medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Full Title: Geospatial modeling of pre-intervention prevalence of *Onchocerca volvulus* infection in Ethiopia as an aid to onchocerciasis elimination

Short title: Geospatial modeling of onchocerciasis infection in Ethiopia

Authors: Himal Shrestha (1), Karen McCulloch (1, 2), Shannon M Hedtke (1), Warwick N Grant (1)

Author affiliations:

(1) Department of Environment and Genetics, School of Agriculture, Biomedicine and Environment, La Trobe University, Bundoora, Australia
(2) WHO Collaborating Centre for Viral Hepatitis, Victorian Infectious Diseases Reference Laboratory, Royal Melbourne Hospital, and Department of Infectious Diseases, University of Melbourne, at the Peter Doherty Institute for Infection and Immunity, Australia

Corresponding author: Shannon Hedtke, S.Hedtke@latrobe.edu.au

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background

Onchocerciasis is a neglected tropical and filarial disease transmitted by the bites of blackflies, causing blindness and severe skin lesions. The change in focus for onchocerciasis management from control to elimination requires thorough mapping of pre-control endemicity to identify areas requiring interventions and to monitor progress. *Onchocerca volvulus* infection prevalence in sub-Saharan Africa is spatially continuous and heterogeneous, and highly endemic areas may contribute to transmission in areas of low endemicity or vice-versa. Ethiopia is one such onchocerciasis-endemic country with heterogeneous *O. volvulus* infection prevalence, and many districts are still unmapped despite their potential for *O. volvulus* infection transmission.

Methodology/Principle findings

A Bayesian geostatistical model was fitted for retrospective pre-intervention nodule prevalence data collected from 916 unique sites and 35,077 people across Ethiopia. We used multiple environmental, socio-demographic, and climate variables to estimate the pre-intervention prevalence of *O. volvulus* infection across Ethiopia and to explore their relationship with prevalence. Prevalence was high in southern and northwestern Ethiopia and low in Ethiopia's central and eastern parts. Distance to the nearest river (-0.015, 95% BCI: -0.025 – -0.005), precipitation seasonality (-0.017, 95% BCI: -0.032 – -0.001), and flow accumulation (-0.042, 95% BCI: -0.07 – -0.019) were negatively associated with *O. volvulus* infection prevalence, while soil moisture (0.0216, 95% BCI: 0.014 – 0.03) was positively associated.

Conclusions/Significance

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Infection distribution was correlated with habitat suitability for vector breeding and associated biting behavior. The modeled pre-intervention prevalence can be used as a guide for determining priority for intervention in regions of Ethiopia that are currently unmapped, most of which have comparatively low infection prevalence.

Author's summary

Areas with unknown onchocerciasis endemicity may pose a threat to the goal of eliminating transmission: they may re-introduce onchocerciasis to areas where interventions have been successful. Additionally, because the vectors (and thus *Onchocerca volvulus* transmission) have specific ecological requirements for growth and development, changes in these ecological factors due to human activities (deforestation, modification of river flows by dam construction, climate change) might change patterns of parasite transmission and endemicity. To estimate the impact of these environmental changes, we must first identify ecological factors that determine transmission and prevalence. We have employed Bayesian geostatistical modeling to create a nationwide *O. volvulus* infection prevalence map for Ethiopia based on pre-intervention nodule prevalence and have explored the effect of environmental variables on *O. volvulus* infection prevalence. We have also identified areas that need additional data to increase the prediction accuracy of the map. We found that hydrological variables such as distance to the nearest river, precipitation seasonality, soil moisture, and flow accumulation are associated significantly with *O. volvulus* infection prevalence. We show that the pre-intervention prevalence can be estimated based on the ecological data and that predicted prevalence can be used as a guide to prioritize pre-intervention mapping.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Mapping infection prevalence is fundamental for control and elimination because it is used to estimate the disease burden and to design and monitor the impacts of interventions. Often the prevalence data are present in the form of point data at different locations and time points, and are aggregated at different administrative levels (1). However, disease risk is a spatially continuous phenomenon that extends across and beyond administrative borders (2). In addition, mapping strategies change depending on the intended endpoint of the intervention (3): when elimination of transmission is the goal, the spatial heterogeneity in disease prevalence has to be quantified accurately so that appropriate interventions can be implemented and, where possible, implementation and monitoring can be informed by the spatial distribution of infection rather than simply along local administrative organizational boundaries. When resources or accessibility to an endemic region are limited, as is the case for many neglected tropical diseases, such thorough data collection may not be possible and methods to extrapolate likely prevalence would be useful.

