Spatial and Spatio-temporal Epidemiology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Statistical methods for linking geostatistical maps and transmission models: Application to lymphatic filariasis in East Africa

Panayiota Touloupou[a,*], Renata Retkute[b], T. Déirdre Hollingsworth[c], Simon E.F. Spencer[d,e]

[a] School of Mathematics, University of Birmingham, Birmingham, UK
[b] Department of Plant Sciences, University of Cambridge, Cambridge, UK
[c] Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK
[d] Department of Statistics, University of Warwick, Coventry, UK
[e] Zeeman Institute, University of Warwick, Coventry, UK

ARTICLE INFO

Article history:
Received 14 February 2020
Revised 6 November 2020
Accepted 6 November 2020
Available online xxx

Keywords:
Bayesian methods
Fine-scale spatial predictions
Linking maps with models
Lymphatic filariasis
Projections
Uncertainty

ABSTRACT

Infectious diseases remain one of the major causes of human mortality and suffering. Mathematical models have been established as an important tool for capturing the features that drive the spread of the disease, predicting the progression of an epidemic and hence guiding the development of strategies to control it. Another important area of epidemiological interest is the development of geostatistical methods for the analysis of data from spatially referenced prevalence surveys. Maps of prevalence are useful, not only for enabling a more precise disease risk stratification, but also for guiding the planning of more reliable spatial control programmes by identifying affected areas. Despite the methodological advances that have been made in each area independently, efforts to link transmission models and geostatistical maps have been limited. Motivated by this fact, we developed a Bayesian approach that combines fine-scale geostatistical maps of disease prevalence with transmission models to provide quantitative, spatially-explicit projections of the current and future impact of control programs against a disease. These estimates can then be used at a local level to identify the effectiveness of suggested intervention schemes and allow investigation of alternative strategies. The methodology has been applied to lymphatic filariasis in East Africa to provide estimates of the impact of different intervention strategies against the disease.

1. Introduction

Geostatistical modelling is increasingly used in epidemiology to combine surveys from multiple locations into a detailed model of local prevalence or incidence (Hay et al. 2009; Moraga et al. 2015; O'Hanlon et al. 2016; Stensgaard et al. 2011; Giorgi et al. 2018). Maps of disease distribution can be used, for example, to plan the development of national scale control strategies by informing policy makers where intervention efforts should be focused (Slater and Michael, 2013; Tatem et al., 2010). Several examples from the literature have shown that spatial heterogeneity is an important epidemiological factor in many diseases (Pullan et al., 2012; Sturrock et al., 2010, for example). However, predictions of future cases are frequently performed on aggregated data, risking the ecological fallacy (Wakefield and Lyons, 2010).

Over the past decades, mathematical models have also been established as an important tool for evaluating the effect of different control strategies by predicting the progression of the disease (Ferguson et al., 2005; Hollingsworth, 2018; Stolk et al., 2018; Tildesley et al., 2009). However, when mathematical modelling is used to evaluate potential intervention strategies, spatial heterogeneity is also frequently ignored (Heesterbeek et al., 2015). Some notable exceptions are the papers by Gibson (1997), Keeling et al. (2001) and Deardon et al. (2010) who considered a spatial model, where the transmission probabilities depend on distances between individuals. In this paper we develop a novel method for taking the output from a geostatistical model and projecting the epidemic dynamics forward in time at the pixel level, under a range of potential intervention strategies, in a computationally efficient way. An important feature of our approach is the ability to capture several sources of uncertainty.

There are only a limited number of studies linking transmission models and geostatistical maps in a way that can dynamically inform policy at a local level. The African Program for Onchocerciasis Control was one of the first groups to develop and

* Corresponding author.
E-mail addresses: P.Touloupou@bham.ac.uk (P. Touloupou), rr614@cam.ac.uk (R. Retkute), Deirdre.Hollingsworth@bdi.ox.ac.uk (T.D. Hollingsworth), S.E.F.Spencer@warwick.ac.uk (S.E.F. Spencer).

https://doi.org/10.1016/j.sste.2020.100391

apply this approach (for example Alley et al., 1994; Plaisier et al., 1991). Kriging was used to extrapolate between survey points and then transmission models were used to project the likely impact of intervention programs. These mapped projections had been extremely useful in informing policy planning over many years and have recently been updated (Tekle et al., 2016). The power of this type of approach to inform policy has been illustrated most notably by Bhatt et al. (2015) in the analysis of the key drivers of successes in malaria interventions over the last 15 years. An important challenge, addressed by our approach, is to appropriately estimate and communicate the projections with their uncertainty. In particular, our method addresses and quantifies a broad range of uncertainties, including uncertainty in the spatial variation in prevalence, transmission parameters, demographics, interventions and even model structure, and propagates them into the uncertainty in future predictions.

Our methodology has many parallels with exact versions of Approximate Bayesian Computation (ABC; Beaumont et al., 2002; Wilkinson, 2013), in which simulations from the model are weighted (or accepted) according to their likelihood of producing the observed data. However, in our framework a likelihood is only available at the survey points, and so instead we weight the simulations by the posterior distribution from a geostatistical model that interpolates between surveys to give a prevalence distribution at each location. To achieve this weighting, we must change the measure of the simulated prevalences from the one induced by the prior on the transmission model parameters, to the posterior distribution from the geostatistical model using the Radon-Nikodym derivative (Billingsley, 1995). However, since this is not available in analytical form, we propose an empirical alternative similar to Goldie and Maller (1999) and references contained within.

The paper is organised as follows. In Section 2 we describe the statistical methodology for combining geo-statistical mapping and transmission modelling, and illustrate its key features with a toy example in Section 3. The proposed method is applied in Section 4 to investigate the impact of intervention programs for lymphatic filariasis in seven countries in Africa. Finally, we conclude with a discussion on limitations of the current method and possible extensions for further research in Section 5.

2. Methods

We develop a Bayesian methodology that captures uncertainty from multiple sources and can be readily applied to different transmission models and intervention strategies to give a distribution of projections across space. The starting point for our analysis is the output from a geostatistical model of disease prevalence, capturing the uncertainty in the spatial distribution of infection. A number of recent studies have adopted a predictive framework known as model-based geostatistics (Diggle and Ribeiro, 2007) for the production of prevalence maps, often employing Bayesian inference for spatial prediction and robust characterisation of uncertainty surrounding those predictions. In particular we assume that the output consists of M Monte Carlo samples from the posterior distribution of the geostatistical model. Although we assume that the spatial distribution represents the pre-control prevalence here, our methodology can easily be generalised beyond this example. In addition, we assume that other geographical information is available for each pixel (with associated uncertainty), such as population size and other demographic data that can be used as an input to the transmission model.

Our methodology consists of 3 steps. First, we generate a large number of simulations from the transmission model, with sufficient variability to capture all of the endemic prevalences observed in the samples from the geostatistical model. Second, for each spatial location in the map we reweight the simulations according to how similar they are to the observed prevalence and other spatial information, such as population data. Finally, we simulate the transmission model further forward in time, possibly under some intervention strategy, and apply the weights to obtain the spatial distribution of the projections. A graphical representation of the method can be found in Fig. 1.

2.1. Step 1: simulating from the transmission model

For each pixel on the map, we assign an informative prior on the model parameters, $\pi_i(\boldsymbol{\theta})$ say for pixel i, representing the uncertainty in our beliefs about the parameters of the transmission model at that location. Next, we define a single proposal distribution over the parameter space, $q(\boldsymbol{\theta})$, capable of producing simulated prevalence levels spanning the values observed in the geostatistical mapping. We then draw J parameter vectors $(\boldsymbol{\theta}_j)$ from the proposal, and for each one we run the model forward in time until it reaches pre-control equilibrium. Denote the resulting prevalence levels by p_j, for $j = 1, \ldots J$. Finally, we calculate an initial $I \times J$ matrix of weights for the I pixels and J simulations according to the usual importance sampling formula, namely $w_{ij}^{(1)} = \pi_i(\boldsymbol{\theta}_j)/q(\boldsymbol{\theta}_j)$.

