit a statistical model to case data using a number of non-climatic predictors, and then subsequently refit the model with a number of additional climate variables to determine if the additional climate information can improve the simulation of the health index in question, taking care to avoid overfitting. Since some of the non-climate variables can be proxies for climate, such as altitude, latitude, or the month of the year to remove seasonality, these studies thus necessarily focus on the impact of climate on interannual variability.

A variant of this approach employed a dynamical modeling framework for malaria, with a susceptible-exposed-infected-recovered (SEIR) compartmental modeling approach with host immunity [54]. First the model's parameters were calibrated using a sequential Monte Carlo methodology, and then the calibration was repeated allowing rainfall to impact one of the vector-related response parameters in an idealized way. The question posed was whether the incorporation of rainfall allowed an improved calibrated fit of the model to the data, which was indeed the case.

The above methodologies attempt to determine whether climate information can improve the fit of a disease model to data, but they do not account for uncertainties in the driving data or indeed the initial conditions. The aim of our study is thus to directly compare the three contributions of initial condition uncertainty, climate boundary condition uncertainty, and model parameter (but not structural) uncertainty. This is accomplished using a modification of the approach of Laneri et al. [54], whereby a dynamical model for malaria transmission is used to simulate observed malaria cases for a highland area, with the initial conditions, disease model parameters and the driving climate data itself calibrated in turn. Thus the goal is to determine whether calibration of model parameters or climate parameters within their respective uncertainties would lead to the best fit of the simulated cases to observations. Additional simulations will attempt to determine the relative uncertainty associated with the entomological initial conditions.

2 Methods

2.1 Model and data

The model experiments are conducted for a tea plantation in a highland location near Kericho in western Kenya which has a long record of monthly malaria case records, and is a region which has received much attention in the literature with regard to the debate concerning the impact of climate variability and trends on transmission [55–59]. The plantation malaria case data used is taken from a previous study [60] and spans the period January 1979 to October 2004. Examining the timeseries of data, a discrete jump in the data series is noted after approximately 1998 (Fig 2a). The cause of this is unknown, although we speculate that the implementation of drip irrigation in the area may have played a role. The small offset is removed by subtracting the minimum number of cases found after 1998 from the entire post 1998 period (Fig 2a). The years with anomalous transmission are highlighted by removing the linear trend from the cases and applying a high pass filter to remove variability on timescales longer than one year (Fig 2b). The one and two standard deviation lines emphasize the outbreak years and months which are marked for clarity, with outbreaks identified in 1981, 1985, 1988, 1990, 1994, 1997–99, and 2002–03. While the majority of the outbreaks occur at the end or shortly after the long rainy season in May, June or July, two outbreaks occurred in February and March (1998 and 2003, respectively), in both cases following an outbreak 6 months previously.

The dynamical model used in this study is referred to as VECTRI [50]. It is a compartmental SEIR model for the transmission in the human host coupled to a compartmental model for the parasite development in the vector in addition to the gonotrophic and larval cycles, similar to the Liverpool Malaria Model [48]. VECTRI in its basic form has been used to simulate and predict transmission intensities at seasonal timescales [5, 61] as well as investigating climate and land use impacts on malaria [23, 24, 62].

New additions to VECTRI are included for the experiments presented in this paper. From v1.4, there is a differential mean bite rate for hosts in the exposed, infected and recovered (EIR) individuals relative to the susceptible category (S), to produce over dispersive biting rates and reflect the fact that some individuals are more attractive to vectors [63–65], are more vulnerable due to clothing and housing standards [66], access to nets, location of housing with respect to water bodies [67–69], and that parasite infection also appears to increase attractiveness of individuals to vectors [70], although the latter effect is offset by increased net use in the case of clinical symptoms. The temperature to which the vector is exposed is calculated as the weighted average of the station measured outdoor temperature and an indoor hut temperature using the parameterization presented by Lunde et al. [71], with the weighting related to the vector endophilicity.

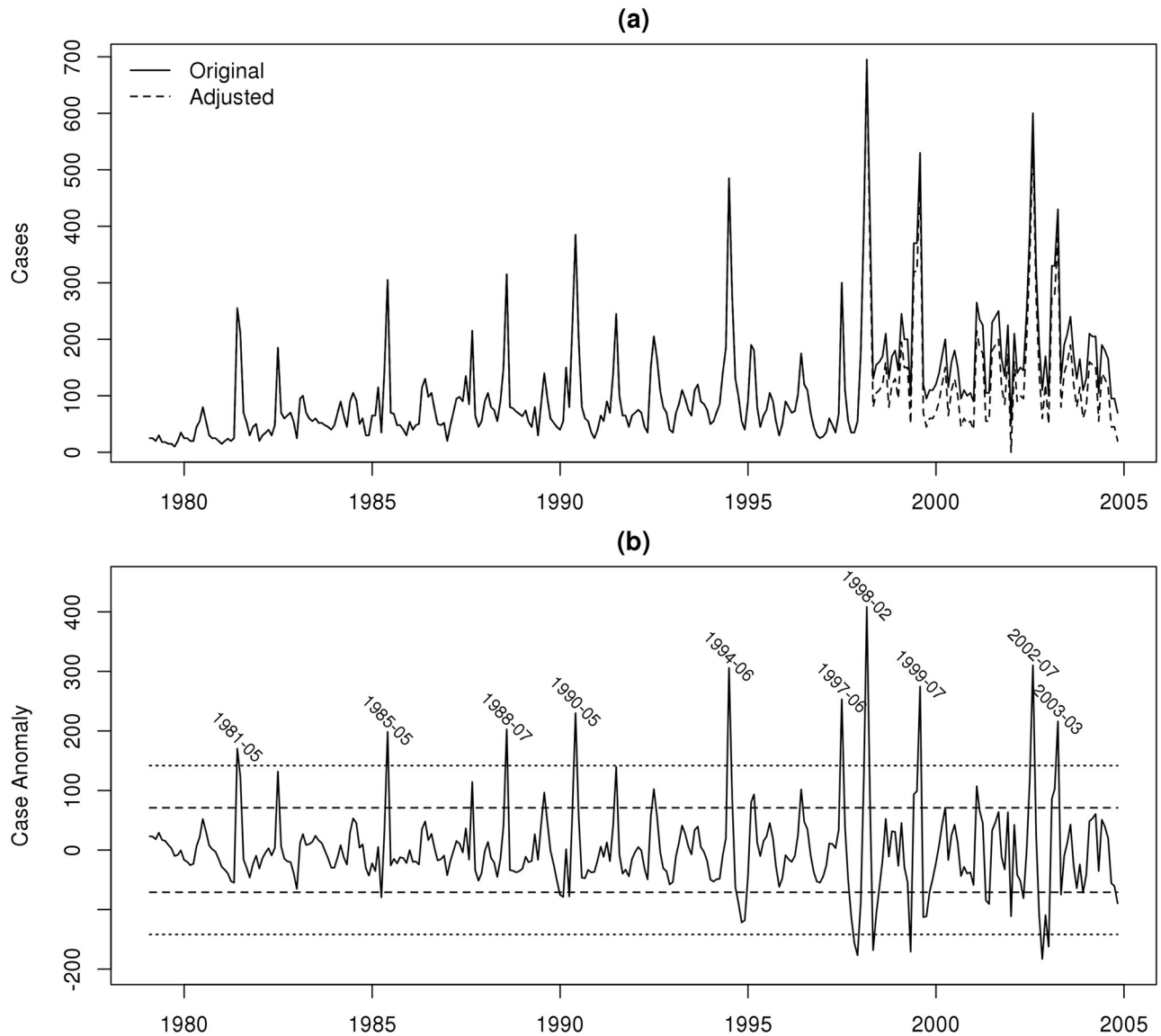


Fig 2. (a) Number of monthly reported malaria cases at plantation, with dashed line showing adjusted series. (b) Anomaly of cases calculated by removing linear trend from adjusted case series and then applying a one year, high pass filter. Dashed and dotted lines show 1 and 2 standard deviation events, respectively, with positive 2σ events labeled.

<https://doi.org/10.1371/journal.pone.0200638.g002>

Each time step of the model, a proportion of the infectious class return to a susceptible state as parasites are cleared. In addition, host immunity has been added to the model following the compartmental model of Laneri et al. [54]. Some infectious individuals are moved to a new recovered immune class at a rate that is governed by the EIR, such that 95% of individuals receiving 100 infectious bites per year ($EIR_a = 100$) will be clinically immune [72, 73]. Immunity is lost according to an exponential decay function at a rate whereby 95% of individuals lose immunity after 3 years. In addition to clinical immunity, the model also accounts for transmission blocking immunity [54, 74], by reducing the host to vector transmission probability in immune individuals [75], (see review in S5 and S7 Tables in Ermert et al. [76]). Another addition to the model is the introduction of a constant birth rate of population, set to 2% per annum here. An equal death rate is assumed, consistent with the assumption of a

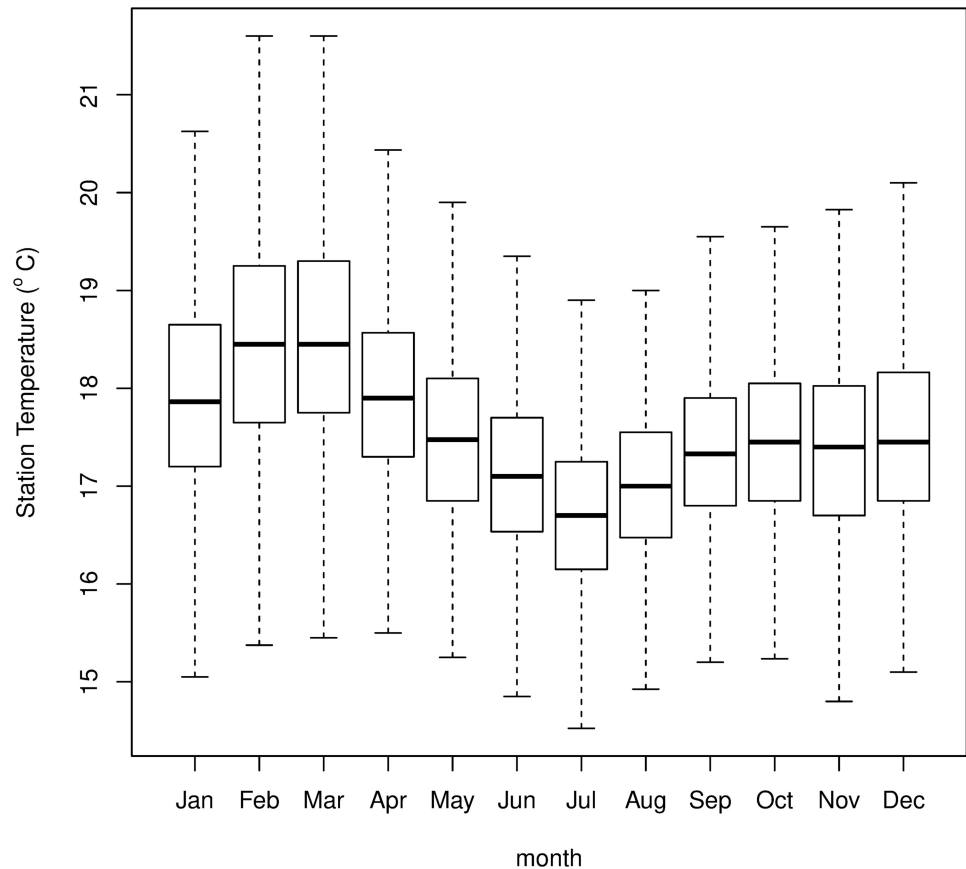


Fig 3. Box-whisker plot of the daily two meter height station temperatures. The horizontal lines give the mean for a given month over the whole period, the boxes representing the lower and upper quartile limits of the daily temperature, while the whiskers mark the minimum and maximum daily value observed over the whole dataset.

<https://doi.org/10.1371/journal.pone.0200638.g003>

unchanging population size. This is a reasonable assumption for a clinic that chiefly serves the workers and families for a fixed-sized plantation, in the absence of data concerning the clinic catchment population. Increasing population size tends to reduce transmission intensity in VECTRI as the model does not account for super-infection.