Using geostatistical modeling techniques, point prevalence data can be transformed into a continuous spatial prevalence map of varying endemicity (2, 4), rather than reporting binary categorization of areas as endemic or non-endemic (5). These continuous maps can extrapolate the prevalence measures to previously unmapped regions based on the spatial autocorrelation between the prevalence measures and the influence of known ecological and socio-demographic factors. In addition, geostatistical models provide unbiased quantification of the uncertainty associated with the prevalence estimates.

Onchocerciasis is a neglected tropical disease caused by infection with a filarial nematode, *Onchocerca volvulus*, that is transmitted by the bites of blackflies (*Simulium* spp.). The vectors

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

have a specific ecological niche: they breed around fast-flowing rivers, requiring high aeration and oxygen content for larval development (6). The flies show diurnal activity and bite humans living in communities near these rivers (7-9). If the biting blackfly carries the infective stage of the parasite (the 3rd stage larvae, or iL3), the larva leaves the blackfly and enters the human host. Inside the human body, the larva develops into an adult worm and forms a nodule, generally localized subcutaneously. People living with onchocerciasis show a range of chronic clinical manifestations, including onchodermatitis, severe itching, rashes, and visual impairment that may culminate in blindness (10). More recently, it has also been linked with epilepsy and nodding syndrome in children (11, 12).

Onchocerciasis is currently targeted for elimination via community-directed mass drug administration with ivermectin (MDAi), either annually, semi-annually, or, in some areas, up to four times a year (13). *Onchocerca volvulus* infection prevalence is measured using counts of microfilariae (mf) in a small of skin biopsy (skin snipping), physical examination for the presence of nodules (nodule palpitation), or antibody tests that detect the presence of antibodies against the parasite Ov16 antigen (14). Rapid Epidemiological Mapping of Onchocerciasis (REMO) uses nodule palpation in combination with geographic information system mapping, and was used by the African Programme for Onchocerciasis Control (APOC) to map prevalence in twenty countries from 1996 to 2012 (15, 16). REMO revealed that the prevalence of *O. volvulus* infection was patchy and heterogenous across Africa (17) and identified areas for ivermectin intervention (3, 15) using a threshold for treatment set at a nodule prevalence of 20%. Onchocerciasis-endemic communities were divided into hypoendemic (nodule prevalence: < 20%), meso-endemic (nodule prevalence: 20–45%), and hyperendemic (nodule prevalence: >

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

45%) (17, 18) based on nodule prevalence. However, there are still many areas that are unmapped and in which the infection prevalence is not known (19).

In onchocerciasis-endemic Ethiopia, mapping of prevalence has been focused on the western districts based on the high incidence of onchocerciasis and because environmental factors favor blackfly breeding in these regions (20). In contrast, eastern Ethiopia has been assumed to be free of *O. volvulus* infection, which has generally proven true (21). However, a recent continent-level mapping (19) found that most of the implementation units that were predicted to be suitable for onchocerciasis in Ethiopia were not mapped, posing a risk to elimination goals. In addition, there is high spatial variability of onchocerciasis endemicity in Ethiopia, ranging from 0% in some areas to as high as 84% in some areas of southwest Ethiopia (21, 22).

MDAi started in some Ethiopian hyperendemic foci in 2002 (20) and, to our knowledge, there has not been coordinated vector control in Ethiopia. The shift to onchocerciasis elimination officially began in 2013 with a goal to eliminate transmission by 2020 (20, 22): the program moved from annual to biannual treatment strategy in all the known endemic areas and scaled up treatment to other additional endemic areas which were not treated previously (22). Cross-border coordination of MDAi between transmission foci in northwestern Ethiopia and bordering Sudan is ongoing (13). In some cases, transmission decline without intervention has also been reported (23) but onchocerciasis persists in some areas despite MDAi for a variety of reasons, including challenges with treatment compliance (24-26), civil unrest (27, 28), and lately the COVID-19 pandemic (29). In addition, there has been variation in the history and the frequency of MDAi. Most of the hyper- and mesoendemic districts have been treated over two decades, while in hypoendemic districts, MDAi started around 2014 following the policy shift from control to elimination (22).