The proposal distribution $q(\boldsymbol{\theta})$ might be uniform over the parameter space in low dimensions, or for higher dimensions it could be developed from pilot simulations, where parameter vectors are sampled uniformly from the support of the priors and for each parameter vector the equilibrium prevalence is simulated from the transmission model. The importance proposal can then be constructed on the parameter space to give more weight to frequently observed prevalence values, and zero weight to implausible prevalence values (e.g. prevalences larger than the maximum observed in the geostatistical model). The efficiency of the proposal can be improved iteratively using adaptive importance sampling techniques (Cornuet et al., 2012; Retkute et al., 2020).

2.2. Step 2: reweighting the simulations to match pixel prevalence distributions

For each pixel the same simulations are reweighted to match the prevalence distribution of that pixel. This prevents unnecessary replication of simulations for pixels that are broadly similar and means that the number of simulations need not increase as the number of pixels increases. More specifically, for pixel $i = 1, \ldots, I$ and simulation $j = 1, \ldots, J$ the new weight is given by:

$$w_{ij}^{(2)} \propto \frac{f(p_j|\boldsymbol{d}_i)}{g(p_j|\boldsymbol{w}_i^{(1)})} w_{ij}^{(1)} \tag{1}$$

where $\boldsymbol{d}_i = (d_{i1}, \ldots, d_{iM})$ is the M dimensional vector of posterior samples of prevalences in pixel i and $\boldsymbol{w}_i^{(1)} = (w_{i1}^{(1)}, \ldots, w_{iJ}^{(1)})$. The function f represents the probability of having prevalence p_j under the geostatistical model and g represents the probability of simulating prevalence p_j from the model with parameter vector drawn from the prior. The ratio f/g therefore represents the usual change of measure formula (Radon-Nikodym derivative). However, since neither of these probability densities are likely to be available in closed form, we use an empirical approximation given by the amount of probability density within $\delta/2$ of p_j:

$$f(p_j|\boldsymbol{d}_i) = \frac{1}{\delta M} \sum_{m=1}^{M} \mathbb{1}_{\{p_j - \delta/2 \le d_{im} \le p_j + \delta/2\}},$$

$$g(p_j|\boldsymbol{w}_i^{(1)}) = \frac{\sum_{k=1}^{J} w_{ik}^{(1)} \mathbb{1}_{\{p_j - \delta/2 \le p_k \le p_j + \delta/2\}}}{\delta \sum_{k=1}^{J} w_{ik}^{(1)}}.$$

Note that as long as $q(\boldsymbol{\theta}) > 0$ implies $\pi_i(\boldsymbol{\theta}) > 0$, then $w_{ij}^{(1)} > 0$ for all j and hence $g(p_j|\boldsymbol{w}_i^{(1)}) > 0$ for all j. The bin width δ controls the trade-off between effective sample size and the fidelity

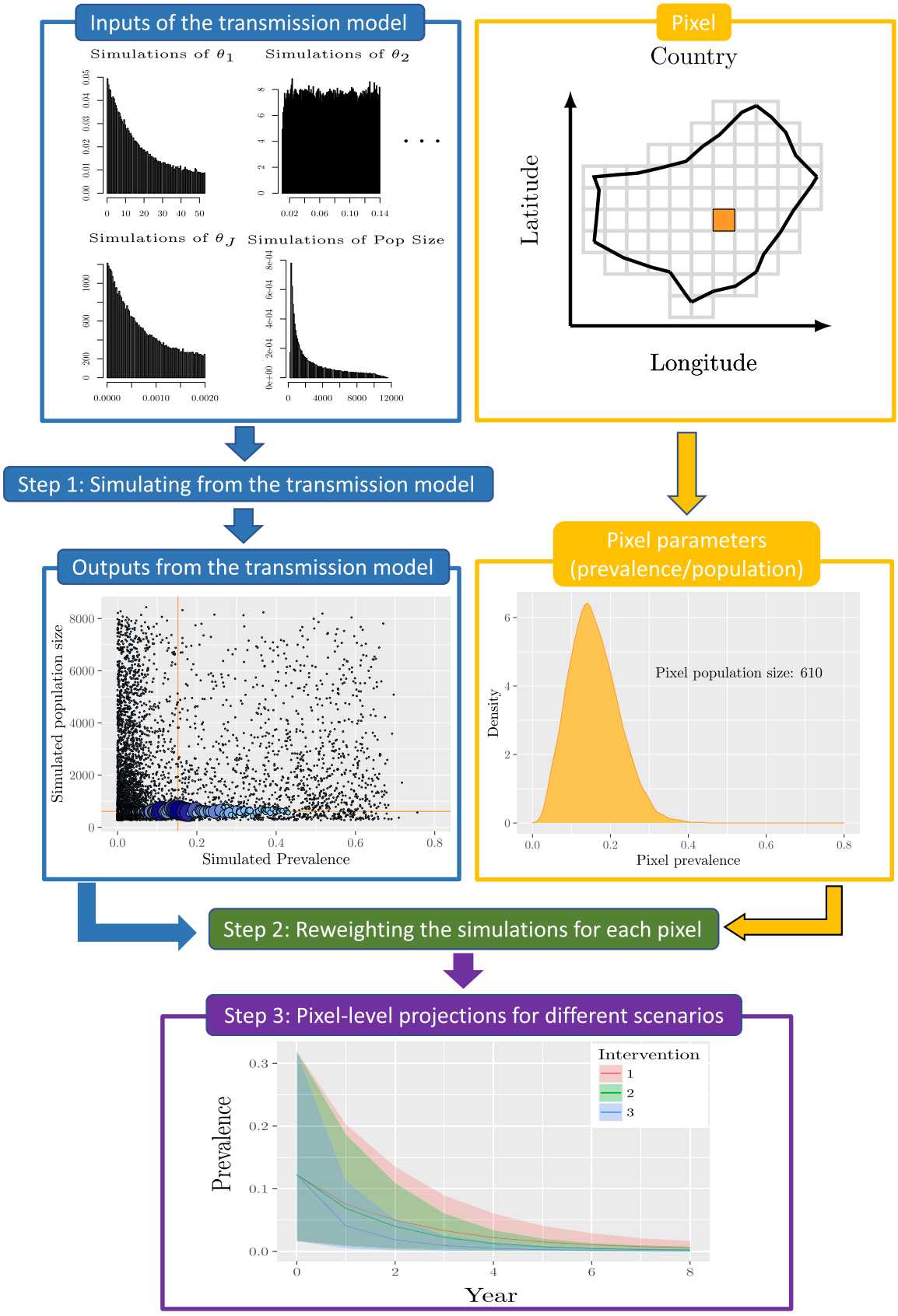


Fig. 1. Methodology for generating mapping results. Using pre-run model simulations (top), we reweight the simulations for each pixel based on the prevalence and population information (middle – red lines represent the median of observed data). Finally, the weights are used to evaluate the impact of different intervention strategies (bottom).

of the distribution of the simulated prevalences to the geostatistical posterior distribution, and should be set as small as possible, whilst producing a reasonable effective sample size, defined as $(\sum_{j=1}^{J} w_{ij}^{(2)})^2/\sum_{j=1}^{J}(w_{ij}^{(2)})^2$. Finally, the weights from Eq. (1) are normalised to give the posterior probabilities (according to the geostatistical model) that simulation j is appropriate for pixel i.

2.3. Step 3: running the simulations forward

The simulations are run forward in time under a given intervention strategy. For each pixel the projected outcomes are weighted according to the normalised weights $\boldsymbol{w}_i^{(2)} = (w_{i1}^{(2)}, \ldots, w_{iJ}^{(2)})$ produced in Step 2. Step 3 is repeated for each intervention strategy under consideration.

2.4. Lemma on the change of measure

In this section we introduce a lemma that generates the reweighting formula in Eq. (1). The lemma is proved in Appendix A.1 of the Supplementary Material (SM).