The model is driven by daily temperature and rainfall measured at the Kericho station data, originally obtained from the Kenyan Meteorological Department [77]. The VECTRI model assumes daily mean temperature to simply be the average of the maximum and minimum values. The long-term, annual-mean temperature of the processed stations data is 17.4°C for the period of the experiment, cooler than the approximate 18°C threshold assumed to be required for sustained *falciparum* transmission (Fig 3). For most months of the year, daily mean temperatures that exceed 18°C are at least an upper quartile occurrence, and these occur more frequently during the latter decade during to a warming trend that occurs at the station [77]. The default model is initialized from an arbitrary state of 5% prevalence of the parasite in the host population. The first year of the model integration is repeated four times in order to allow the model parameters to adjust from the artificial and arbitrary initial conditions to a state that is appropriate for the location. The first three repetitions are discarded in the analysis. The experiments presented in this paper will demonstrate that this period is adequate to prevent initial conditions from impacting the simulations. An additional ensemble of experiments are conducted with varying initial conditions in order to investigate initial conditions uncertainty.

2.2 Soft constraint genetic algorithm calibration

There are a wide variety of calibration methods available to perform both local and global parameter space calibration, including Bayesian methods, adjoint methods, ensemble Kalman filtering and sequential Monte Carlo genetic algorithms [78, 79]. Genetic algorithms are an effective, if non-optimal, global search strategy that have the attraction of not requiring a tangent linear or adjoint model of the full code, and can find global function minima of highly nonlinear systems. As the concept is relatively straightforward to understand and implement, the technique is utilized in a wide range of fields, including atmospheric models, engineering problems, machine learning and hydrology [80–84]. This simplicity is also an advantage if the intention is to use a model in decision support, such that decision makers can easily appreciate the underlying concepts of the tools they are using.

The method employed here is a standard genetic algorithm (GA) approach, with selection, cross over and mutation, illustrated schematically in Fig 4. An ensemble of m VECTRI models M_1, M_2, \dots, M_m are created, each of which has a set of n parameters K_1, K_2, \dots, K_n which are to be calibrated. These parameters can be *model* parameters that determine the way malaria transmission is modeled. The calibration process is also applied to the initial conditions [85] and, in a relatively novel development, also to *climate* uncertainty parameters. These introduce a scale factor or offset to the rainfall and temperature data to account for measurement and spatial sampling uncertainty. Each parameter is initialized using a random value sampled from a Gaussian distribution using the respective default value $K_{i,\mu}$ and the prior uncertainty estimate $K_{i,\sigma}$ thus $K_i \sim \mathcal{N}(K_{i,\mu}, K_{i,\sigma}^2)$. Parameters can be subject to upper and lower bounds to prevent unphysical parameter settings: $K_i \in [K_{i,min}, K_{i,max}]$.

The ensemble is integrated to produce m predictions which can be compared to the data source. A fitness function \mathcal{L} is constructed for each member which determines the probability of the member’s parameter settings being passed to model members of the subsequent “child” generation. The GA allows for cross-over, thus rather than selecting whole models, the child generation can have between 1 and a specified maximum of β parents, where $\beta \in \mathbb{Z} \cap [1, m]$. In this work, $\beta = m$ is adopted.

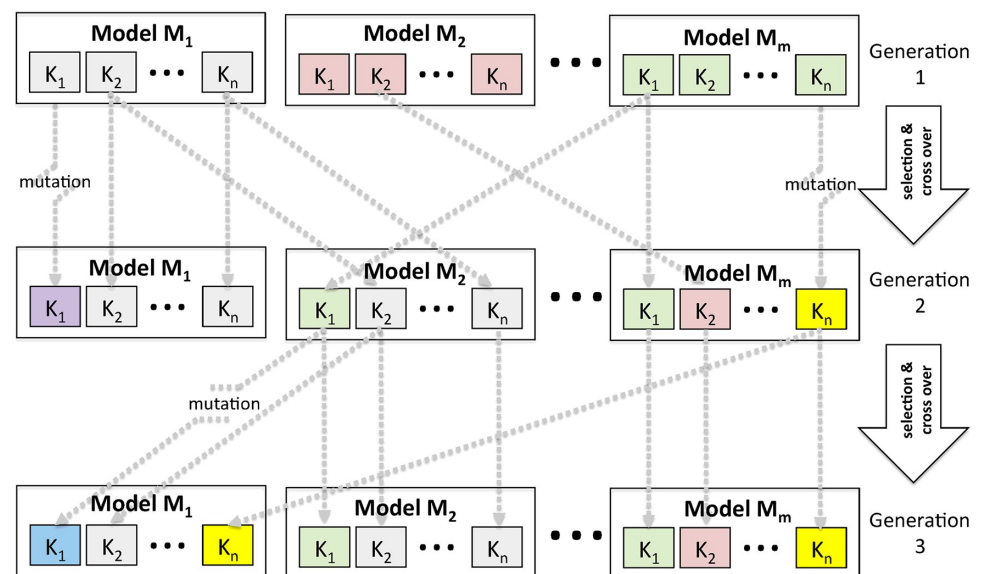


Fig 4. Schematic of genetic algorithm approach, see text for details.

<https://doi.org/10.1371/journal.pone.0200638.g004>

Parameters can undergo mutation to ensure that adequate genetic variety is retained in the ensemble and prevent the ensemble collapsing to a single member, usually referred to as ensemble degeneracy, and that the global parameter space is explored. The mutation is accomplished by adding a random perturbation to each parameter thus

$$K_i = K_i + \mathcal{H}(P - \lambda_U)\gamma\lambda_i, \tag{1}$$

where $\lambda_U \sim \mathcal{U}[0, 1]$ is a random number sampled from a uniform distribution, \mathcal{H} is the Heaviside function, $\lambda_i \sim \mathcal{N}(0, K_{i,\sigma}^2)$ is a random variable sampled from a normal distribution of zero mean and standard deviation equal to the prior uncertainty. P is the probability of mutation and γ is the mutation scale factor, and as is usual in GA, both the mutation probability and magnitude decline over generations to a lower rate, with a relaxation timescale of τ (units are number of generations). This allows the algorithm to find the global minimum of the model fitness in the early stages of the search. Thus for generation G , if $\alpha = e^{-\frac{G}{\tau}}$, then

$$\gamma = \alpha\gamma_0 + (1 - \alpha)\gamma_\infty, \tag{2}$$

$$P = \alpha P_0 + (1 - \alpha)P_\infty. \tag{3}$$

From test simulations, we select $\gamma_0 = 1$, $\gamma_\infty = 0.1$, $P_0 = 0.5$ and $P_\infty = 0.02$, although the conclusions are not affected by variations in these parameters, as they were only found to impact the number of generations required to converge to the final solution, not the solution itself, within reasonable variation.

The final step is to define the fitness function \mathcal{L} . The choice of function is flexible and can represent any metric of model skill, such as accuracy of predicting monthly cases, interannual cases variability, or aspects that describe the malaria seasonality [86], depending on the model aspect that the user wishes to maximize with the calibration process. Indeed, a fitness function can be constructed using an appropriately weighted average of a basket of different measures. Here the r^2 correlation coefficient goodness-of-fit is applied comparing the model's simulated monthly cases per 1000 prediction C_s to the observed total case number C_o .

The fitness function for one model ensemble member is defined thus:

$$\mathcal{L} = \eta\mathcal{H}(\overline{C_s} - \epsilon)r^2(C_s, C_o)\prod_{i=1}^n P(K_i) \tag{4}$$

The term η is a normalization factor such that the sum of the ensemble fitness is unity, the second Heaviside function term on the right-hand side is to penalize any model with settings that prevent malaria cases exceeding a small threshold value ϵ . Thus model settings that results in zero or near zero transmission are prevented from passing their parameter settings to the next generation, increasing the efficiency of the algorithm. The third term represents the r^2 correlation goodness of fit between C_s and C_o . The fitness function has one further contribution which represents a penalty function depending on the departures of the model parameters from their default values, $\prod_{i=1}^n P(K_i)$. Without this, calibration techniques would search for the optimal model parameters that maximize the skill of the model.

If a good fit to data can only be obtained using extreme model parameter values that lie far outside the accepted range, this would indicate that the model suffers from structural errors, such as neglecting or poorly representing a key processes. The GA calibration technique imposed here therefore implements a penalty function for departures in parameter settings from the default value, accounting for the prior uncertainty, in much the same way that variational assimilation systems instigate a penalty function for departures from the first-guess

model integration [87]. Thus the calibration search is effectively performed within an n -sphere of a dimension determined by the parameter priors, centered on the vector describing the parameter prior best-estimate values. This has the added advantage of reducing the search space, making the algorithm tractable for a high dimensional search, but with the obvious drawback that optimal solutions may well be missed if the assessment of the default parameter and/or its uncertainty is inaccurate.

The departure penalty is based on each respective parameter setting's probability. In this initial investigation the uncertainty of all model parameters is assumed to be Gaussian. It is possible that further improvements in the approach could be achieved by allowing some parameters to asymmetrical error distributions (possibly by assuming Gaussian error characteristics of transformed variables [88]). If a parameter is poorly constrained, it may be appropriate to allow a range of values to be adopted without penalty, although setting large Gaussian errors will have a similar effect. Thus the 2-tailed probability of parameter value K_i is:

$$P(K_i) = 1 - 2 \left| \int_{-\infty}^{K_i} \mathcal{N}(K_{i,\mu}, K_{i,\sigma}^2) dK_i - \frac{1}{2} \right|. \quad (5)$$

The product of the probabilities is used in the cost function to produce an overall probability of the parameter vector. The fitness vector will be insensitive to the number of parameters calibrated since the vector of fitness values for each ensemble member ($\mathcal{L}_1, \mathcal{L}_2 \dots \mathcal{L}_m$) is normalized by $\eta = (\sum_{j=1}^m \mathcal{L}_j)^{-1}$, so that the normalized vector of fitness sums to unity and each fitness \mathcal{L} represents the probability of the model member of passing on its parameter settings to the child generation. Using the fitness vector, child generation models then generate β uniform random numbers $\mathcal{U}[0, 1]$ to select their β parents for each parameter.

2.3 Uncertainty assessment

The temperature and rainfall data have various sources of uncertainty, such as instrumental random error and biases, trend errors and sampling error [89]. Temperature error is accounted for by allowing a constant offset and a linear trend to the driving data. Station instrumentation random error and bias for temperature are quite limited, on the order of 0.2°C [90]. These errors are likely dominated by representative errors, that is, the temperature data measured at the station may not be representative of the temperature everywhere within the plantation catchment area. These will spatially vary due to changes in altitude across the steep terrain, heterogeneous land cover creating microclimates [91–93] and vectors' exposure to temperature [94].

Considering the spatial uncertainty, altitude variations over complex topography can easily result in temperature variations on the order of several degrees, and land cover variations within a clinic catchment can also contribute to temperature heterogeneity [92, 95–97]. A very approximate estimate is made of such effects using the root mean squared differences between a gridded analysis of climate, the ERA interim reanalysis [98], and the daily station temperatures, which are 1.4°C. There are many other sources of uncertainty regarding the climate. For example, endophilic vectors rest indoors and are subject to different temperatures, but the impact will depend on the location and type of building, and the degree of endophilicity of the key malaria vectors in Africa is also highly uncertain [59, 99]. This is also an example where the boundary between climate and model uncertainty is blurred, since the VECTRI model accounts for hut temperatures [71]. Thus part of this climate uncertainty is reflected in model uncertainty in the setting of the parameters related to this scheme.

Taking the above uncertainties into account, the tolerance for the temperature offset is set 2.5°C. It is acknowledged that while this may be a reasonable estimate for spatial

representativeness uncertainty for a typical health clinic catchment, it may be an overestimation in the context of a plantation where workers and their families are often housed in a single on-site residential location. As the simulations span more than two decades, it is possible that changes in urbanisation and other land cover changes may also impact the assessment of temperature trend over the experiment period. A tolerance is thus also allowed on the temperature trend of 0.02°C per annum, which is of the same order as the observed trend.

Rainfall measurement errors are difficult to judge, despite a large literature inter-comparing various satellite-based retrievals and station data in Africa [100–109], and rainfall is also highly spatially heterogeneous in the tropics [110, 111] implying sampling error can be large. The uncertainty is thus set to 30% of the mean.