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

There is no national-level baseline endemicity map of *O. volvulus* infection prevalence for Ethiopia, which has created difficulty in quantifying the effect of MDAi on a national scale. Baseline/pre-control endemicity is an important indicator of morbidity and a predictor for the time required for elimination (18, 30-32). In addition, the prevalence measures before intervention provide an unbiased relationship between the infection prevalence and environmental variables. Prevalence and onchocerciasis suitability mapping for Ethiopia in previous studies (17, 19) have been done as part of continental-scale research, although Zouré et al.(17) did not consider environmental factors, and Cromwell et al. (19) used presence-absence data which do not capture the magnitude of the prevalence. Although these studies helped to place *O. volvulus* infection prevalence or risk in a broader ecological and epidemiological context, we have focused on a spatial scale which offers us greater flexibility to explore ecological patterns unique to Ethiopia by incorporating both the magnitude of prevalence and associated ecological variables (33). We develop a geostatistical model for the distribution of pre-intervention nodule prevalence of *O. volvulus* infection prevalence in Ethiopia using this approach that considers spatial variation in environmental and socio-demographic variables. Furthermore, we identify the most important environmental and socio-demographic variables contributing to *O. volvulus* infection prevalence, and present estimates of uncertainty in the predicted prevalence that can be used to target areas for further mapping efforts.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

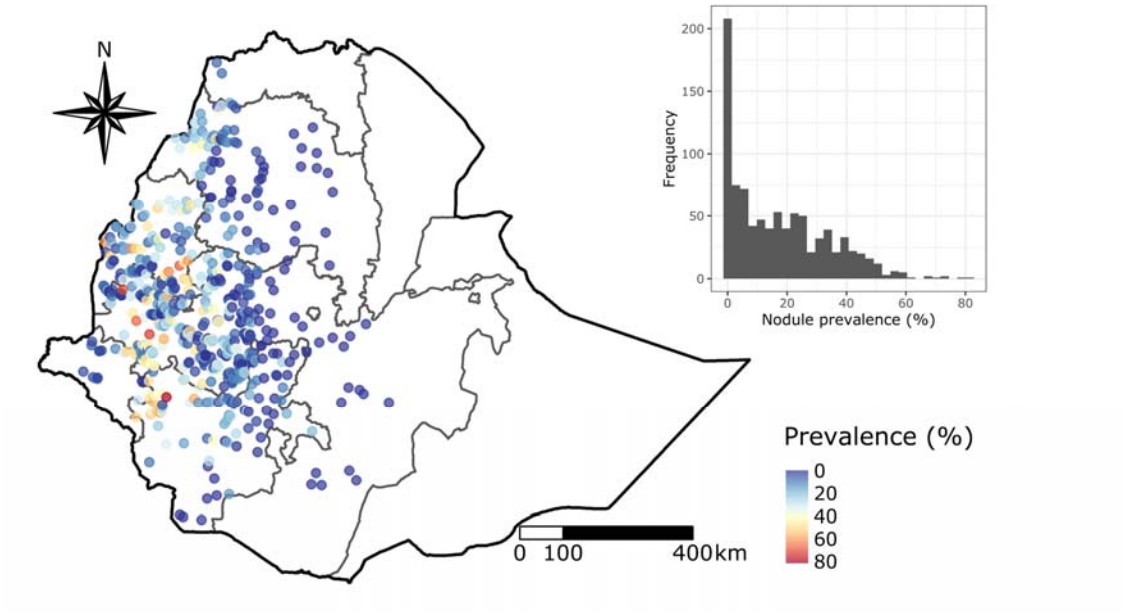
Methods

Prevalence data

Onchocerca volvulus infection prevalence data with site-specific coordinates for Ethiopia were obtained from the publicly available Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN) database (34). Nodule prevalence data were collected as part of REMO mapping before the initiation of MDAi, from the year 2001 to 2012, examining the presence of palpable onchocercal nodules in 30 to 50 adults randomly selected from each surveyed village (15, 20). The protocols for the REMO assessment are available in the published guidelines (35). There were 927 geopositioned coordinates for nodule prevalence in 36,010 people. Any observations from the same geographic coordinates at different times were aggregated by adding the number of cases observed and the number of total tests done before calculating the prevalence.