Lemma 1. *Let $p : \mathbb{R}^d \to [0, 1]$ denote a deterministic model that produces a prevalence $p(\boldsymbol{\theta})$ from a vector of parameters $\boldsymbol{\theta}$. Let $\pi(\boldsymbol{\theta})$ be a prior distribution over the parameters that induces a prior distribution over prevalences, which we denote by $g(p)$. Suppose that there exists a differentiable and invertible function $\boldsymbol{\phi} : \mathbb{R}^d \to \mathbb{R}^d$ that admits the prevalence as its first argument, ie. $\boldsymbol{\phi}(\boldsymbol{\theta}) = (p(\boldsymbol{\theta}), \boldsymbol{q}(\boldsymbol{\theta}))$ for some $\boldsymbol{q}(\boldsymbol{\theta})$. Finally suppose that we wish to change the probability measure over prevalences from g to another measure f that is absolutely continuous with respect to g. Then the resulting measure over the parameter space is given by $h(\boldsymbol{\theta}) = \frac{f(p(\boldsymbol{\theta}))}{g(p(\boldsymbol{\theta}))}\pi(\boldsymbol{\theta})$.*

Notes:

1. The same approach can be applied for stochastic transmission models as long as the model is defined on a separate probability space $(\Omega, \mathcal{F}, P)$ to the prior. For stochastic models we fix $\omega \in \Omega$ and consider the transmission model as a deterministic map $\boldsymbol{\phi}(\boldsymbol{\theta}, \omega)$, applying the Lemma and then integrating over Ω.
2. The condition that f must be absolutely continuous with respect to g means that whenever $g(p) = 0$ then we must also have $f(p) = 0$. In other words, when the prior probability of a prevalence is zero then the map measure of prevalence must also be zero. This has important implications for the implementation of our method, discussed further in Appendix A.2.

2.5. Alternative empirical Radon-Nikodym derivatives

In Step 2 of our algorithm (described in Section 2.2) we proposed an empirical estimate of the Radon-Nikodym derivative f/g based on using the prevalences within $\delta/2$ of p_j. Clearly there are many possible alternative estimates that could be used and there are two in particular that are worthy of further discussion. The first is based on histograms and the second is based on minimising a discrepancy measure.

2.5.1. Histogram-based empirical Radon-Nikodym derivative

If we consider a fixed partition of the prevalence space into bins (as if we were constructing a histogram) then it is straightforward to calculate the Radon-Nikodym derivative f/g for each bin as the proportion of posterior samples in the bin divided by the proportion of the weight belonging to simulations that fall in the bin. More precisely, given a finite set of disjoint intervals with union [0,1] then if prevalence p_j falls in interval $\mathcal{I}(p_j)$ we have that

$$f(p_j|\boldsymbol{d}_i) = \frac{1}{M|\mathcal{I}(p_j)|}\sum_{m=1}^{M} \mathbb{1}_{\{d_{im} \in \mathcal{I}(p_j)\}},$$

$$g(p_j|\boldsymbol{w}_i^{(1)}) = \frac{\sum_{k=1}^{J} w_{ik}^{(1)} \mathbb{1}_{\{p_k \in \mathcal{I}(p_j)\}}}{|\mathcal{I}(p_j)| \sum_{k=1}^{J} w_{ik}^{(1)}},$$

where $|\mathcal{I}(p_j)|$ is the length of the interval containing p_j.

The main advantage of this estimate is computational – since all of the simulations in the same interval have the same ratio (for a given pixel) then instead of having to calculate J ratios we need only calculate one per interval. A secondary advantage is that the weighted histogram of the simulation prevalences will be identical to the histogram of the posterior prevalence distribution. However, the relative weightings within each bin are unchanged and so a different choice of bins will reveal that the two distributions are different.

2.5.2. Discrepancy-based empirical Radon-Nikodym derivative

A second alternative empirical Radon-Nikodym derivative can be defined to minimise the difference between the empirical cumulative distribution functions (cdfs) of the posterior prevalences and the weighted simulated prevalences. Let $F(x|\boldsymbol{d}_i) = \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}_{\{d_{im} \le x\}}$ be the empirical cdf of the map prevalence distribution for pixel i and $H(x|\boldsymbol{w}_i^{(2)}) = \sum_{j=1}^{J} w_{ij}^{(2)} \mathbb{1}_{\{p_j \le x\}}$ be the empirical cdf of the final weighted distribution of simulated prevalences, then we can choose $\boldsymbol{w}_i^{(2)}$ to minimise some distance $||F(\cdot|\boldsymbol{d}_i) - H(\cdot|\boldsymbol{w}_i^{(2)})||$. For example, we may wish to minimise $\int_0^1 |F(x|\boldsymbol{d}_i) - H(x|\boldsymbol{w}_i^{(2)})|\,\mathrm{d}x$ or $\int_0^1 \left(F(x|\boldsymbol{d}_i) - H(x|\boldsymbol{w}_i^{(2)})\right)^2 \mathrm{d}x$. In this paper we have focussed on the latter of these, for details of the calculation we refer the reader to the SM Appendix A.3.

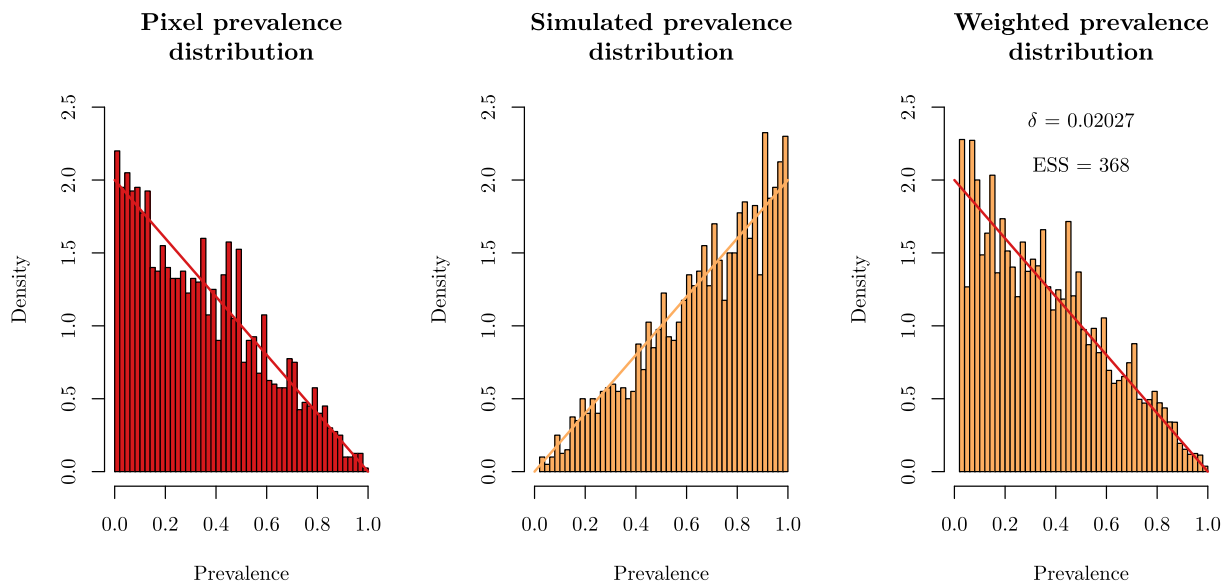
3. Simulation studies: a toy example

In this section we provide a toy example to assess the performance of the proposed method under different settings. Particular focus was given on how the method was affected by the value of δ, by the choice of the proposal distribution of the parameters and the empirical estimate of the Radon-Nikodym derivative. A full description of the analysis can be found in SM Appendix B and here we summarize the key results.