The prior uncertainty is also required for model parameters. Although many life cycle-related parameters are known from laboratory and/or field studies, often only a single study is available and ad-hoc estimates of the prior uncertainty are necessary. Here the priors are set to 20% of the mean for most life-cycle parameters (Table 1) following Lindstrom and coauthors [31]. Parameters pertaining to the model’s treatment of hydrological processes [112, 113] are poorly constrained and are subject to a higher level of uncertainty to allow the calibration algorithm maximum scope to find the global optimum.

The initial conditions for the malaria model consist of the entomological parameters, such as vector and larvae densities as well as the circumsporozoite protein rate (CSPR), and epidemiological parameters, such as the parasite ratio (PR) and immunity levels in the host population. It was found that the calibration procedure did not converge when calibrating the malaria model initial conditions. To demonstrate why this was the case, an additional experiment was conducted, in which an ensemble of 100 VECTRI integrations are run, each initialized with a PR value ranging from 0.01 to 1.0, with an increment of 0.01. No model spin up period was applied in these experiments.

3 Results

3.1 Full calibration

The relative importance of the different sources of uncertainty (model, initial conditions and driving climate) are assessed by assessing the quality of fit of the converged constrained genetic algorithm ensemble. The convergence of the algorithm is first examined for the case where all climate, model parameters and initial conditions are calibrated, in order to confirm that the

Table 1. Default model parameters (K_{μ}), the uncertainty estimate (K_{σ}), and the ensemble mean of the calibrated values (\bar{K}) for the experiment in which only the model parameters are calibrated (“model”) and the experiment where model, climate and initial conditions are calibrated (termed “all”). Only those parameters that differ from their default value in the final calibrated state are given.

Parameter K	K_{μ}	K_{σ}	\bar{K} (model)	\bar{K} (all)
Infiltration and evaporation loss (mm day ⁻¹)	250	150	118	239
Clearance rate for non-immune subjects (days)	20.	20.	27.9	20.3
Degree day length of sporogonic cycle	111.	20.	95.9	110
Success rate for vectors to find blood meal (day ⁻¹)	0.5	0.3	0.61	0.51
Base [‡] larval daily survival rate	0.92	0.1	0.999	0.985
Base [‡] vector daily survival rate	0.96	0.2	1.0	1.0

[‡] The base survival rates for the vector and larvae are both subsequently modified by temperature-dependent mortality schemes [50].

<https://doi.org/10.1371/journal.pone.0200638.t001>

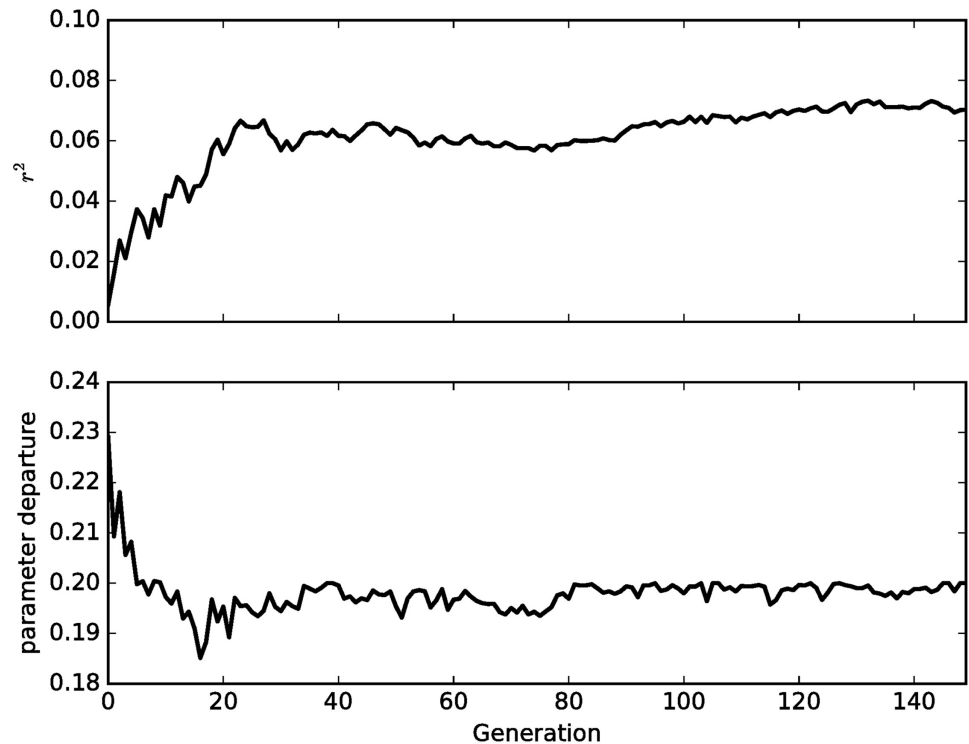


Fig 5. Adjustment of r^2 correlation and a measure of the mean normalized parameter departure

$$\frac{1}{mm} \sum_{i=1}^m \sum_{j=1}^n \frac{|K_j(t) - K_{j,0}|}{K_{j,0}}$$

<https://doi.org/10.1371/journal.pone.0200638.g005>

model is able to capture at least some of the observed interannual variability and malaria outbreaks (Fig 5). The ensemble mean skill is maximized after approximately 20 generations, and while the mean parameter departure is stable after approximately the same time. The algorithm makes further adjustments to the parameter settings over subsequent generations which do not increase the departure penalty but improve the ensemble fitness, which reaches a maximum after 80 generations.

The simulation of cases per 1000 population of the VECTRI model ensemble is compared to the observed cases in Fig 6, where the shading indicates the ensemble spread, with bands for the middle tercile, the 20–80th percentiles (excluding the 16 lowest and 16 highest models) and the 5–95th percentiles. Although the model simulation is poor in the initial period from 1979 to 1983, the simulation is able to reproduce the seasonality of cases numbers for a number of years, in particular the seasons of 1990, 1991 and 1992. The model is also able to reproduce the enhanced outbreak of 1998, which was associated with rainfall anomalies and enhanced temperature associated with the strong El Niño event that year [91], but the timing of the event is inaccurate. In both the 1987/88 and 1998/1999 strong El Niño years, the model tends to maintain a higher level of cases over a period of one to two years that is not observed. The overall ensemble-mean correlation approaches 0.25 which is statistically significant given the length of the timeseries.

The nature of the observed and modeled variability can be assessed by conducting a wavelet analysis [114], which is accomplished using the WaveletComp package of R [115]. The wavelet power spectra of the normalized and detrended observed and simulated timeseries of malaria clinical cases (Fig 7a and 7b) shows some similarities between the two. Both the observations and model unsurprisingly have a statistically significant peak at the 12 month annual cycle. In

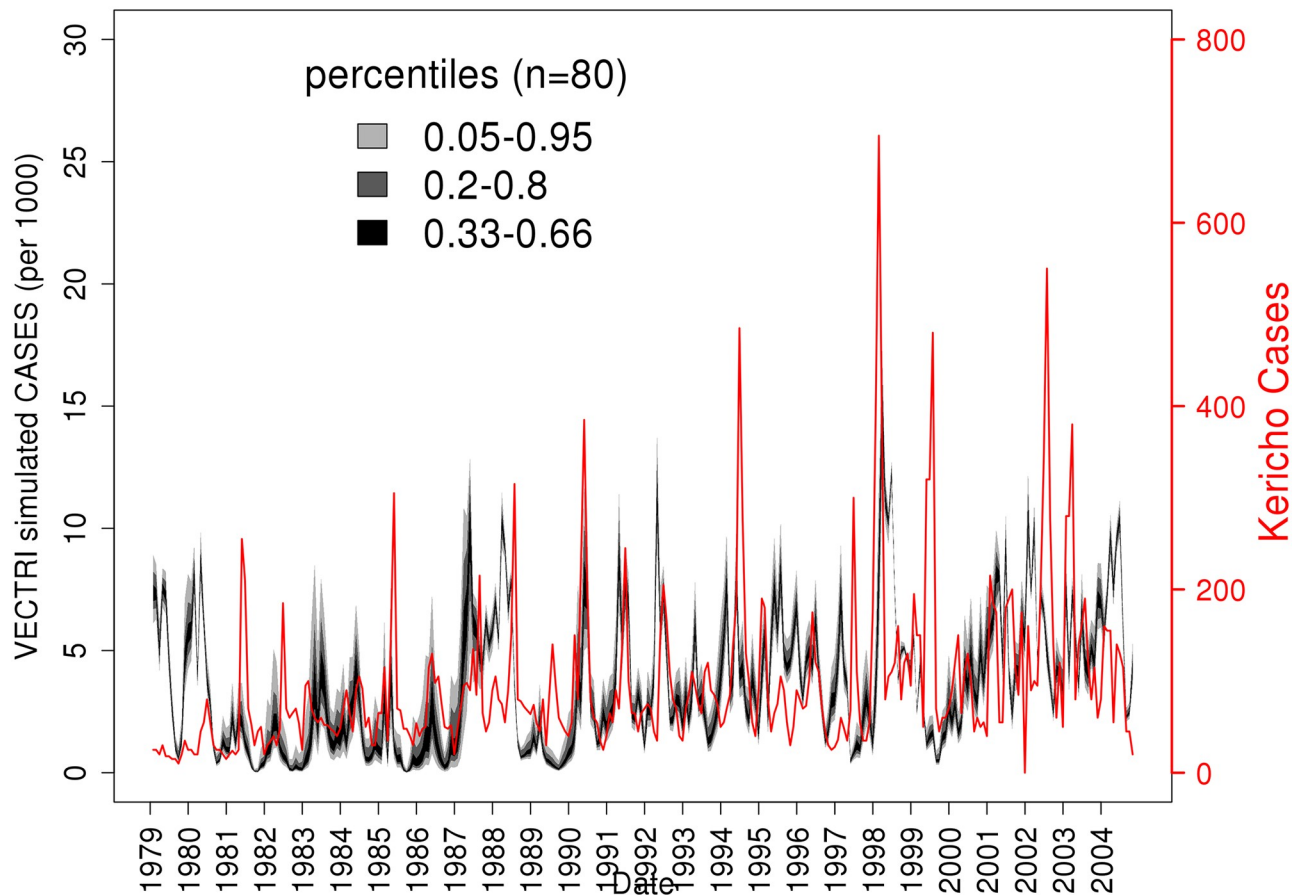


Fig 6. Timeseries of malaria cases for the Kericho plantation (red line, right axis) and the simulations of malaria cases per 1000 population from the calibrated 80 member ensemble of VECTRI integrations. Both model parameter and climate information is calibrated.

<https://doi.org/10.1371/journal.pone.0200638.g006>

in addition both have a clear mode at approximately 4 to 5 years, related to the impact of ENSO, although this is relatively stronger in the model than the observations. One interesting disparity between the two is that the observations sometimes show strong transmission anomalies after only one of the two rainy seasons experienced in this location, which results in a stronger mode at a period of around 6 months and also 18 months. The model instead often maintains transmission between consecutive rainy seasons, which results in a much weaker 6 month mode which is not statistically significant at the 10% level. This is associated with the fact that the model did not reproduce the two pairs of outbreaks that were separated by 6 months in 1997/98 and 2002/03. This could also indicate a model structural error such as an overly long memory associated with parasite clearance. The lack of stochastic forcing term in the model is also likely to be a factor, leading to transmission being overly regular. The lower panels examine the coherency between the model and observations, and highlight the agreement at the annual cycle (Fig 7c and 7d) period. Coherency is high across timescales in the early 1990s when the two timeseries match very closely. The periods of high coherency at ENSO timescales match the periods when high wavelet power is observed in the observational record.