Although the database contains both nodule prevalence data and skin mf prevalence data, in this analysis, only nodule prevalence data were considered for the geospatial analysis because of the limited number of skin mf sampled (n = 126) before MDAi, which limits its utility for identifying associations between prevalence and environmental variables. That said, skin mf data were used to assess the correlation between skin mf as a measure of *O. volvulus* infection and the nodule prevalence measure, which revealed two outlier observations with very high nodule prevalence but with low skin mf prevalence. These were excluded from the dataset because the low skin mf prevalence could not be attributed to a reasonable cause, such as MDAi, as the data were collected before ivermectin distribution. Thus, the final dataset contained nodule prevalence data from 916 unique sites and 35,077 people (Fig 1).

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .



165

166 **Fig 1. Sites and the nodule prevalence measured during Rapid Epidemiological Mapping of**
167 **Onchocerciasis (REMO) in Ethiopia**. The grey boundary on the map represents administrative
168 regions. The inset figure shows the histogram of the prevalence.

169 **Environmental, climate, and socio-demographic variables**

170 Variables relevant to *O. volvulus* infection prevalence and *Simulium* ecology based on published
171 literature were assembled from different sources and were exported as a raster layer at a
172 resolution of 1 km using Google Earth Engine (6, 36-39) (S1 Table). Raster layers with higher
173 resolution were downsampled using a mean aggregation method, whereas the raster layers with
174 lower resolution were resampled to align with 1 km resolution (40) to prepare a raster stack of
175 uniform resolution. Raster data were processed using the *raster* package in R version 4.1.0 (41-
176 43). Downloaded raster variables were reprojected to a standard projection, World Geodetic
177 System 1984 (WGS84). The raster covariates were cropped to the boundary of Ethiopia (S1 Fig),
178 and a raster stack of covariates was prepared. The measurement of different covariates at each
179 sample site was extracted from the raster stack.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Variable selection

Thirty-two variables were grouped into six major categories: elevation, temperature, precipitation, socio-demographic, hydrological, and vegetation (S1 Table), and the initial selection of covariates was conducted separately for each category. During the initial rounds of variable selection, multi-collinearity was assessed among the variables by calculating the Spearman's rank correlation matrix and a variable inflation factor (VIF) for the linear model, including the variables, using the *GGally* and *car* packages in R (44, 45). Next, any variables with an absolute correlation coefficient less than 0.8 with other variables within the group were selected (46). For the set of covariates with a correlation coefficient greater than 0.8 and a VIF greater than 10, only one of the covariates was selected (33, 47). The VIF measures how easily a given predictor can be predicted from a linear regression based on other predictors. The predictor with the lowest VIF score was selected among the set of correlated covariates. The final covariates yielded a correlation matrix of less than 0.8 (S2 Fig) and a VIF factor of less than 10 (S2 Table). Based on this initial round of analysis, a set of 16 covariates were selected.

Model fit was assessed based on the Deviance Information Criterion (DIC) and Widely Applicable Information Criterion (WAIC) scores (48). We ran a univariate regression model and calculated the DIC and WAIC scores for the respective univariate models. A covariate yielding the least DIC and WAIC scores from each category was selected. Combinations of other variables were further explored if their inclusion further optimized the model fit scores. Eight covariates from the pool of 16 possible were selected for downstream geostatistical analysis (S3 Table).