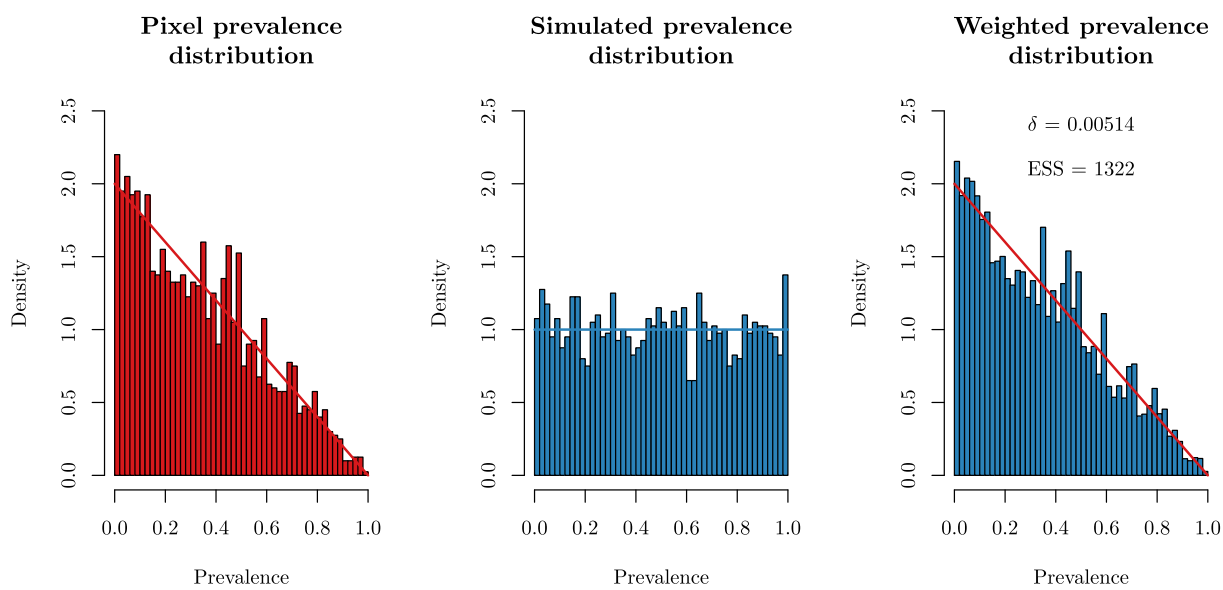
Suppose that the prior distribution is $\pi(\theta_1, \theta_2) = 2$ if $0 < \theta_2 < \theta_1 < 1$ and zero otherwise. Plots illustrating this prior are given in SM Fig. B.2. For simplicity, assume that the transmission model has equilibrium prevalence given by $p(\theta_1, \theta_2) = \theta_1$ so that the induced prior over prevalences is the marginal for θ_1, ie. $g(p) = 2p$ for $0 < p < 1$, which is a Beta(2,1) distribution. Further, suppose that we are given $M = 2000$ samples from a pixel with prevalence measure $f(p) = 2(1 - p)$ for $0 < p < 1$, representing a Beta(1,2) distribution. This challenging example allows us to assess how the methodology performs when there are few simulations with low weights in the area of high posterior probability close to $p = 0$.

Simulations were conducted to investigate the accuracy and efficiency of the proposed method under different settings, where the observed pixel and simulated prevalence data are obtained from the toy model. Fig. 2 shows how $J = 2000$ simulations from a proposal (centre histogram) can be reweighted (right histogram) to resemble the pixel prevalence distribution (left histogram). In Fig. 2(a) the proposal is from the prior, whilst in Fig. 2(b) the proposal is $U(0, 1)$. The improvement due to the proposal distribution having good support in all areas of the posterior distribution was demonstrated by the substantial increase in effective sample size (ESS; from 368 to 1322), despite a much smaller value of δ.

Fig. 3 illustrates how the performance changes as δ is increased. The left figure shows the distance (given by integrated squared difference) between the empirical cumulative distribution functions of the weighted simulations and the samples from the pixel posterior; and the right figure shows the effective sample sizes. The



(a) The proposal yields a simulated prevalence distribution equal to its marginal prior, Beta(2,1).



(b) The proposal yields a uniform simulated prevalence distribution.

Fig. 2. The estimated weighted prevalence distribution for the suggested value of δ (right panel) is compared to the true pixel prevalence distribution (left panel), under different proposal distributions for the prevalence (middle panel): (a) Beta(2,1) and (b) $U(0,1)$. The target densities are also shown on each panel.

corresponding results from the discrepancy based empirical Radon-Nikodym derivative (see Section 2.5.2) are shown as horizontal dashed lines and provide the minimum distance possible between the cdfs. The results show that smaller values of δ reproduced the empirical cdf more accurately, unless δ was so small that very few simulations were included in each estimate of the density g (the denominator in Eq. (1)). After some experimentation (see SM Appendix B.1.1) we chose to set δ to be the smallest value for which at least three simulations were included in each estimate of g. These values are illustrated by vertical lines on Fig. 3, and can be seen to come close to achieving the minimum possible squared distance between the empirical cdfs, but with larger ESS.

In our simulations we have evaluated the performance of the method for the distance-based empirical Radon-Nikodym derivative, which is based on using the prevalences within a certain distance given by $\delta/2$ of p_j. We also investigated alternative derivatives, discussed in Section 2.5, and the results are summarised in Table B.1 of the SM. Overall, we observed that the discrepancy-based derivative provides the best possible distance between the two cdfs, but at a cost of a lower ESS in all the scenarios considered. When the proposal was uniform and had simulations in all areas of the posterior distribution then the histogram-based derivative performed better than the distance-based derivative both in terms of accuracy and ESS. However, the situation was

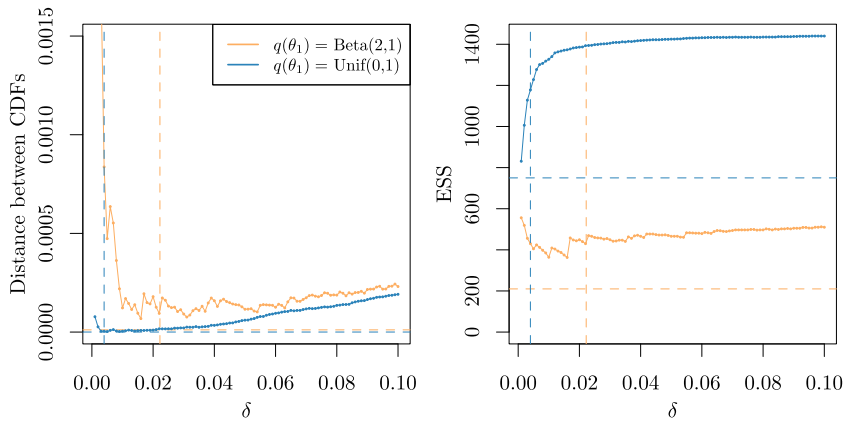


Fig. 3. Distance between the two cumulative distribution functions (cdfs) (left panel) and effective sample size (ESS) (right panel) obtained under different values of δ and choice of proposal distribution for parameter θ_1, for one randomly selected simulated dataset. Orange solid line represents a prevalence proposal distribution equal to the marginal prior, i.e. Beta(2,1), whereas the blue line corresponds to a $U(0,1)$ proposal. In both cases, the pixel prevalences were drawn from a Beta(1,2) distribution. Dashed vertical lines represent the suggested value of δ for each scenario considered. Dashed horizontal lines correspond to the minimum possible distance (left panel) and its associated ESS (right panel).

reversed when there were areas in the proposal with few simulations. In that scenario, the distance-based derivative was found to have lower integrated squared distance and higher ESS compared to the histogram-based derivative. Therefore, we used the distance-based empirical Radon-Nikodym derivative in our applications, since it was more robust to weaknesses in the proposal.

4. Application to lymphatic filariasis data

In this section, we apply the proposed approach for the analysis of real data for lymphatic filariasis (LF) in East Africa. LF is caused by a mosquito-borne macro-parasite, which was historically endemic in many tropical countries, with over a billion people at risk of infection, and millions affected by the disease suffer from disability, stigma and associated social and economic consequences (Ramaiah and Ottesen, 2014). LF is one of the neglected tropical diseases (NTDs) targeted for elimination as a public health problem by 2020 (WHO, 2012), with new guidelines currently being developed for 2030. Global efforts to eliminate LF as a public health problem, through the use of mass drug administration (MDA) of treatments with an excellent safety record, have reduced prevalence to low levels in many settings (Ramaiah and Ottesen, 2014). While many countries have successfully scaled-up their programs, there remain a number of questions on how best to scale up treatment to assist priority countries in optimising interventions to accelerate elimination. Therefore, there is an urgent need to provide detailed estimates of the impact of current and future control programs for donor and policy planning.