3.2 Impact of climate versus model parameter uncertainty

Two further experiments were conducted that calibrated the model parameters and the climate parameters separately. When the model parameters are calibrated, despite the fact that over 17

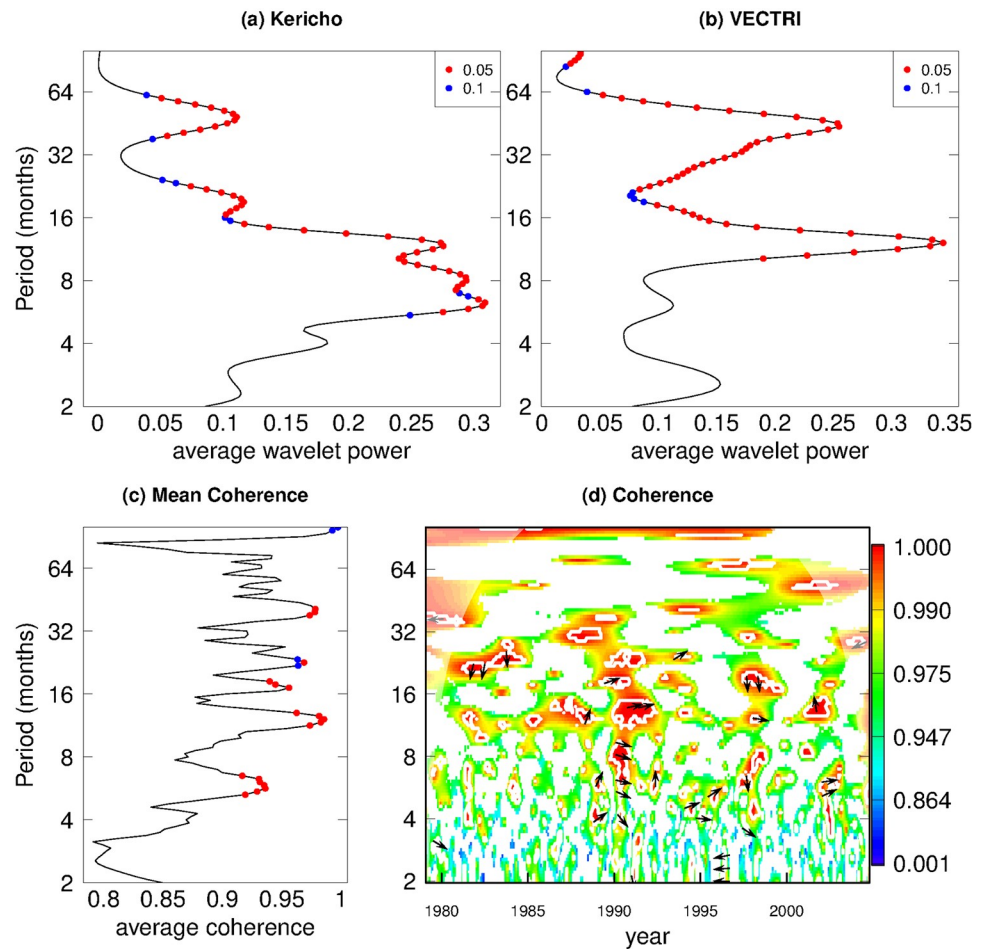


Fig 7. Wavelet spectral power of (a) Kericho cases data and (b) Fully calibrated VECTRI model cases simulation and (c) mean and (d) time-evolution of wavelet coherence between the two time series. In panel (d), areas with $p > 0.33$ are masked for clarity.

<https://doi.org/10.1371/journal.pone.0200638.g007>

parameters are able to adjust, the calibrated model is unable to sustain continuous transmission through the earlier part of the series (Fig 8a). Outbreaks begin to occur in the late 1980s due to the presence of a weak warming trend in the station data and it is only after the warming of 1998 that sustained transmission can begin. In Fig 8b the result of the temperature and rainfall calibration is revealed to be very similar to the results with all parameters calibrated. Thus, simply allowing the calibration process to apply a constant offset to the air and pond water temperature, and to scale the temperature trend and precipitation amount, leads to a vastly improved simulation of the malaria case variability, even using the default, uncalibrated malaria transmission model. In fact, a further pair of experiments which separated the rainfall and temperature calibration show that the temperature calibration is by far the most important (not shown); and only allowing the rainfall to calibrate via a scaling parameter, the model is unable to simulate malaria transmission at all due to the cold temperatures at this station location.

Table 1 (right columns) gives the mean of the calibrated model parameters for the model-only and all-parameters experiments, for the key parameters which were perturbed more than 10% of the prior uncertainty when calibrated in at least one experiment. Calibrated parameters

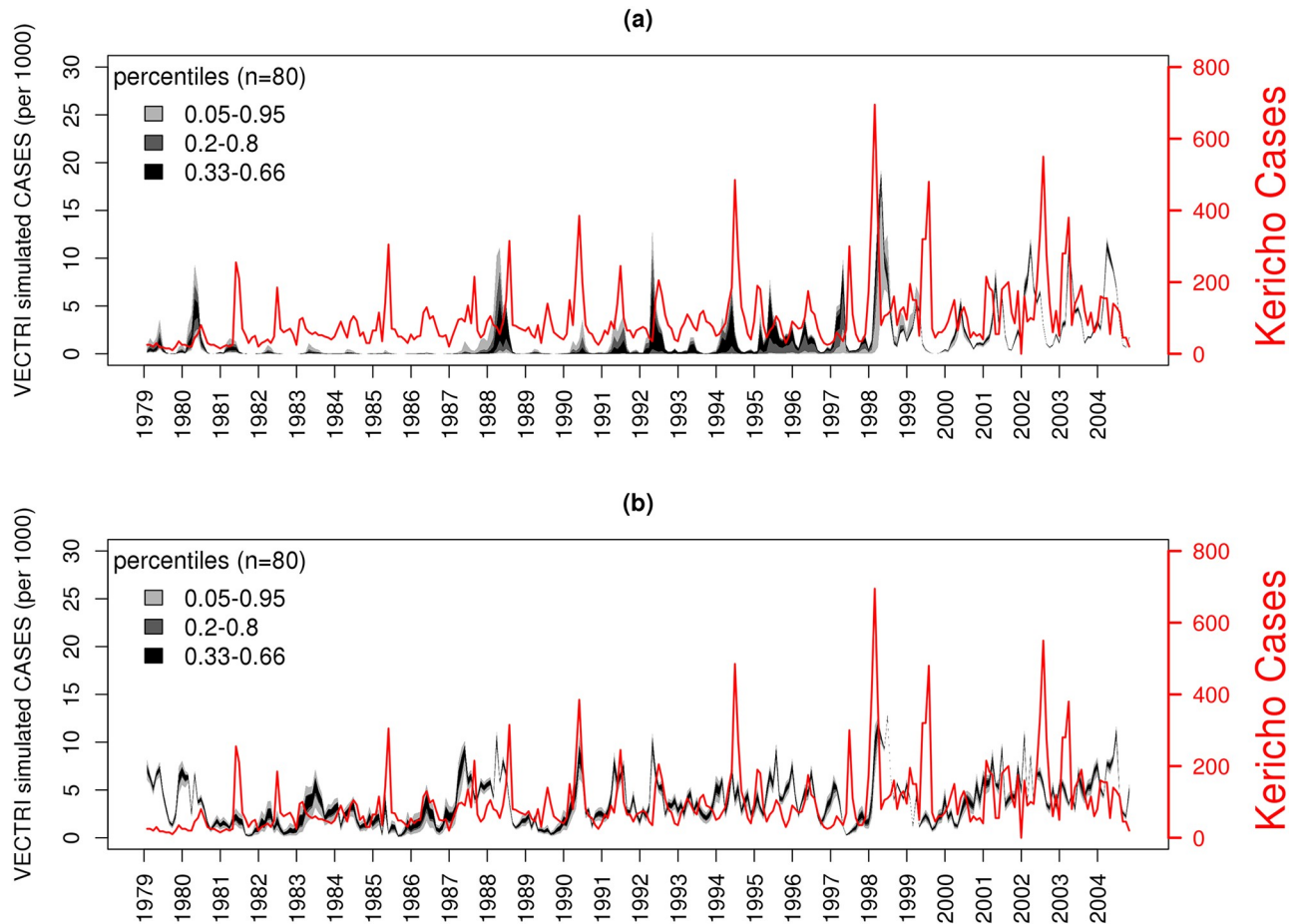


Fig 8. Timeseries of malaria cases for the Kericho plantation (red line, right axis) and the simulations of malaria cases per 1000 population from the calibrated 80 member ensemble of VECTRI integrations, where (a) only malaria model parameters are calibrated and (b) only climate information is calibrated.

<https://doi.org/10.1371/journal.pone.0200638.g008>

that are perturbed significantly from their default values are all perturbed in the direction of increased transmission intensity. This is expected, as it is recalled that the cold mean temperatures imply no transmission in the default model. For example, in the model-only calibration experiment, parasite clearance rates increase from 20 to 28 days, the sporogonic cycle decreases from 111 to 96 days and the underlying survival rates for larvae and vector are both increased to unity (which are subsequently reduced by temperature-related functions [50]); all changes that increase transmission intensity and the number of clinical cases modeled. There is also a strong reduction in the infiltration and evaporation rate of ponds, which would increase their duration after rains and increase vector productivity. Without the soft constraint, the calibration process may have produced a closer reproduction of the observed case data, but at the expense of unreasonable settings for these parameters.

Many of the parameters are calibrated to values close to their default, indicating that these variables have a limited influence on the simulation quality as measured by the r^2 correlation. Perturbing these parameters from their default values does little to improve the fit of the simulation to the observed cases but increases the penalty function for the parameter departure. For example, the parameter that specifies the number of eggs laid per batch by each gravid female is calibrated to a mean of 78 eggs per batch, very close to the default value of 80, with a

Table 2. Default model parameters (K_μ), the uncertainty estimate (K_σ), and the ensemble mean of the calibrated values (\bar{K}) for the experiment in which only the climate parameters are calibrated (“climate”) and the experiment where model, climate and initial conditions are calibrated (termed “all”).

Parameter K	K_μ	K_σ	\bar{K} (climate)	\bar{K} (all)
Temperature offset (°C)	0.0	2.5	2.3	1.1
Temperature trend (°C decade ⁻¹)	0.0	0.02	0.022	0.017
Water temperature offset (°C)	0.0	2.0	0.25	0.23
Rainfall scale factor	1.0	0.3	0.98	1.0

<https://doi.org/10.1371/journal.pone.0200638.t002>

perturbation far smaller than the prior. The parameters governing the behavior of the immunity scheme are also insensitive to the calibration process.

For the climate calibration, the temperature offset is the key parameter, and it is calibrated with a value approaching its specified uncertainty of 2.3°C (Table 2). The method also slightly enhances the temperature trend by a further 0.02°C decade⁻¹ and a small water temperature offset of 0.25°C is applied, with substantial spread across the ensemble. The rainfall is left largely unchanged. The temperature offset is much smaller in magnitude when both model and climate parameters are calibrated, with the mean temperature calibrated warmer by 1.1°C in this case due to the additional contribution of the model calibration.

3.3 Impact of initial conditions

The calibration procedure did not converge when calibrating the initial conditions. The reason for this is clear when the evolution of the PR is examined for the 100-member VECTRI ensemble of integrations, which each was initialized with a different PR value ranging from 0.01 to 1.0. The solutions converge over a period of approximately two years (Fig 9), indicating that the system is asymptotically stable. This result is expected from the extensive literature on regional climate and numerical weather prediction models, driven by lateral boundary conditions for meteorological variables such as winds, temperature and humidity. Initial condition uncertainty has been shown to be highly constrained in a regional model forced by lateral boundary conditions [116], and initial conditions uncertainty reduces over time in both low [117] and high [118] resolution experiments. These lateral boundary conditions are analogous to the time-evolving climate boundary conditions used for the disease model. The simulations here suggest that initial conditions are important in the seasonal prediction early warning context, and still lead to significant spread between the models in the second year, but can be safely neglected as a source of uncertainty in multi-year simulations or climate change experiments. The rate of convergence is dependent on the temperature, which is demonstrated in Fig 9 by two further ensembles in which an arbitrary additional temperature offset is used to give a mean temperature of 19.9°C and 24.4°C. At 24.4°C the ensemble spread collapses almost to zero after just 4 months, as this temperature is close to the optimum temperature for transmission [6] and PR values are close to their potential maximum. Thus initial condition uncertainty is relatively more important in cooler highland locations than warmer lowland endemic areas.

4 Discussion

In this work, the various contributions of model and climate uncertainty to overall uncertainty in modeled malaria-related health outcomes was assessed. The approach was to use a genetic algorithm to calibrate the driving climate data and the malaria model parameters, each constrained by an estimate of parameter uncertainty.