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Geo-statistical modelling framework

A Bayesian geostatistical model was implemented using the Integrated Nested Laplace Approximation (INLA) approach, which has been reported to be computationally efficient for posterior distribution calculation and has been employed in recent large-scale geostatistical models (33, 46, 49, 50). Geostatistical approaches assume a positive spatial correlation between observations; i.e., the observations nearer to each other are more related than the farther ones. Information from neighboring pixels can then be utilized to allow smoothing of extreme values due to small sample sizes and give reliable and robust estimates from sparse data (33, 51). Further, the hierarchical structure of the model permits the estimation of covariate effects, spatial covariance structure, and the prediction of missing data (33). These models incorporate both fixed and random effects. The fixed effects determine the influence of covariates on *O. volvulus* infection prevalence, while the random effects account for the spatial variation that determines anomalous regions of high and low prevalence (46). This model can thus identify the relationship between infection prevalence data and several predictors and quantify spatial dependence via the covariance matrix of a Gaussian process facilitated by adding random effects to the observed locations (49).

Model fitting

Conditional on the true prevalence $P(x_i)$ at location $x_i = 1,2,3 \ldots . n$, the number of cases (Y_i) observed out of the total number of people tested (N_i) were assumed to follow a binomial distribution.

$$Y_i|P(\mathbf{x}_i) \sim Binomial\big(N_i, P(\mathbf{x}_i)\big)$$

The log odds of prevalence is modeled as

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

$$logit\big(P(\mathbf{x}_i)\big) \sim \beta_0 + X_i^T\beta + S(\mathbf{x}_i)$$

where β_0 is the intercept, $X_i^T\beta$ are the vectors of covariates and their coefficients. $S(\mathbf{x}_i)$ is a spatial random effect with zero-mean Gaussian process following the Matérn covariance function which is defined by the equation:

$$\mathrm{Cov}\left(S(\mathbf{x}_i), S(\mathbf{x}_j)\right) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}\left(\kappa||\mathbf{x}_i - \mathbf{x}_j||\right)^{\nu} K_{\nu}\left(\kappa||\mathbf{x}_i - \mathbf{x}_j||\right)$$

Here, K_ν is the modified Bessel function of the second kind and order $\nu > 0$, ν is the smoothness parameter, and σ^2 is the marginal variance (46). $\kappa > 0$ is the scaling parameter related to the practical range ρ, the distance at which the correlation between two points is approximately zero. However, if $\rho = 8\nu\kappa$, at this range, the spatial correlation is close to 0.1 (9, 46). Default priors were used for the intercept parameter, effect parameters for the covariates, and the hyperparameters in the model as defined in Moraga, p. 35–37 (2).

Accounting for excess zero prevalence

The binomial distribution is governed by only a single parameter which does not address overdispersion. To account for the excess zero prevalence in the data (Fig 1), zero-inflated binomial models (ZIB) Type 0 and Type 1 were also considered. There are structural zeros (prevalence reported to be zero based on reality) and sample zeros (prevalence reported to be zero based on chance) in any probability distribution (48, 52). Type 0 model considers only the structural zeros, while Type 1 considers both the structural and sample zeros. With ZIB Type 0 model, the probability density function for the observed cases is

$$p(Y_i|P(\mathbf{x}_i), P_0) = P_0 I(Y_i = 0) + (1 - P_0) I(Y_i > 0)\binom{N_i}{Y_i} P(\mathbf{x}_i)^{Y_i}\big(1 - P(\mathbf{x}_i)\big)^{N_i - Y_i}$$

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Here, P_0 is the proportion of structural zeros, $(1 - P_0)$ is the proportion of sample zeros, and $I(Y_i = 0)$ is the indicator variable. When both structural zeros and sample zeros are considered, i.e., Type 1, the observations follow the probability density function:

$$p(Y_i|P(\mathbf{x}_i), P_0) = P_0 I(Y_i = 0) + (1 - P_0)\left(\frac{N_i}{Y_i}\right) P(\mathbf{x}_i)^{Y_i}\left(1 - P(\mathbf{x}_i)\right)^{N_i - Y_i}$$

To determine the best fit model for the nodule prevalence data, model fit statistics (DIC and WAIC) were calculated for each model, viz. binomial, ZIB Type I, and ZIB Type 0.