For LF, the intervention strategy for most of Africa is to have yearly MDA at 65% coverage for 5 years, followed by an assessment of transmission and, if necessary, further rounds of treatment. In areas where MDA has not yet started, alternative strategies may be required to meet the WHO target, i.e. the prevalence being less than 1% (WHO, 2012) as soon as possible. Enhanced strategies include MDA at high coverage or twice-yearly treatment (Stolk et al., 2018). By bringing together statistical mapping and transmission modelling, we aim to provide high-resolution quantification of the likely impact of control programs and predictions on both future impact and demand for interventions, allowing policy makers to more effectively target available resources.

4.1. The mathematical model of LF transmission dynamics

In this section we describe the mathematical model of lymphatic filariasis transmission, TRANSFIL (Irvine et al., 2015), that is used throughout the paper. TRANSFIL is an individual-based model of LF infection in human populations, with each host having their own adult worm and microfilariae (mf) burden, as well as mosquito bite risk and treatment history. A full description of the model is provided in Irvine et al. (2015) and in Appendix C of the SM, so here we provide only a brief overview and the updated aspects of it.

Each human is assumed to have their own burden of male and female worms denoted by W_i^m and W_i^f, respectively. The times at which human i acquires female and male adult worms are given by two inhomogenous Poisson processes, both with rate:

$$\tfrac{1}{2}\lambda b_i(V/H)\psi_1\psi_2 s_2 h(a),$$

where λ is the number of bites per mosquito, V/H is the ratio of vectors to hosts, ψ_1 is the probability that a third-stage larvae (L3, the infectious stage) leaves the host during a bite, ψ_2 is the probability that the L3 enters the host, s_2 is the proportion of L3 that develop into adult worms within the host and $h(a)$ is the biting rate for a human with age a. Both male and female worms are introduced to a human according to a bite risk b_i drawn from a gamma distribution with mean 1 and shape parameter k. Thus, the degree of parasite aggregation amongst humans can be quantified by this shape parameter. Finally, we assume that each worm has a constant death rate μ.

Microfilariae concentration in the peripheral blood, denoted by M_i, is also modelled for each individual according to the following equation:

$$\frac{\mathrm{d}M_i}{\mathrm{d}t} = \alpha W_i^f \mathbb{1}_{\{W_i^m>0\}} - \gamma M_i,$$

with α being the production rate of mf per worm, γ the constant death rate of mf and the indicator function $\mathbb{1}_{\{W_i^m>0\}}$ is one if there are male worms and zero if not. Larvae development occurs when mf enter the mosquito during a blood meal from an infected host. Different functional forms have been found to describe the relationship between the number of mf ingested and the number that develop within the mosquito. For *Anopheles*, which is the genus of the most dominant vector species in East Africa, this relationship is expressed as:

$$L(m) = \kappa_{s2}\left(1 - e^{-r_2 m/\kappa_{s2}}\right)^2,$$

where m is the concentration of mf per 20 µL taken during a blood meal and r, κ_s denote the saturation values related to the uptake function as detailed in Gambhir and Michael (2008). The equilib-

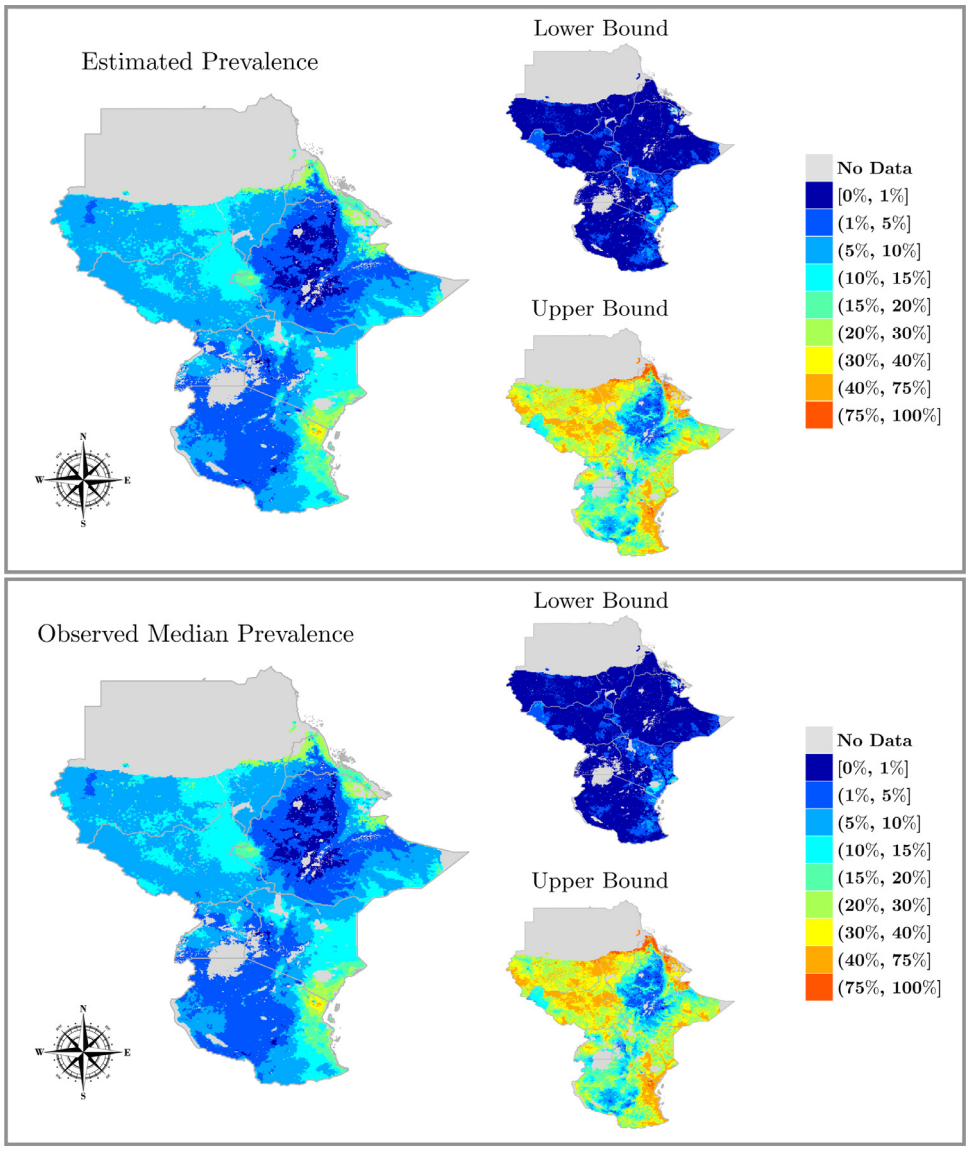


Fig. 4. Accuracy assessment of our method. The true distribution of LF prevalence (lower panel) is compared to the estimated prevalence distribution (upper panel) using our proposed methodology predicted at 5×5 km resolution. Point estimates along with lower (2.5%) and upper (97.5%) percentiles are presented.

rium value for L3 in a mosquito is given by:

$$L^* = \frac{\lambda g \tilde{L}}{\sigma + \lambda \psi_1},$$

where λ is the number of bites per mosquito, g is the proportion of mosquitoes which pick up infection when biting an infected host, σ is the death rate of mosquitoes and $\tilde{L}$ is the average number of larvae per mosquito.

Each human begins life with zero infection and a bite-rate of exposure b_i. The human death rate is denoted by τ and is assumed to be constant throughout an individuals lifetime with a cut-off at age 100. When an individual dies another one is born in order to keep the population size constant.