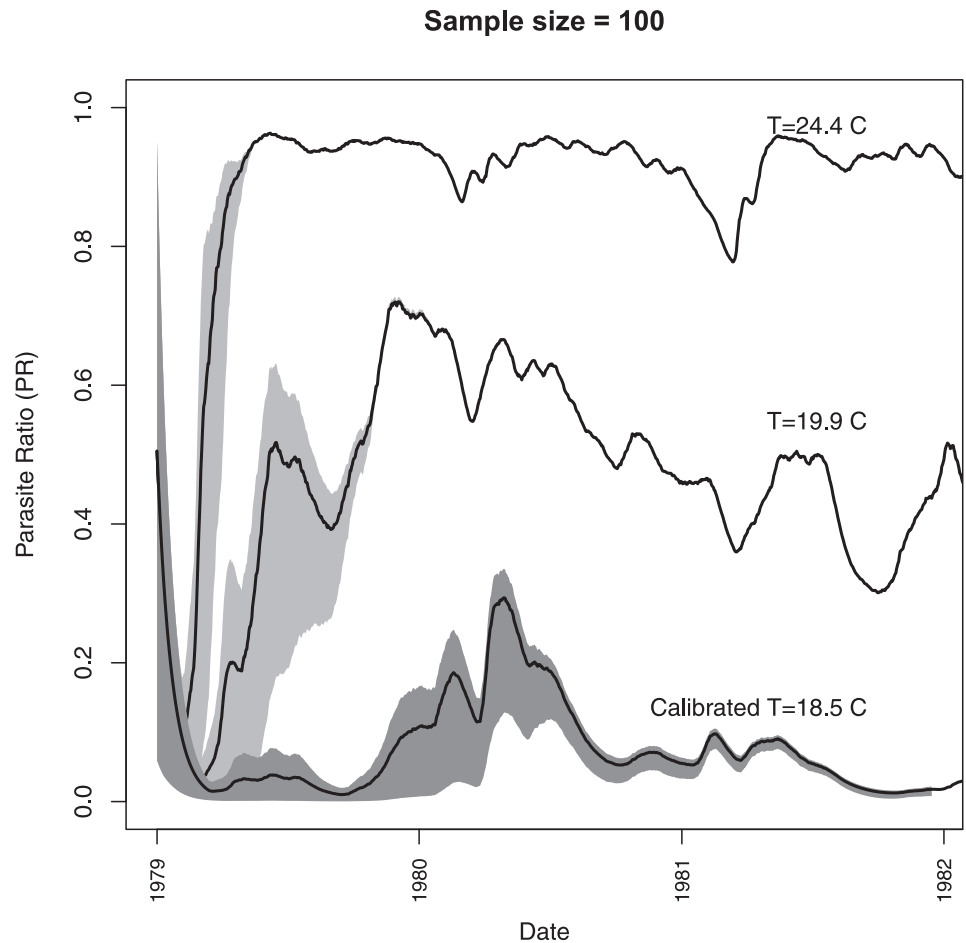


Fig 9. Evolution of PRd for three experiments, each with an ensemble size of 100 integrations that differ only in their initial conditions of PRd, equi-spaced between 0.01 and 1.0. The gray shading marks the 10 to 90 percentile of the ensemble, and the thick line indicates the ensemble mean. The first experiment uses the calibrated model and climate parameters and has a mean temperature of 18.5 °C, while the other two add an additional (arbitrary) temperature offset to give a mean of 19.9 and 24.4 °C, respectively.

<https://doi.org/10.1371/journal.pone.0200638.g009>

The first task was to determine if the model could provide a reasonable simulation of malaria. In its default setting the model did not simulate malaria due to the cold mean temperatures recorded at the plantation station. However, once the model parameters were calibrated it provided a reasonable simulation of the malaria case data and its seasonality and the outbreaks of 1986, 1990, 1991, 1992, and the exceptional year of 1998 are well captured. The positive correlation between the model simulations and observations was statistically significant. There are three notable outbreaks in observations that the model is unable to capture, specifically in 1999, 2002 and 2003. Shanks et al. [57] discuss the 2002 outbreak at some length, in particular the fact that no outbreak was observed at a nearby plantation (Fig 4 of reference [57]). They noted that both plantations switched to malaria treatment for outpatients from chloroquine to sulfadoxine-pyrimethamine during the period between 1999 and 2000 and were consequentially unable to explain the differences between the plantations. The year 2002 was associated with elevated cases at the district hospital, reflecting the plantation data presented here [119]. Hay et al. [119] showed that precipitation was normal in Kisii and Gucha districts in 2002, while Nandi and Kericho appeared to have anomalous wet conditions in May

2002. The rainfall anomaly in May 2002 was 3.22 mm day^{-1} , which while not the largest in the dataset, still represents a 90th percentile event, as the standard deviation of the monthly anomaly is 2.35 mm day^{-1} . The temperatures anomalies are limited. Thus there is the indication that, while the VECTRI model reproduces the seasonal cycle of malaria cases in response to the rainy season well, interannual variations in rainfall do not drive as much malaria variability as expected, possibly highlighting a model structural error that can not be easily addressed with parameter calibration alone. It is also emphasized that the calibration process only allows perturbations that are independent of time (with the exception of the linear temperature trend) and thus the fit also indicates that the structure of the underlying model is reasonable.

Some parameters did not change radically from their default values with the calibration process. This does not imply that these parameters do not affect the model simulations, only that, relative to their respective uncertainties, these parameters have a smaller influence on the simulation outcome for the given metric chosen to assess simulation quality. The calibration method will tend to perturb those parameters lead to the greatest improvement of simulation skill for the smallest departure penalty. This is subtly different to the study of model Jacobians, which identifies the parameters to which a model output is sensitive to, but not necessarily leading to improved skill. These experiments illustrate the power of departure constraint on the GA calibration, since if the priors were an accurate reflection of our present state of knowledge concerning a particular process then the method shows which processes should be prioritized for research to reduce these uncertainties.

The next question posed was, if we are able to calibrate the driving climate dataset within a reasonable range of the assessed uncertainty, is it possible to achieve a better fit to the observed series of malaria cases than if the malaria model parameters are fitted within their own respective uncertainties? In both experiments, the parameters were calibrated in such a way as to increase transmission intensities. Vector and larval survival rates were increased, for example, and the sporogonic cycle reduced. However, the calibration of malaria model parameters was not able to match the data as effectively as calibrating the climate data, in particular the temperature.

While instrumental random errors and biases are quite limited, the representativeness error can be quite substantial, particularly over complex topography. Thus, vectors in two locations that may be quite close to each other (in the same health centre catchment on even the plantation) may experience quite different temperatures due to differences in altitude, land cover and vegetation. A difference of 1 or 2 degrees can imply a very significant difference in terms of transmission, particularly near the 16 to 18°C threshold for sustaining transmission or in the range of temperatures where transmission intensity changes rapidly as a function of temperature [120], even if the transmission-temperature relationship is still hotly debated [6, 121]. As an example, if the sporogonic cycle for *falciparum* is assumed to obey a degree day relationship with a threshold of 18°C and a length 111 degree days [120], then a temperature range of 18.5 to 19.5°C across a health clinic catchment, representing an uncertainty of just 1°C, would be equivalent to an uncertainty in sporogonic cycle length ranging from 37 to 111 days. A comparison of a 1°C temperature uncertainty to a 20% uncertainty in the sporogonic cycle is shown in Fig 10. The simple calculation shows that while climate uncertainty would dominate at cooler temperatures, the model uncertainty of setting the degree day length for the sporogonic process is larger once the temperature exceeds around 23°C. This emphasizes the fact that the relative contributions of model uncertainty and climate uncertainty are a function of the locality and its mean climate. Climate is the most important in the experiments presented here due to the highland location chosen for the study.

As interannual temperature variability is limited in the tropics relative to higher latitudes, there is sometimes the tendency to neglect the impact of its temporal and small scale spatial

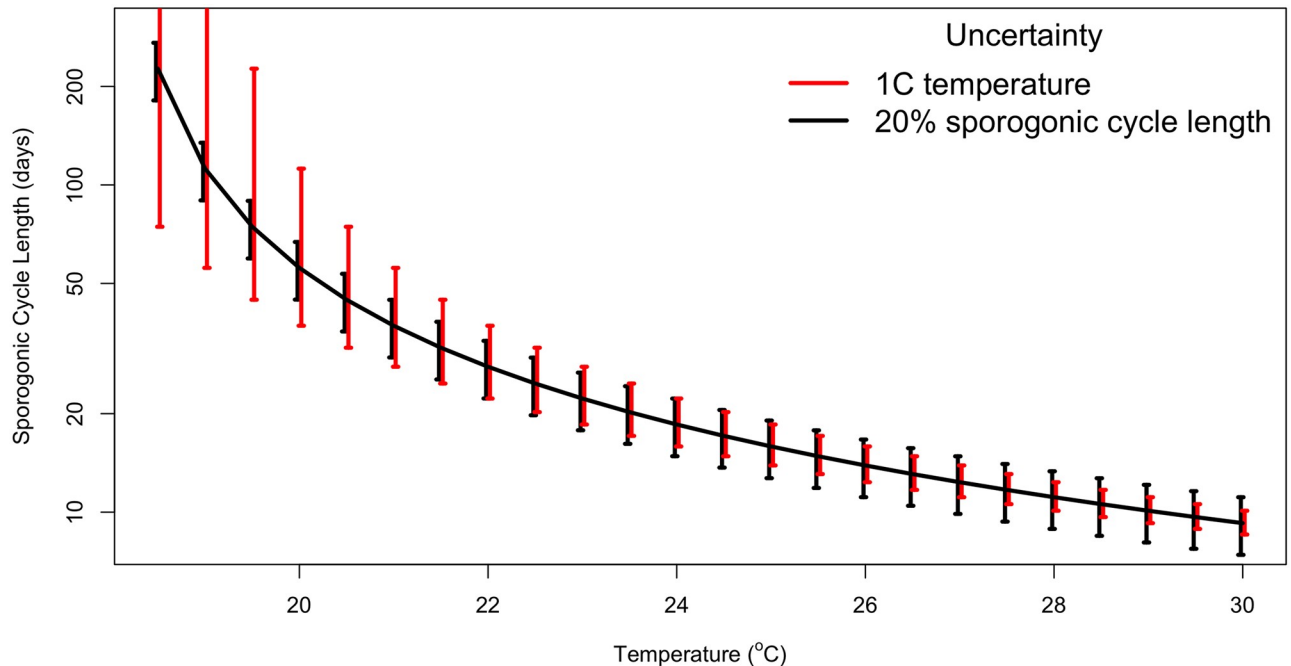


Fig 10. Comparison of range of uncertainty in the sporogonic cycle length (days) that would arise from assuming a 1°C uncertainty in temperature (red error bars) or a 20% uncertainty in the degree day length (black), as a function of mean temperature. The threshold temperature is assumed to be 18°C for this process.

<https://doi.org/10.1371/journal.pone.0200638.g010>

variability on transmission fluctuations. For example, some earlier work on predicting outbreaks, particularly associated with El Niño, used a univariate analysis focused on rainfall variability alone [91, 122, 123], neglecting the fact that tropical temperatures are usually $O(1)^\circ\text{C}$ warmer in El Niño years, which would also contribute to transmission intensification in highland areas. Likewise, while there is evidence from detailed mapping exercises that proximity to water bodies greatly amplifies risk of infection [67, 68], such analyses often neglect the impact of spatial variations in temperature that are associated with changing altitude, proximity to water and the associated changes in land cover. For example, one analysis dismissed temperature as a contributor to contrasts in malaria transmission between two highland locations since a statistical test using daily data deemed there to be no significant difference between annual mean temperatures of 26.1°C and 28.3°C [124], when in fact these differences could have a significant impact in terms of malaria transmission.

Statistical modeling approaches can attempt to separate temperature and rainfall contributions over small scales [125] but even these resort to interpolated gridded products which will neglect small scale variations in temperature and possibly underestimate its impact. Here it is seen that simply calibrating temperatures with a tolerance of 2.5°C and a tendency of 0.2C decade^{-1} is adequate to progress from a model that simulates no transmission at all, to one which is able to reproduce many of the seasonal changes in malaria in a highlands location.