Mesh construction

We assume an underlying spatially continuous variable for the observed geostatistical data, which can be modeled with Gaussian random fields. We used the Stochastic Partial Differential Equation (SPDE) approach in the *INLA* package to fit a spatial model and to predict each variable of interest at an unsampled location (2, 53). An approximate solution to SPDE can be found using the finite element method. The finite element representation of the Matérn field is used as a linear combination of basis functions defined on a triangulation of domain D (54). Domain D is subdivided into a triangulated mesh which is formed first by placing the triangle's vertices at the sample locations and then adding other vertices around the regions of spatial prediction.

We constructed the finite element mesh for SPDE approximation to the Gaussian process regression using the boundary of Ethiopia. Triangulation meshes with different cut-off parameters and the maximum length for the triangle inside and outside the boundary were tested for their model fit and computation cost. The mesh that yielded the lowest DIC and WAIC scores without significantly increasing computational cost was chosen.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

259 **Cross-validation**

260 K-fold cross-validation (with k = 10) was run to observe the differences in the predictive
261 accuracy of the candidate models. Different measures of predictive accuracy were calculated by
262 assessing the relationship between the predicted and observed prevalence in the validation
263 dataset. In 10-fold cross-validation, the dataset is divided into ten random sections (or folds) (39,
264 46). Then, model validation runs are performed on each fold (10% of the data labeled the
265 validation dataset) after fitting the data on the remaining nine folds (90% of the data labeled the
266 training dataset). Thus, ten validation runs are performed for 10-fold cross-validation. During
267 each validation run, both Spearman's rank correlation coefficient and the Root Mean Square
268 Error (RMSE) between the observed data and the predicted data for validation samples were
269 calculated to assess accuracy.

270 **Prediction**

271 The posterior distribution of prevalence was estimated at 5 km resolution, accounting for the
272 effect of the variables and the spatial covariance structure. The covariate raster stack was
273 aggregated to 5 km spatial resolution by taking either the mean or sum of the raster cells. The
274 mean of raster cells was calculated for all continuous covariates except population count, for
275 which the sum was calculated. Aggregated data were used to ease the computational burden
276 associated with geospatial prediction at higher resolution. In addition, we calculated the
277 aggregated mean, and the range of predicted prevalence values within a district/implementation
278 unit (IUs). The predicted prevalence map was also used to assess the relationship with the
279 environmental variables fitting the Generalized Additive Model (GAM) curve.

280

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Results

We formulated a Bayesian geostatistical model using INLA to estimate the nationwide pre-intervention prevalence of *O. volvulus* infection in Ethiopia. Nodule prevalence data from 916 unique geopositioned sites were combined with eight different environmental and socio-demographic covariates to construct the geostatistical model. Most of the prevalence data were from western Ethiopia, as eastern Ethiopia is largely unmapped for *O. volvulus* infection prevalence. The mean and the standard deviation of the observed prevalence across the sampling locations in Ethiopia was 17.24±16.32% ranging from 0 to 81.48%. There were 204 sites with zero prevalence (Fig 1).

Model selection and fitting

Four different types of model were tested for the nodule prevalence data viz. binomial without spatial structure, binomial with spatial structure, ZIB type 1 and ZIB type 2, both with spatial structure. These were done without including any environmental and socio-demographic variables in the model. The binomial model that did not account for spatial effects showed higher DIC (9806.988) and WAIC (9816.581) scores (S3 Fig). The addition of spatial effect and accounting for zero inflation with a Type I zero-inflated binomial model decreased the DIC and WAIC scores to 5661.098 and 5916.715, respectively. Thus, ZIB Type I with spatial structure was chosen for modeling the prevalence data. To optimize the SPDE mesh, six different triangulation meshes with different parameters were tested for their model fit and computation cost (S4 Fig, S4 Table). The mesh C yielded the best model fit scores (DIC = 4538.12; WAIC = 4652.22). However, the mesh E yielded a comparable model fit (DIC = 4572.74; WAIC = 4710.781) but was computationally more efficient (45.38 s vs. 1667.33 s) and therefore, mesh E was chosen for fitting the model.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We selected environmental and socio-demographic variables based on the model fit scores of the univariate model. Isothermality was selected from the group of temperature variables, precipitation seasonality from the group of precipitation, and similarly, population density, distance to the nearest river, slope, and Normalized Difference Vegetation Index (NDVI) were selected from the group of socio-demographic, hydrological, and vegetation groups of covariates, respectively. Other combinations were also explored and the inclusion of covariates like soil moisture and flow accumulation further reduced the DIC and WAIC scores (S3 Table).