During an intervention campaign, the impact of MDA is simulated for an individual by reducing their mf concentration and their male and female worm burden according to the estimated drug efficacies from the literature (Ismail et al., 1998; Michael et al., 2004). In addition, there is a period after MDA during which the production of mf for that individual is diminished. Furthermore, the individuals' compliance after multiple rounds of treatment is modelled based on the paper by Griffin et al. (2010), where the authors model the probability of an individual making the same decision as in the previous round of treatment.

Finally, we extended the model to include a very low rate of importation of infection from outside the population being modelled, otherwise the equilibrium distribution (steady state), that is used as the starting point of the simulations, is just the degenerate distribution where no-one is infected. The interventions reduce the prevalence over time, and so we reduce the importation rate after intervention in proportion to the reduction in prevalence seen in pilot simulations. Lists of the model parameters are provided in Tables C.3 and C.4 of the SM.

4.2. Implementation details

The starting point for our analysis was the spatial map (pixel scale 5×5 km) providing the predicted distribution of the LF prevalence based on mf data, generated through a Bayesian geostatistical modelling approach described by Moraga et al. (2015). The top panel of Fig. 4 shows the median of the posterior distribution of the prevalence obtained at each pixel, along with estimates of lower (2.5%) and upper (97.5%) percentiles. In particular, we anal-

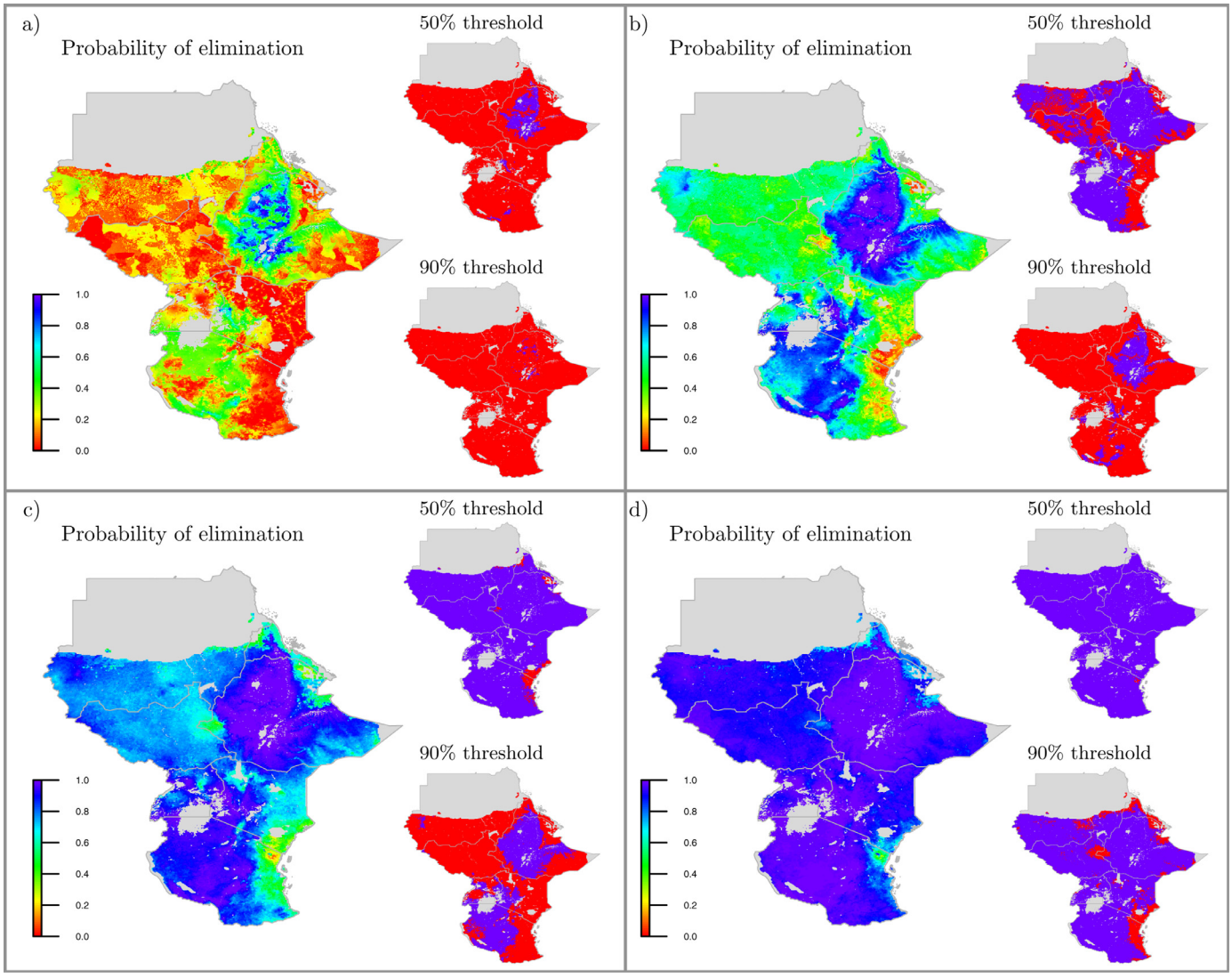


Fig. 5. Probability of less than 1% prevalence after 5 years under: a) no intervention; annual MDA with coverage of b) 65%; c) 80% and d) biannual MDA at 65% coverage predicted at 5×5 km resolution, for Ethiopia, Sudan, South Sudan, Eritrea, Kenya, Tanzania and Uganda. Right of panels: Pixels that achieve elimination (blue) or do not achieve elimination (red) using different probability thresholds.

ysed the following seven African countries: Ethiopia, Sudan, South Sudan, Eritrea, Kenya, Tanzania and Uganda. We linked each pixel to the corresponding population estimates obtained from the Gridded Population of the World (Worldpop, 2010), which provides the estimated number of people in each pixel. We avoided handling pixels with either very small or very large populations as the transmission model was not thought to be appropriate in these environments (Irvine et al., 2015; Smith et al., 2017). More specifically, for small populations we pooled pixels with less than 300 people together, ensuring that the merged pixels belong to the same country and that the groups contain as few pixels as possible. We excluded pixels with population estimates over 10 000 from the analysis; resulting in 1.7% of the pixels being excluded.

The stochastic model of LF transmission TRANSFIL was used to investigate and compare the impact of different control strategies. In order to simulate the entire range of observed baseline mf prevalence levels, with values up to 95%, we assumed that four parameters of the mathematical model were spatially varying: the population size, the vector to host ratio, the aggregation parameter of individual exposure to mosquitoes and the importation rate, using prior distributions informed from data, pilot simulations and previous analyses. We assumed that the parameter prior was the same in each spatial location (discussed in more details in Appendix C.2 of the SM) except for the population size, which was assumed to be a log normal distribution, ie $\log(n) \sim \mathcal{N}\left(\log(N_i), \sigma^2\right)$, where N_i is the reported population of pixel i (adjusted population from Worldpop, 2010) and σ is the sample standard deviation of the log population estimates available in WorldPop.

The proposal density of the population sizes, $q(n)$, was designed so that each simulation contributed an equal amount to the effective sample size of a set of pixels with populations $\{260, 261, \ldots, 10\,000\}$. This was achieved by calculating the effective sample size of an initial proposal, namely, $q_0(n) \propto 1$. The remaining population sizes (10 001–11 550) were taken to decrease linearly from $q(10\,000)$ to zero. Since the uncertainty in the log-normal prior is much greater for large populations, fewer simulations are needed in these regions. The final proposal was obtained from 10 iterations of $q_i(n) \propto q_{i-1}(n)/\mathrm{ESS}_{i-1}(n)$, where $\mathrm{ESS}_i(n)$ is the effective sample size of a simulation with n individuals from the proposal q_i (see Fig. C.7(a) of the SM).