Given the importance of temperature uncertainty, what could be the possible approaches to reduce this? One of the key uncertainties was related to the spatial representativeness of isolated temperature data, which could be addressed by running VECTRI at finer spatial resolutions, if the source of the spatial heterogeneity could be modeled and verified. VECTRI already uses topography to statistically downscale temperature to a finer grid mesh, usually on the order of 10km [50]. This could be extended to finer resolutions, although spatial information on precipitation is not yet available at scales finer than 10km over Africa from the famine early

warning system (FEWS) and more recently, the global precipitation measurement (GPM) products. However, topographical effects are not the only drivers of fine-scale temperature heterogeneity. Other factors such as land cover and canopy effects on temperature introduce further heterogeneity, and one approach to incorporate this would be to use the tile approach of land surface schemes, whereby VECTRI would be integrated for each tile portion of a grid-cell for each land cover type. At still smaller scales, factors such as the indoor temperatures of habitations are important, and harder to model with certainty. VECTRI can account for in-hut temperature impacts, but the data is not available to set the constants in these schemes accurately, and incorporating the spatial heterogeneity is more challenging. These uncertainties highlight the need of a modeling approach that can incorporate uncertainty in a calibration approach.

This preliminary study uses a fixed, ad hoc estimate of the model parameter uncertainty due to the fact that for many of the relevant biological processes, few analogous laboratory or field studies exist. Additional repetition of past laboratory studies would greatly help modeling efforts by allowing process uncertainty to be better estimated and incorporated.

5 Conclusions

Initial condition, model parameter and driving climate uncertainty were compared using a constrained genetic algorithm to calibrate a malaria model used to simulate cases in a highland location in south-west Kenya. As expected from earlier work with regional climate models, the experiments showed that initial condition uncertainty is only significant for the first two years of the simulation, and thus is relevant in the health early warning system context but not for simulations on decadal or longer timescales. The spatial uncertainty of temperature was found to be the dominant source of uncertainty, even though modification of parameters related to temperature-sensitive processes also improved the simulation quality. A simple analysis showed that this result is valid only for cold highland locations close to the lower threshold for sustained malaria transmission. At warmer temperatures, the model parameter uncertainty can become relatively more important. A number of sources of uncertainty have not been considered, in particular the model structural uncertainty. This could be achieved by employing a multi-model experimental setup with additional modeling systems such as the Liverpool Malaria Model [48]. Moreover the challenge of how to incorporate model uncertainty into the decision making process within the presented framework is left open for future consideration.

Acknowledgments

This project was initiated during two visits of the lead author to IRI which were financially supported by NIH. Bradfield Lyon and Israel Ukawuba are thanked for providing post-processed and quality controlled climate and health data and advice thereon. MCT Directs the IRI-WHO Collaborating Center (430) Malaria Early Warning Systems and Other Climate Sensitive Diseases and acknowledges the support of WHO in facilitating engagement with national policy makers and practitioners that generates questions for policy relevant research.

Author Contributions

Conceptualization: Adrian M. Tompkins, Madeleine C. Thomson.

Data curation: Adrian M. Tompkins, Madeleine C. Thomson.

Formal analysis: Adrian M. Tompkins.

Funding acquisition: Madeleine C. Thomson.

Investigation: Adrian M. Tompkins.

Methodology: Adrian M. Tompkins.

Software: Adrian M. Tompkins.

Visualization: Adrian M. Tompkins.

Writing – original draft: Adrian M. Tompkins.

Writing – review & editing: Madeleine C. Thomson.

References

1. Palmer TN. Predicting uncertainty in forecasts of weather and climate. *Rep Progress Phys.* 2000; 63:71–116. <https://doi.org/10.1088/0034-4885/63/2/201>
2. Lowe R, Coelho CA, Barcellos C, Carvalho MS, Catao RDC, Coelho GE, et al. Evaluating probabilistic dengue risk forecasts from a prototype early warning system for Brazil. *elife.* 2016; 5:e11285. <https://doi.org/10.7554/eLife.11285> PMID: 26910315
3. Tompkins AM, Lowe R, Nissan H, Martiny N, Roucou P, Thomson MC, et al. Predicting climate impacts on health at sub-seasonal to seasonal timescales. In: Robertson AW, Vitart F, editors. *The gap between weather and climate forecasting: sub-seasonal to seasonal prediction.* Elsevier; 2018. p. in press.
4. Chaulagai CN, Moyo CM, Koot J, Moyo HB, Sambakunsi TC, Khunga FM, et al. Design and implementation of a health management information system in Malawi: issues, innovations and results. *Health policy and planning.* 2005; 20(6):375–384. <https://doi.org/10.1093/heapol/czi044> PMID: 16143590
5. Tompkins AM, Di Giuseppe F. Potential predictability of malaria using ECMWF monthly and seasonal climate forecasts in Africa. *J Appl Meteor Clim.* 2015; 54:521–540.
6. Lunde TM, Bayoh MN, Lindtjørn B, et al. How malaria models relate temperature to malaria transmission. *Parasit Vectors.* 2013; 6(1):1–10. <https://doi.org/10.1186/1756-3305-6-20>
7. Hawkins E, Sutton R. The potential to narrow uncertainty in regional climate predictions. *Bull Amer Meteor Soc.* 2009; 90(8):1095. <https://doi.org/10.1175/2009BAMS2607.1>
8. Palmer T, Shutts G, Hagedorn R, Doblas-Reyes F, Jung T, Leutbecher M. Representing model uncertainty in weather and climate prediction. *Annu Rev Earth Planet Sci.* 2005; 33:163–193. <https://doi.org/10.1146/annurev.earth.33.092203.122552>
9. Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, et al. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature.* 2005; 433:403–406. <https://doi.org/10.1038/nature03301> PMID: 15674288
10. Molteni F, Buizza R, Palmer TN, Petroliagis T. The ECMWF ensemble prediction system: Methodology and validation. *Q J R Meteorol Soc.* 1996; 122:73–119. <https://doi.org/10.1002/qj.49712252905>
11. Buizza R, Miller M, Palmer TN. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Q J R Meteorol Soc.* 1999; 125:2887–2908. <https://doi.org/10.1002/qj.49712556006>
12. Krishnamurti TN, Kishtawal C, Zhang Z, LaRow T, Bachiochi D, Williford E, et al. Multimodel ensemble forecasts for weather and seasonal climate. *Journal of Climate.* 2000; 13(23):4196–4216. [https://doi.org/10.1175/1520-0442\(2000\)013%3C4196:MEFFWA%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013%3C4196:MEFFWA%3E2.0.CO;2)
13. Piccolo C, Cullen M. Ensemble data assimilation using a unified representation of model error. *Mon Wea Rev.* 2016; 144(1):213–224. <https://doi.org/10.1175/MWR-D-15-0270.1>
14. Kay J, Deser C, Phillips A, Mai A, Hannay C, Strand G, et al. The community earth system model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull Amer Meteor Soc.* 2015; 96(8):1333–1349. <https://doi.org/10.1175/BAMS-D-13-00255.1>
15. Tebaldi C, Knutti R. The use of the multi-model ensemble in probabilistic climate projections. *Phil Trans Roy Soc Lon A.* 2007; 365(1857):2053–2075. <https://doi.org/10.1098/rsta.2007.2076>
16. Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl GA. Challenges in combining projections from multiple climate models. *Journal of Climate.* 2010; 23(10):2739–2758. <https://doi.org/10.1175/2009JCLI3361.1>
17. Katz RW. Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Res.* 2002; 20(2):167–185. <https://doi.org/10.3354/cr020167>

18. Morse AP, Doblas-Reyes FJ, Hoshen MB, Hagedorn R, Palmer TN. A forecast quality assessment of an end-to-end probabilistic multi-model seasonal forecast system using a malaria model. *Tellus A*. 2005; 57(3):464–475. <https://doi.org/10.3402/tellusa.v57i3.14668>
19. Thomson MC, Doblas-Reyes FJ, Mason SJ, Hagedorn R, Connor SJ, Phindela T, et al. Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*. 2006; 439:576–579. <https://doi.org/10.1038/nature04503> PMID: 16452977
20. Yamana TK, Bombliès A, Eltahir EA. Climate change unlikely to increase malaria burden in West Africa. *Nature Climate Change*. 2016;. <https://doi.org/10.1038/nclimate3085>
21. Wallace DI, Southworth BS, Shi X, Chipman JW, Githeko AK. A comparison of five malaria transmission models: benchmark tests and implications for disease control. *Malar J*. 2014; 13(1):1–16. <https://doi.org/10.1186/1475-2875-13-268>
22. Ruiz D, Brun C, Connor SJ, Omumbo JA, Lyon B, Thomson MC. Testing a multi-malaria-model ensemble against 30 years of data in the Kenyan highlands. *Malar J*. 2014; 13(1):1. <https://doi.org/10.1186/1475-2875-13-206>
23. Caminade C, Kovats S, Rocklov J, Tompkins AM, Morse AP, Colón-González FJ, et al. Impact of climate change on global malaria distribution. *Proc Nat Acad Sci*. 2014; 111(9):3286–3291. <https://doi.org/10.1073/pnas.1302089111> PMID: 24596427
24. Leedale J, Tompkins AM, Caminade C, Jones AE, Nikulin G, Morse AP. Projecting malaria hazard from climate change in eastern Africa using large ensembles to estimate uncertainty. *Geospat Health*. 2016; 11 <https://doi.org/10.4081/gh.2016.393>
25. Piontek F, Müller C, Pugh TA, Clark DB, Deryng D, Elliott J, et al. Multisectoral climate impact hotspots in a warming world. *Proc Nat Acad Sci*. 2014;p. <https://doi.org/10.1073/pnas.1222471110>
26. Li SL, Bjørnstad ON, Ferrari MJ, Mummah R, Runge MC, Fonnesbeck CJ, et al. Essential information: Uncertainty and optimal control of Ebola outbreaks. *Proc Nat Acad Sci*. 2017; 114(22):5659–5664. <https://doi.org/10.1073/pnas.1617482114> PMID: 28507121
27. Blower SM, Dowlatabadi H. Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *International Statistical Review/Revue Internationale de Statistique*. 1994;p. 229–243.
28. Hoffman FO, Hammonds JS. Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Analysis*. 1994; 14(5):707–712. <https://doi.org/10.1111/j.1539-6924.1994.tb00281.x> PMID: 7800861
29. Van de Velde N, Brisson M, Boily MC. Modeling human papillomavirus vaccine effectiveness: quantifying the impact of parameter uncertainty. *American journal of epidemiology*. 2007; 165(7):762–775. <https://doi.org/10.1093/aje/kwk059> PMID: 17276976
30. Gouache D, Bensadoun A, Brun F, Pagé C, Makowski D, Wallach D. Modelling climate change impact on *Septoria tritici blotch* (STB) in France: accounting for climate model and disease model uncertainty. *Agricultural and forest meteorology*. 2013; 170:242–252. <https://doi.org/10.1016/j.agrformet.2012.04.019>
31. Lindström T, Tildesley M, Webb C. A Bayesian ensemble approach for epidemiological projections. *PLoS Comput Biol*. 2015; 11(4):e1004187. <https://doi.org/10.1371/journal.pcbi.1004187> PMID: 25927892
32. Duan Q, Sorooshian S, Gupta V. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*. 1992; 28(4):1015–1031. <https://doi.org/10.1029/91WR02985>
33. Gupta HV, Sorooshian S, Yapo PO. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resour Res*. 1998; 34(4):751–763. <https://doi.org/10.1029/97WR03495>
34. Boyle DP, Gupta HV, Sorooshian S. Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resour Res*. 2000; 36(12):3663–3674. <https://doi.org/10.1029/2000WR900207>
35. Vrugt JA, Gupta HV, Bouten W, Sorooshian S. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour Res*. 2003; 39(8):1201. <https://doi.org/10.1029/2002WR001642>
36. Larrañaga P, Karshenas H, Bielza C, Santana R. A review on evolutionary algorithms in Bayesian network learning and inference tasks. *Information Sciences*. 2013; 233:109–125. <https://doi.org/10.1016/j.ins.2012.12.051>
37. Lowe R, Bailey TC, Stephenson DB, Jupp TE, Graham RJ, Barcellos C, et al. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in South-east Brazil. *Statistics in medicine*. 2013; 32(5):864–883. <https://doi.org/10.1002/sim.5549> PMID: 22927252