K-fold cross-validation (k = 10) was done for three different models: one without environmental covariates, one with six covariates, and the other with an additional two covariates (flow accumulation and soil moisture), which revealed that model 3 was superior to model 0 and 1 (S5 Fig). For model 3, calculating the Spearman rank correlation coefficient between the observed prevalence and the predicted prevalence ranged from 0.48 to 0.70 with a median of 0.66. Similarly, the RMSE ranged from 11.09 to 15.1, with a median of 13.18. This suggested a good model fit and accuracy for predictions across the validation datasets.

Model parameters

The regression coefficients were estimated for each covariate included in the model. Since INLA is a Bayesian technique, the regression coefficients are a probability distribution rather than point estimates. A negative coefficient estimate implies a negative association of the variable with the prevalence and vice versa. The significance of the estimates was determined as described in Moraga et al. (55). The association was deemed significant only if both the 95% BCI values were below 0 for negative association and above 0 for positive association.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Out of 8 covariates considered for the final model, four covariates were significantly associated (based on 95% BCI) with *O. volvulus* infection prevalence (Table 1). Soil moisture was positively associated (0.0216, 95% BCI: 0.014 – 0.03) with *O. volvulus* infection prevalence, whereas distance to the nearest river (-0.015, 95% BCI: -0.025 – -0.005), precipitation seasonality (-0.017, 95% BCI: -0.032 – -0.001), and flow accumulation (-0.042, 95% BCI: -0.07 – -0.019) were negatively associated with *O. volvulus* infection prevalence. The regression coefficient of significant variables was at least one to two orders of magnitude greater than the non-significant ones.

Table 1. Mean coefficient estimates and 95% Bayesian credible interval (BCI) for the environmental and socio-demographic variables in the model. Regression coefficients for a particular covariate represent the change in $logit(P)$ for a unit change in that covariate given that all other variables are kept constant.

	Regression coefficients	
Variables	**Mean**	**95% BCI**
Distance to the nearest river	-0.01508	(-0.0252, -0.005)
Soil moisture	0.02158	(0.0135, 0.0297)
Flow accumulation	-0.04225	(-0.0703, -0.0186)
Precipitation seasonality	-0.01645	(-0.0324, -0.0005)
Vegetation index	0.00168	(-0.0072, 0.0105)
Slope	0.00097	(-0.0071, 0.009)
Population density	0.00005	(-0.0003, 0.0003)
Isothermality	-0.00185	(-0.0287, 0.0249)
Intercept	-1.94027	(-4.9233, 1.0416)
Hyper-parameters		
Zero probability parameter	0.33345	(0.327, 0.342)
$\theta 1$ for spatial field	-1.33992	(-1.417, -1.286)
$\theta 2$ for spatial field	-0.09636	(-0.17, -0.046)

95% BCI includes 0.025 quantiles and the 0.975 quantiles of the probability distribution of the coefficients

Hyperparameters defining the SPDE mesh were used to calculate the spatial effect and project the spatial field (S6 Fig). The spatial effect indicates the intrinsic spatial variability in the prevalence estimates, helping us understand the data's spatial structure (47). Further, the spatial field also represents the spatial effect that was not accounted for by the covariates included in the

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

model (55). The mean spatial field is higher in western Ethiopia while it is lower in central Ethiopia and eastern Ethiopia, along with the high standard deviation of the spatial field in the eastern parts.