A significant merit of our approach is that it can be easily applied in parallel which can be utilised to speed up implementation,

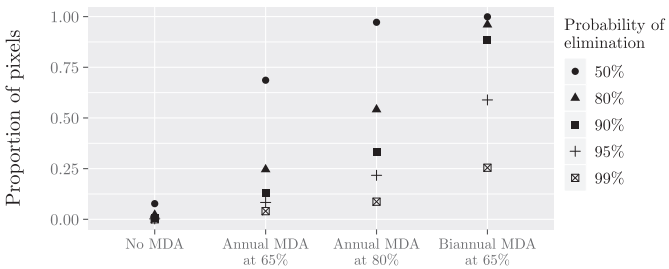


Fig. 6. Proportion of pixels with prevalence less than 1% using different probability thresholds after 5 years under four intervention strategies.

especially in applications involving a large number of pixels. This is because we are treating each pixel independently and therefore the computation of the weights can be undertaken in parallel. In our application, the computation time of this step was approximately 8 hours using a 112 core computer cluster (around 30 s on a single core for each pixel).

4.3. Results

In this section, the Bayesian approach presented in Section 2 was applied to the LF data. Firstly, we assessed the accuracy of the method, defined as the ability of the transmission model weighted simulations to reproduce the pre-control (baseline) geostatistical map, by comparing the observed and the estimated distribution of the baseline (equilibrium) mf prevalences at each pixel. Fig. 4 illustrates the median map (with 2.5 and 97.5 percentiles), along with the corresponding maps of the observed data. Overall, the results show that the maps are almost identical, indicating that the method is able to reproduce the distribution of the observed baseline prevalence in each pixel. In addition, in the left panel of Fig. D.8 of the SM we compared the estimated number of people per pixel with the observed value, which were in close agreement indicating that the proposed method accurately reproduced the number of people in each pixel. In the right panel of Fig. D.8 of the SM, we examined the ESS per pixel, which represents the effective number of simulations per pixel and is a measure of how well the method performs. We observed that the pixels with high prevalence (which may require a change of intervention strategy) have high ESS.

Secondly, the methodology was applied to evaluate the impact of different intervention programs for LF in East Africa. In particular, four treatment scenarios were simulated: no interventions; the standard 65% coverage annual MDA (aMDA); 80% coverage aMDA; or biannual MDA (bMDA) at 65% coverage, in order to investigate how these affect the probability of elimination after 5 years (Fig. 5). Adopting a prevalence of less than 1% as the threshold set by WHO as a global target for determining LF transmission elimination, our analysis predicted that the recommended strategy of 5 rounds of aMDA at 65% is not enough for eliminating the disease in all pixels, with probability of elimination above 90% only for 13% of the pixels (see also Fig. 6). Moreover, when more intensive treatments were implemented, i.e. more frequent MDA or higher coverage, the probability of elimination significantly increased compared to aMDA programme at 65%. In particular, bMDA at 65% coverage was the most effective of all strategies considered and was able to achieve elimination in 88% of the pixels, with at least 90% probability. However, the proportion of pixels which achieved elimination after 5 years reduced to 59% and 25% when the probability threshold was increased to 95% and 99%, respectively, illustrating that the policy is sensitive to uncertainty.

Finally, predictions of mf prevalence for the first and fifth year of intervention were summarized by calculating the estimated prevalence at each pixel, together with the 2.5th and 97.5th percentiles in Figs. D.9 and D.10 of the SM, for each of the four scenarios. Very similar observations were made on the predictions of mf prevalence for the first 5 years of intervention.

5. Discussion

This study highlights the value of integrating geostatistical prevalence maps and transmission models for providing predictions on the impact of interventions aiming to eliminate transmission at a local scale. The main contribution is the development of new statistical tools through which existing research in mapping and predictive modelling are combined in a computationally efficient and flexible way which correctly accounts for uncertainty in these different techniques. Although we focus and apply our methodology on LF transmission, it can be applied to other infectious diseases.

We have shown that the current strategy of 5 annual rounds of MDA at 65% coverage will not be sufficient to eliminate the disease in most areas. We also found that a change in the current MDA strategy, such as increasing the coverage and frequency of MDA, will be required if LF elimination is to be accelerated in East Africa. This suggests that it may be necessary to employ different enhanced intervention plans at a fine scale, according to the characteristics of each area, in order to achieve the WHO elimination targets.

However, for the results presented here we assumed that no interventions have been applied in East Africa prior to the prevalence survey. While this assumption is correct for most areas, MDA programs began to be implemented in a few districts of Africa since 2000 and in many more districts thereafter. Therefore, one of the next steps will be to account for previous MDA programmes as spatio-temporal covariate information in the transmission model. Apart from MDA, insecticide treated bednets have been used in some countries (data can be extracted from the Malaria Atlas Project), the use of which has been shown to be an effective additional measure for control of the disease (Bockarie et al., 2009). Integrating geostatistical maps with transmission models with these additional covariates is more complicated as the simulations must include the appropriate historical interventions.

An additional challenge is the gap between reported and true coverage with an MDA. Where there are parasitological data against which to test the expected and achieved impact of reported coverages, they have been shown to be unreliable (Budge et al., 2016). This will pose a particular challenge to interpreting historic coverage and a challenge in communicating future projections. This work represents our initial framework and future research will be required to extend the methodology to capture these more complex settings.

A limitation of our statistical approach is that it doesn't capture the spatial correlations in the predictions, since each pixel is weighted independently to produce a marginal posterior for each pixel. This approach means that we lose the spatial autocorrelation that was captured in the original geostatistical model and, furthermore, that there is no way for nearby pixels to interact during the simulations, for example, to account for movements of humans or vectors. A more sophisticated approach would be to use the spatial autocorrelations from the geostatistical model, alongside any available movement or connectivity data, to define a transmission kernel that describes spatial spread. This kernel could be used within a single meta-population model describing the transmission dynamics across the whole map. At present, such an approach would be computationally infeasible at the country scale, but may become possible in future through improvements in methodology and advances in high-performance computing.

Acknowledgments

The authors gratefully acknowledge funding of the NTD Modeling Consortium by the Bill and Melinda Gates Foundation [OPP1152057, OPP1053230, OPP1156227, OPP1186851]. The views, opinions, assumptions or any other information set out in this article should not be attributed to the Bill and Melinda Gates Foundation or any person connected with the Bill and Melinda Gates Foundation. The authors thank Rachel L. Pullan and Jorge Cano for sharing the geostatistical maps of LF prevalence and Michael A. Irvine and Paul Brown for improving the model code. We also thank Dr. Nick Golding for providing helpful comments on our manuscript.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sste.2020.100391. In the supplementary material, we provide further details on the methodology and the transmission model as well as additional plots and results of the real data analysis.