38. Lowe R, Carvalho MS, Coelho CA, Barcellos C, Bailey TC, Stephenson DB, et al. Interpretation of probabilistic forecasts of epidemics. *The Lancet Infectious Diseases*. 2015; 15(1):20. [https://doi.org/10.1016/S1473-3099\(14\)71031-X](https://doi.org/10.1016/S1473-3099(14)71031-X) PMID: 25541169
39. Chitnis N, Hyman JM, Cushing JM. Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model. *Bull Math Biol*. 2008; 70:1272–1296. <https://doi.org/10.1007/s11538-008-9299-0> PMID: 18293044
40. Chitnis N, Smith T, Steketee R. A mathematical model for the dynamics of malaria in mosquitoes feeding on a heterogeneous host population. *Journal of Biological Dynamics*. 2008; 2(3):259–285. <https://doi.org/10.1080/17513750701769857> PMID: 22876869
41. Chitnis N, Schapira A, Smith T, Steketee R. Comparing the effectiveness of malaria vector-control interventions through a mathematical model. *Am J Trop Med Hyg*. 2010; 83(2):230–240. <https://doi.org/10.4269/ajtmh.2010.09-0179> PMID: 20682861
42. Fillion L, Mahfouf JF. Jacobians of an operational prognostic cloud scheme. *Mon Wea Rev*. 2003; 131:2838–2856. [https://doi.org/10.1175/1520-0493\(2003\)131%3C2838:JOAOPC%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131%3C2838:JOAOPC%3E2.0.CO;2)
43. Lu X, Wang YP, Ziehn T, Dai Y. An efficient method for global parameter sensitivity analysis and its applications to the Australian community land surface model (CABLE). *Agricultural and forest meteorology*. 2013; 182:292–303. <https://doi.org/10.1016/j.agrformet.2013.04.003>
44. Yin X, Liu J, Wang B. Nonlinear ensemble parameter perturbation for climate models. *J Climate*. 2015; 28(3):1112–1125. <https://doi.org/10.1175/JCLI-D-14-00244.1>
45. Gubbins S, Carpenter S, Baylis M, Wood JL, Mellor PS. Assessing the risk of bluetongue to UK livestock: uncertainty and sensitivity analyses of a temperature-dependent model for the basic reproduction number. *Journal of the Royal Society Interface*. 2008; 5(20):363–371. <https://doi.org/10.1098/rsif.2007.1110>
46. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. 2013; 496(7446):504–507. <https://doi.org/10.1038/nature12060> PMID: 23563266
47. Tompkins AM, Larsen L, McCreesh N, Taylor DM. To what extent does climate explain variations in reported malaria cases in early 20th century Uganda? *Geospat Health*. 2016; 11. <https://doi.org/10.4081/gh.2016.407> PMID: 27063740
48. Hoshen MB, Morse AP. A weather-driven model of malaria transmission. *Malar J*. 2004; 3:32 <https://doi.org/10.1186/1475-2875-3-32> PMID: 15350206
49. Bomblies A, Duchemin JB, Eltahir EAB. A mechanistic approach for accurate simulation of village scale malaria transmission. *Malar J*. 2009; 8:223 <https://doi.org/10.1186/1475-2875-8-223> PMID: 19799793
50. Tompkins AM, Ermert V. A regional-scale, high resolution dynamical malaria model that accounts for population density, climate and surface hydrology. *Malaria Journal*. 2013; 12 <https://doi.org/10.1186/1475-2875-12-65>
51. Lowe R, Bailey TC, Stephenson DB, Graham RJ, Coelho CAS, Sá Carvalho M, et al. Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Comput Geosci*. 2011; 37:371–381. <https://doi.org/10.1016/j.cageo.2010.01.008>
52. Lowe R, Chirombo J, Tompkins AM. Relative importance of climatic, geographic and socio-economic determinants of malaria in Malawi. *Malar J*. 2013; 12(1) <https://doi.org/10.1186/1475-2875-12-416> PMID: 24228784
53. Colón-González FJ, Tompkins AM, Biondi R, Bizimana JP, Namanya DB. Assessing the effects of air temperature and rainfall on malaria incidence: an epidemiological study across Rwanda and Uganda. *Geospat Health*. 2016; 1
54. Laneri K, Bhadra A, Ionides EL, Bouma M, Dhiman RC, Yadav RS, et al. Forcing versus feedback: epidemic malaria and monsoon rains in northwest India. *PLoS Comput Biol*. 2010;p. <https://doi.org/10.1371/journal.pcbi.1000898> PMID: 20824122
55. Hay SI, Cox J, Rogers DJ, Randolph SE, Stern DI, Shanks GD, et al. Climate change and the resurgence of malaria in the East African highlands. *Nature*. 2002; 415:905–909. <https://doi.org/10.1038/415905a> PMID: 11859368
56. Hay SI, Rogers DJ, Randolph SE, Stern DI, Cox J, Shanks GD, et al. Hot topic or hot air? Climate change and malaria resurgence in East African highlands. *Trends Parasitol*. 2002; 18:530–534. [https://doi.org/10.1016/S1471-4922\(02\)02374-7](https://doi.org/10.1016/S1471-4922(02)02374-7) PMID: 12482536
57. Shanks GD, Hay SI, Omumbo JA, Snow RW. Malaria in Kenya's western highlands. *Emerg Infect Dis*. 2005; 11(9):1425. <https://doi.org/10.3201/eid1109.041131> PMID: 16229773
58. Pascual M, Ahumada JA, Chaves LF, Rodó X, Bouma M. Malaria resurgence in the East African highlands: Temperature trends revisited. *Proc Nat Acad Sci*. 2006;p. 5829–5834. <https://doi.org/10.1073/pnas.0508929103> PMID: 16571662

59. Alonso D, Bouma MJ, Pascual M. Epidemic malaria and warmer temperatures in recent decades in an East African highland. *Proc R Soc Lond B Biol Sci.* 2011; 278:1661–1669. <https://doi.org/10.1098/rspb.2010.2020>
60. Stern DI, Gething PW, Kabaria CW, Temperley WH, Noor AM, Okiro EA, et al. Temperature and malaria trends in highland East Africa. *PLoS ONE.* 2011; 6(9):e24524. <https://doi.org/10.1371/journal.pone.0024524> PMID: 21935416
61. Tompkins AM, Di Giuseppe F, Colón-González FJ, Namanya DB. A planned operational malaria early warning system for Uganda provides useful district-scale predictions up to 4 months ahead. In: Shumake-Guillemot J, Fernandez-Montoya L, editors. *Climate services for health: Case studies of enhancing decision support for climate risk management and adaptation.* WHO/WMO Geneva; 2016. p. 130–131.
62. Tompkins AM, Caporaso L. Assessment of malaria transmission changes in Africa due to the climate impact of land use change using CMIP5 earth system models. *Geospat Health.* 2016; 11 <https://doi.org/10.4081/gh.2016.380>
63. Lindsay SW, Adiamah JH, Miller JE, Pleass RJ, Armstrong JRM. Variation in attractiveness of human subjects to malaria mosquitoes (Diptera: Culicidae) in The Gambia. *Journal of medical entomology.* 1993; 30(2):368–373. <https://doi.org/10.1093/jmedent/30.2.368> PMID: 8459413
64. Knols BGJ, de Jong R, Takken W. Differential attractiveness of isolated humans to mosquitoes in Tanzania. *Trans R Soc Trop Med Hyg.* 1995; 89:604–606. [https://doi.org/10.1016/0035-9203\(95\)90406-9](https://doi.org/10.1016/0035-9203(95)90406-9) PMID: 8594668
65. Mukabana WR, Takken W, Coe R, Knols BGJ. Host-specific cues cause differential attractiveness of Kenyan men to the African malaria vector *Anopheles gambiae*. *Malar J.* 2002; 1:17 <https://doi.org/10.1186/1475-2875-1-17> PMID: 12513703
66. Lwetoijera DW, Kiware SS, Mageni ZD, Dongus S, Harris C, Devine GJ, et al. A need for better housing to further reduce indoor malaria transmission in areas with high bed net coverage. *Parasit Vectors.* 2013; 6:57. <https://doi.org/10.1186/1756-3305-6-57> PMID: 23497471
67. Carter R, Mendis KN, Roberts D. Spatial targeting of interventions against malaria. *Bull World Health Organ.* 2000; 78(12):1401–1411. PMID: 11196487
68. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, et al. Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS medicine.* 2012; 9(1):e1001165. <https://doi.org/10.1371/journal.pmed.1001165> PMID: 22303287
69. Kienberger S, Hagenlocher M. Spatial-explicit modeling of social vulnerability to malaria in East Africa. *International journal of health geographics.* 2014; 13(1):29. <https://doi.org/10.1186/1476-072X-13-29> PMID: 25127688
70. Lacroix R, Mukabana WR, Gouagna LC, Koella JC. Malaria infection increases attractiveness of humans to mosquitoes. *PLoS Biol.* 2005; 3(9):e298. <https://doi.org/10.1371/journal.pbio.0030298> PMID: 16076240
71. Lunde TM, Korecha D, Loha E, Sorteberg A, Lindtjörn B. A dynamic model of some malaria-transmitting anopheline mosquitoes of the Afrotropical region. I. Model description and sensitivity analysis. *Malar J.* 2013; 12(1)
72. Filipe JAN, Riley EM, Drakeley CJ, Sutherland CJ, Ghani AC. Determination of the processes driving the acquisition of immunity to malaria using a mathematical transmission model. *PLoS Comput Biol.* 2007; 3:e255 <https://doi.org/10.1371/journal.pcbi.0030255> PMID: 18166074
73. Doolan DL, Dobano C, Baird JK. Acquired immunity to malaria. *Clin Microbiol Rev.* 2009; 22:13 <https://doi.org/10.1128/CMR.00025-08> PMID: 19136431
74. Klein EY, Smith DL, Boni MF, Laxminarayan R. Clinically immune hosts as a refuge for drug-sensitive malaria parasites. *Malar J.* 2008; 7(1):1. <https://doi.org/10.1186/1475-2875-7-67>
75. Boudin C, Diop A, Gaye A, Gadiaga L, Gouagna C, Safeukui I, et al. *Plasmodium falciparum* transmission blocking immunity in three areas with perennial or seasonal endemicity and different levels of transmission. *Am J Trop Med Hyg.* 2005; 73:1090–1095. <https://doi.org/10.4269/ajtmh.2005.73.1090> PMID: 16354818
76. Ermert V, Fink AH, Jones AE, Morse AP. Development of a new version of the Liverpool Malaria Model. I. Refining the parameter settings and mathematical formulation of basic processes based on a literature review. *Malar J.* 2011; 10 <https://doi.org/10.1186/1475-2875-10-35>
77. Omumbo JA, Lyon B, Waweru SM, Connor SJ, Thomson MC. Raised temperatures over the Kericho tea estates: revisiting the climate in the East African highlands malaria debate. *Malar J.* 2011; 10(1):12. <https://doi.org/10.1186/1475-2875-10-12> PMID: 21241505
78. Lee Y, Park S, Chang DE. Parameter estimation using the genetic algorithm and its impact on quantitative precipitation forecast. In: *Annales Geophysicae.* vol. 24. Copernicus GmbH; 2006. p. 3185–3189. <https://doi.org/10.5194/angeo-24-3185-2006>

79. Karydis V, Capps S, Russell A, Nenes A. Adjoint sensitivity of global cloud droplet number to aerosol and dynamical parameters. *Atmos Chem Phys*. 2012; 12(19):9041–9055. <https://doi.org/10.5194/acp-12-9041-2012>
80. Liong SY, Chan WT, ShreeRam J. Peak-flow forecasting with genetic algorithm and SWMM. *Journal of Hydraulic Engineering*. 1995; 121(8):613–617. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1995\)121:8\(613\)](https://doi.org/10.1061/(ASCE)0733-9429(1995)121:8(613))
81. Wang Q. Using genetic algorithms to optimise model parameters. *Environmental Modelling & Software*. 1997; 12(1):27–34. [https://doi.org/10.1016/S1364-8152\(96\)00030-8](https://doi.org/10.1016/S1364-8152(96)00030-8)
82. Ihshaish H, Cortés A, Senar MA. Parallel multi-level genetic ensemble for numerical weather prediction enhancement. *Procedia Computer Science*. 2012; 9:276–285. <https://doi.org/10.1016/j.procs.2012.04.029>
83. Wang L, He WP, Liao LJ, Wan SQ, He T. A new method for parameter estimation in nonlinear dynamical equations. *Theoretical and Applied Climatology*. 2015; 119(1-2):193–202. <https://doi.org/10.1007/s00704-014-1113-3>
84. Lazzús JA, Rivera M, López-Caraballo CH. Parameter estimation of Lorenz chaotic system using a hybrid swarm intelligence algorithm. *Physics Letters A*. 2016; 380(11):1164–1171.
85. Ionides EL, Bretó C, King AA. Inference for nonlinear dynamical systems. *Proc Nat Acad Sci*. 2006; 103(49):18438–18443. <https://doi.org/10.1073/pnas.0603181103> PMID: 17121996
86. Reiner RC, Geary M, Atkinson PM, Smith DL, Gething PW. Seasonality of *Plasmodium falciparum* transmission: a systematic review. *Malar J*. 2015; 14(1):1. <https://doi.org/10.1186/s12936-015-0849-2>
87. Courtier P, Thepaut JN, Hollingsworth A. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q J R Meteorol Soc*. 1994; 120:1367–1387. <https://doi.org/10.1002/qj.49712051912>
88. Holm E, Andersson E, Beljaars A, Lopez P, Mahfouf JF, Simmons AJ, et al. Assimilation and Modeling of the Hydrological Cycle: ECMWF's Status and Plans. available at <http://www.ecmwf.int/publications>: European Centre for Medium-Range Weather Forecasts,; 2002. pp55.
89. Brohan P, Kennedy JJ, Harris I, Tett SF, Jones PD. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J Geophys Res*. 2006; 111(D12). <https://doi.org/10.1029/2005JD006548>
90. Folland CK, Rayner NA, Brown S, Smith T, Shen S, Parker D, et al. Global temperature change and its uncertainties since 1861. *Geophys Res Lett*. 2001; 28(13):2621–2624. <https://doi.org/10.1029/2001GL012877>
91. Lindblade KA, Walker ED, Onapa AW, Katungu J, Wilson ML. Highland malaria in Uganda: prospective analysis of an epidemic associated with El Niño. *Trans R Soc Trop Med Hyg*. 1999; 93(5):480–487. [https://doi.org/10.1016/S0035-9203\(99\)90344-9](https://doi.org/10.1016/S0035-9203(99)90344-9) PMID: 10696401
92. Lindblade KA, Walker ED, Onapa AW, Katungu J, Wilson ML. Land use change alters malaria transmission parameters by modifying temperature in a highland area of Uganda. *Tropical Medicine & International Health*. 2000; 5(4):263–274. <https://doi.org/10.1046/j.1365-3156.2000.00551.x>
93. Munga S, Yakob L, Mushinzimana E, Zhou G, Ouna T, Minakawa N, et al. Land use and land cover changes and spatiotemporal dynamics of anopheline larval habitats during a four-year period in a highland community of Africa. *Am J Trop Med Hyg*. 2009; 81(6):1079–1084. <https://doi.org/10.4269/ajtmh.2009.09-0156> PMID: 19996440
94. Paaijmans KP, Blanford S, Bell AS, Blanford JI, Read AF, Thomas MB. Influence of climate on malaria transmission depends on daily temperature variation. *Proc Nat Acad Sci*. 2010; 107:15135–15139. <https://doi.org/10.1073/pnas.1006422107> PMID: 20696913
95. Minakawa N, Munga S, Atieli F, Mushinzimana E, Zhou G, Githeko AK, et al. Spatial distribution of Anopheline larval habitats in Western Kenyan highlands: effects of land cover types and topography. *Am J Trop Med Hyg*. 2005; 73(1):157–165. <https://doi.org/10.4269/ajtmh.2005.73.157> PMID: 16014851
96. Afrane YA, Lawson BW, Githeko AK, Yan G. Effects of microclimatic changes caused by land use and land cover on duration of gonotrophic cycles of *Anopheles gambiae* (Diptera: Culicidae) in western Kenya highlands. *J Med Entomol*. 2005; 42(6):974–980. <https://doi.org/10.1093/jmedent/42.6.974> PMID: 16465737
97. Afrane YA, Zhou G, Lawson BW, Githeko AK, Yan G. Effects of microclimatic changes caused by deforestation on the survivorship and reproductive fitness of *Anopheles gambiae* in western Kenya highlands. *Am J Trop Med Hyg*. 2006; 74(5):772–778. <https://doi.org/10.4269/ajtmh.2006.74.772> PMID: 16687679
98. Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc*. 2011; 137(656):553–597. <https://doi.org/10.1002/qj.828>

99. Paaijmans KP, Read AF, Thomas MB. Understanding the link between malaria risk and climate. *Proc Nat Acad Sci*. 2009; 106(33):13844–13849. <https://doi.org/10.1073/pnas.0903423106> PMID: 19666598
100. McCollum AG J R, Ba MB. Discrepancy between gauge and satellite estimates of rainfall in equatorial Africa. *J Appl Meteor*. 2000; 39:666–679. <https://doi.org/10.1175/1520-0450-39.5.666>
101. Thorne V, Grimes D, Dugdale G. Comparison of TAMSAT and CPC rainfall estimates with raingauges, for southern Africa. *International Journal of Remote Sensing*. 2001; 22(24):1951–1974. <https://doi.org/10.1080/01431160118816>
102. Law KB, Janowiak JE, Huffman GJ. Verification of rainfall estimates over Africa using RFE, NASA MPA-RT, and CMORPH. available at <http://www.cpc.noaa.gov/products/fews/>; NOAA Climate Prediction Center; 2002.
103. Nicholson SE, Some B, McCollum J, Nelkin E, Klotter D, Berte Y, et al. Validation of TRMM and other rainfall estimates with a high-density gauge dataset for West Africa. Part II: Validation of TRMM rainfall products. *J Appl Meteor*. 2003; 42(10):1355–1368. [https://doi.org/10.1175/1520-0450\(2003\)042%3C1355:VOTAOR%3E2.0.CO;2](https://doi.org/10.1175/1520-0450(2003)042%3C1355:VOTAOR%3E2.0.CO;2)
104. Adeyewa ZD, Nakamura K. Validation of TRMM radar rainfall data over major climatic regions in Africa. *J Appl Meteor*. 2003; 42(2):331–347. [https://doi.org/10.1175/1520-0450\(2003\)042%3C0331:VOTRRD%3E2.0.CO;2](https://doi.org/10.1175/1520-0450(2003)042%3C0331:VOTRRD%3E2.0.CO;2)
105. Ali A, Amani A, Diedhiou A, Lebel T. Rainfall estimation in the Sahel. Part II: Evaluation of rain gauge networks in the CILSS countries and objective intercomparison of rainfall products. *J Appl Meteor*. 2005; 44(11):1707–1722. <https://doi.org/10.1175/JAM2305.1>
106. Dinku T, Ceccato P, Grover-Kopec E, Lemma M, Connor SJ, Ropelewski CF. Validation of satellite rainfall products over East Africa's complex topography. *Int J Remote Sens*. 2007; 28:1503–1526. <https://doi.org/10.1080/01431160600954688>
107. Villarini G, Mandapaka PV, Krajewski WF, Moore RJ. Rainfall and sampling uncertainties: A rain gauge perspective. *J Geophys Res*. 2008; 113(D11):D111102. <https://doi.org/10.1029/2007JD009214>
108. Tompkins AM, Adebisi AA. Using CloudSat cloud retrievals to differentiate satellite-derived rainfall products over West Africa. *J Hydrometeor*. 2012; 13:1810–1816. <https://doi.org/10.1175/JHM-D-12-039.1>
109. Dinku T, Alessandrini S, Evangelisti M, Rojas O. A description and evaluation of FAO satellite rainfall estimation algorithm. *Atmos Res*. 2015; 163:48–60. <https://doi.org/10.1016/j.atmosres.2015.01.020>
110. Taylor CM, Lebel T. Observational evidence of persistent convective-scale rainfall patterns. *Mon Wea Rev*. 1998; 126:1597–1607. [https://doi.org/10.1175/1520-0493\(1998\)126%3C1597:OEOPCS%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126%3C1597:OEOPCS%3E2.0.CO;2)
111. Taylor CM, Parker DJ, Harris PP. An observational case study of mesoscale atmospheric circulations induced by soil moisture. *Geophys Res Lett*. 2007; 34(15). <https://doi.org/10.1029/2007GL030572>
112. Asare EO, Tompkins AM, Amekudzi LK, Ermert V. A breeding site model for regional, dynamical malaria simulations evaluated using in situ temporary ponds observations. *Geospat Health*. 2016; 11 <https://doi.org/10.4081/gh.2016.390>
113. Asare E, Tompkins AM, Bomblies A. Evaluation of a simple puddle breeding site model for malaria vectors using high resolution explicit surface hydrology simulations. *PLoS ONE*. 2016; 11:<http://dx.doi.org/10.1371/journal.pone.0150626>.
114. Cazelles B, Chavez M, de Magny GC, Guégan JF, Hales S. Time-dependent spectral analysis of epidemiological time-series with wavelets. *Journal of the Royal Society Interface*. 2007; 4(15):625–636. <https://doi.org/10.1098/rsif.2007.0212>
115. Roesch A, Schmidbauer H. WaveletComp: Computational Wavelet Analysis; 2014. R package version 1.0. Available from: http://www.hs-stat.com/projects/WaveletComp/WaveletComp_guided_tour.pdf.
116. Warner TT, Peterson RA, Treadon RE. A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull Amer Meteor Soc*. 1997; 78(11):2599–2617. [https://doi.org/10.1175/1520-0477\(1997\)078%3C2599:ATOLBC%3E2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078%3C2599:ATOLBC%3E2.0.CO;2)
117. Wu W, Lynch AH, Rivers A. Estimating the uncertainty in a regional climate model related to initial and lateral boundary conditions. *J Climate*. 2005; 18(7):917–933. <https://doi.org/10.1175/JCLI-3293.1>
118. Vié B, Nuissier O, Ducrocq V. Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: uncertainty on initial conditions and lateral boundary conditions. *Mon Wea Rev*. 2011; 139(2):403–423. <https://doi.org/10.1175/2010MWR3487.1>
119. Hay SI, Were EC, Renshaw M, Noor AM, Ochola SA, Olusanmi I, et al. Forecasting, warning, and detection of malaria epidemics: a case study. *The Lancet*. 2003; 361:1705–1706. [https://doi.org/10.1016/S0140-6736\(03\)13366-1](https://doi.org/10.1016/S0140-6736(03)13366-1)

120. Craig MH, Snow RW, le Sueur D. A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitol Today*. 1999; 15:105–111. [https://doi.org/10.1016/S0169-4758\(99\)01396-4](https://doi.org/10.1016/S0169-4758(99)01396-4) PMID: [10322323](https://pubmed.ncbi.nlm.nih.gov/10322323/)
121. Mordecai EA, Paaijmans KP, Johnson LR, Balzer C, Ben-Horin T, Moor E, et al. Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecol Lett*. 2013; 16(1):22–30. <https://doi.org/10.1111/ele.12015> PMID: [23050931](https://pubmed.ncbi.nlm.nih.gov/23050931/)
122. Kilian A, Langi P, Talisuna A, Kabagambe G. Rainfall pattern, El Niño and malaria in Uganda. *Trans R Soc Trop Med Hyg*. 1999; 93(1):22–23. [https://doi.org/10.1016/S0035-9203\(99\)90165-7](https://doi.org/10.1016/S0035-9203(99)90165-7) PMID: [10492781](https://pubmed.ncbi.nlm.nih.gov/10492781/)
123. Grover-Kopec E, Kawano M, Klaver RW, Blumenthal B, Ceccato P, Connor SJ. An online operational rainfall-monitoring resource for epidemic malaria early warning systems in Africa. *Malar J*. 2005; 4:6 <https://doi.org/10.1186/1475-2875-4-6> PMID: [15663795](https://pubmed.ncbi.nlm.nih.gov/15663795/)
124. Wanjala CL, Waitumbi J, Zhou G, Githeko AK. Identification of malaria transmission and epidemic hot-spots in the western Kenya highlands: its application to malaria epidemic prediction. *Parasites & vectors*. 2011; 4(1):1.
125. Kleinschmidt I, Sharp B, Clarke G, Curtis B, Fraser C. Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in KwaZulu Natal, South Africa. *American Journal of Epidemiology*. 2001; 153(12):1213–1221. <https://doi.org/10.1093/aje/153.12.1213> PMID: [11415957](https://pubmed.ncbi.nlm.nih.gov/11415957/)