References

Alley, E.S., Plaisier, A.P., Boatin, B.A., Dadzie, K.Y., Remme, J., Zerbo, G., Samba, E.M., 1994. The impact of five years of annual ivermectin treatment on skin microfilarial loads in the onchocerciasis focus of Asubende, Ghana. Trans. R. Soc. Trop. Med. Hyg. 88 (5), 581–584.
Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. Genetics 162 (4), 2025–2035.
Bhatt, S., Weiss, D.J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K.E., Moyes, C.L., Henry, A., Eckhoff, P.A., et al., 2015. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature 526 (7572), 207–211.
Billingsley, P., 1995. Probability and Measure. Wiley, New York.
Bockarie, M.J., Pedersen, E.M., White, G.B., Michael, E., 2009. Role of vector control in the global program to eliminate lymphatic filariasis. Annu. Rev. Entomol. 54, 469–487.
Budge, P.J., Sognikin, E., Akosa, A., Mathieu, E.M., Deming, M., 2016. Accuracy of coverage survey recall following an integrated mass drug administration for lymphatic filariasis, schistosomiasis, and soil-transmitted helminthiasis. PLoS Negl. Trop. Dis. 10 (1), e0004358.
Cornuet, J.-.M., Marin, J.-.M., Mira, A., Robert, C.P., 2012. Adaptive multiple importance sampling. Scand. J. Stat. 39 (4), 798–812.
Deardon, R., Brooks, S.P., Grenfell, B.T., Keeling, M.J., Tildesley, M.J., Savill, N.J., Shaw, D.J., Woolhouse, M.E.J., 2010. Inference for individual-level models of infectious diseases in large populations. Stat. Sin. 20 (1), 239.
Diggle, P., Ribeiro, P.J., 2007. Model-based Geostatistics. Springer, New York.
Ferguson, N.M., Cummings, D.A.T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., Burke, D.S., 2005. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 437 (7056), 209–214.
Gambhir, M., Michael, E., 2008. Complex ecological dynamics and eradicability of the vector borne macroparasitic disease, lymphatic filariasis. PLoS One 3 (8), e2874.
Gibson, G.J., 1997. Markov chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology. J. R. Stat. Soc. 46 (2), 215–233.
Giorgi, E., Diggle, P.J., Snow, R.W., Noor, A.M., 2018. Geostatistical methods for disease mapping and visualisation using data from spatiotemporally referenced prevalence surveys. Int. Stat. Rev. 86 (3), 571–597.
Goldie, C.M., Maller, R.A., 1999. Generalized densities of order statistics. Stat. Neerl. 53 (2), 222–246.
Griffin, J.T., Hollingsworth, T.D., Okell, L.C., Churcher, T.S., White, M., Hinsley, W., Bousema, T., Drakeley, C.J., Ferguson, N.M., Basáñez, M.-.G., et al., 2010. Reducing plasmodium falciparum malaria transmission in Africa: a model-based evaluation of intervention strategies. PLoS Med. 7 (8), e1000324.
Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H., Elyazar, I.R.F., Brooker, S., et al., 2009. A world malaria map: plasmodium falciparum endemicity in 2007. PLoS Med. 6 (3), e1000048.
Heesterbeek, H., Anderson, R.M., Andreasen, V., Bansal, S., De Angelis, D., Dye, C., Eames, K.T.D., Edmunds, W.J., Frost, S.D.W., Funk, S., et al., 2015. Modeling infectious disease dynamics in the complex landscape of global health. Science 347 (6227), aaa4339.
Hollingsworth, T.D., 2018. Counting down the 2020 goals for 9 neglected tropical diseases: what have we learned from quantitative analysis and transmission modeling? Clin. Infect. Dis. 66 (suppl_4), S237–S244.
Irvine, M.A., Reimer, L.J., Njenga, S.M., Gunawardena, S., Kelly-Hope, L., Bockarie, M., Hollingsworth, T.D., 2015. Modelling strategies to break transmission of lymphatic filariasis - aggregation, adherence and vector competence greatly alter elimination. Parasites Vectors 8 (1), 547.
Ismail, M.M., Jayakody, R.L., Weil, G.J., Nirmalan, N., Jayasinghe, K.S.A., Abeyewickrema, W., Sheriff, M.H.R., Rajaratnam, H.N., Amarasekera, N.D.D.M., De Silva, D.C.L., et al., 1998. Efficacy of single dose combinations of albendazole, ivermectin and diethylcarbamazine for the treatment of bancroftian filariasis. Trans. R. Soc. Trop. Med. Hyg. 92 (1), 94–97.
Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J., Grenfell, B.T., 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science 294 (5543), 813–817.
Michael, E., Malecela-Lazaro, M.N., Simonsen, P.E., Pedersen, E.M., Barker, G., Kumar, A., Kazura, J.W., 2004. Mathematical modelling and the control of lymphatic filariasis. Lancet Infect. Dis. 4 (4), 223–234.
Moraga, P., Cano, J., Baggaley, R.F., Gyapong, J.O., Njenga, S.M., Nikolay, B., Davies, E., Rebollo, M.P., Pullan, R.L., Bockarie, M.J., et al., 2015. Modelling the distribution and transmission intensity of lymphatic filariasis in sub-Saharan Africa prior to scaling up interventions: integrated use of geostatistical and mathematical modelling. Parasites Vectors 8 (1), 560.
O'Hanlon, S.J., Slater, H.C., Cheke, R.A., Boatin, B.A., Coffeng, L.E., Pion, S.D.S., Boussinesq, M., Zouré, H.G.M., Stolk, W.A., Basáñez, M.-.G., 2016. Model-based geostatistical mapping of the prevalence of Onchocerca volvulus in West Africa. PLoS Negl. Trop. Dis. 10 (1), e0004328.
Plaisier, A.P., Van Oortmarssen, G.J., Remme, J., Alley, E.S., Habbema, J.D., 1991. The risk and dynamics of onchocerciasis recrudescence after cessation of vector control. Bull. World Health Organ. 69 (2), 169–178.
Pullan, R.L., Sturrock, H.J.W., Magalhaes, R.J.S., Clements, A.C.A., Brooker, S.J., 2012. Spatial parasite ecology and epidemiology: a review of methods and applications. Parasitology 139 (14), 1870–1887.
Ramaiah, K.D., Ottesen, E.A., 2014. Progress and impact of 13 years of the global programme to eliminate lymphatic filariasis on reducing the burden of filarial disease. PLoS Negl. Trop. Dis. 8 (11), e3319.
Retkute, R., Touloupou, P., Basanez, M.-G., Hollingsworth, T. D., Spencer, S. E. F., 2020. Integrating geostatistical maps and transmission models using adaptive multiple importance sampling. medRxiv, 10.1101/2020.08.03.20146241. 10.1101/2020.08.03.20146241
Slater, H., Michael, E., 2013. Mapping, Bayesian geostatistical analysis and spatial prediction of lymphatic filariasis prevalence in Africa. PLoS One 8 (8), e71574.
Smith, M.E., Singh, B.K., Irvine, M.A., Stolk, W.A., Subramanian, S., Hollingsworth, T.D., Michael, E., 2017. Predicting lymphatic filariasis transmission and elimination dynamics using a multi-model ensemble framework. Epidemics 18, 16–28.
Stensgaard, A.-.S., Vounatsou, P., Onapa, A.W., Simonsen, P.E., Pedersen, E.M., Rahbek, C., Kristensen, T.K., 2011. Bayesian geostatistical modelling of malaria and lymphatic filariasis infections in Uganda: predictors of risk and geographical patterns of co-endemicity. Malar. J. 10 (1), 298.
Stolk, W.A., Prada, J.M., Smith, M.E., Kontoroupis, P., De Vos, A.S., Touloupou, P., Irvine, M.A., Brown, P., Subramanian, S., Kloek, M., et al., 2018. Are alternative strategies required to accelerate the global elimination of lymphatic filariasis? Insights from mathematical models. Clin. Infect. Dis. 66 (Supplement_4), S260–S266.
Sturrock, H.J.W., Gething, P.W., Clements, A.C.A., Brooker, S., 2010. Optimal survey designs for targeting chemotherapy against soil-transmitted helminths: effect of spatial heterogeneity and cost-efficiency of sampling. Am. J. Trop. Med. Hyg. 82 (6), 1079–1087.
Tatem, A.J., Smith, D.L., Gething, P.W., Kabaria, C.W., Snow, R.W., Hay, S.I., 2010. Ranking of elimination feasibility between malaria-endemic countries. Lancet 376 (9752), 1579–1591.
Tekle, A.H., Zouré, H.G.M., Noma, M., Boussinesq, M., Coffeng, L.E., Stolk, W.A., Remme, J.H.F., 2016. Progress towards onchocerciasis elimination in the participating countries of the African Programme for Onchocerciasis Control: epidemiological evaluation results. Infect. Dis. Poverty 5 (1), 66.
Tildesley, M.J., Bessell, P.R., Keeling, M.J., Woolhouse, M.E.J., 2009. The role of pre-emptive culling in the control of foot-and-mouth disease. Proc. R. Soc. B 276 (1671), 3239–3248.
Wakefield, J., Lyons, H., 2010. Spatial aggregation and the ecological fallacy. In: Gelfand, A.E., Diggle, P., Guttorp, P., Fuentes, M. (Eds.), Handbook of Spatial Statistics. CRC Press, pp. 541–558.
WHO, 2012. Accelerating work to overcome the global impact of neglected tropical diseases: a roadmap for implementationGeneva (WHO/HTM/NTD/2012).
Wilkinson, R.D., 2013. Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. Stat. Appl. Genet. Mol. Biol. 12 (2), 129–141.
Worldpop, 2010. The AfriPop demography project (www.worldpop.org.uk).