Model prediction

The predicted prevalence map shows spatial heterogeneity in *O. volvulus* infection prevalence in Ethiopia (Fig 2). Predicted *O. volvulus* infection prevalence is concentrated in the western parts of Ethiopia, with three to four hotspots in southwest Ethiopia. There is a relatively low prevalence of infection in eastern Ethiopia and near to zero prevalence in central Ethiopia. The range of predicted mean prevalence was 0.39 to 55.27%. Similarly, the lower limit of predicted infection prevalence ranged from 0 to 47.28%, while the upper limit of the predicted prevalence ranged from 1.41 to 65.32%. The correlation between the observed and the predicted prevalence was 0.71 (S7 Fig). Due to the geostatistical smoothing effect, some observations with higher prevalence were underestimated and vice-versa.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

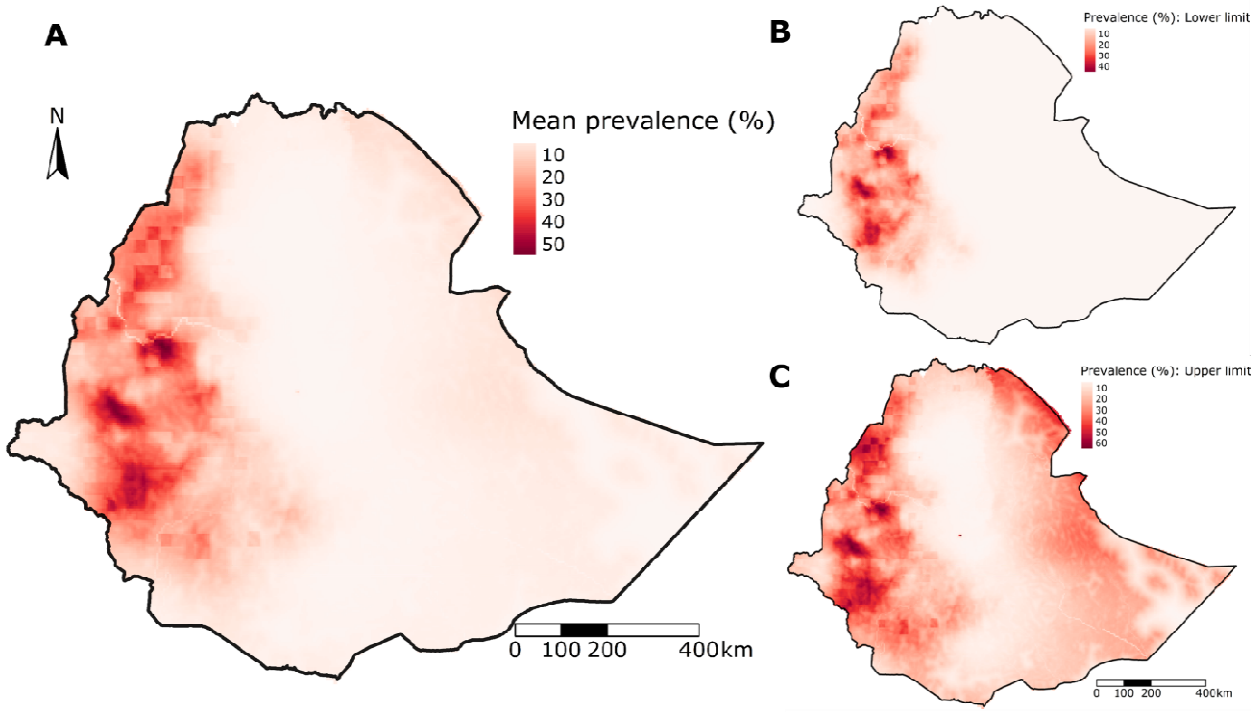


Fig 2. *Onchocerca volvulus* infection prevalence map in Ethiopia generated from the geostatistical model. The mean (A), the lower limit (B), and the upper limit (C) of *O. volvulus* infection prevalence. The prediction interval of the prevalence map is generated from the calculated 95% BCI of fitted prevalence values.

The uncertainty in the prevalence estimates was derived using the standard deviation of the posterior distribution. The uncertainty map shows that the presence of data influenced the uncertainty in the prevalence estimates; i.e., areas with the ground truth data have lower uncertainty (Fig 3A). The uncertainty was higher in eastern Ethiopia due to the lack of ground truth data from those sites. Most of central Ethiopia and some areas in eastern Ethiopia, regardless of the absence of the data, showed low prevalence with lower uncertainty (Fig 3B). There were areas with a high prevalence that had different levels of uncertainty in western Ethiopia. The regions with higher uncertainty almost always corresponded with sparse data from those regions.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

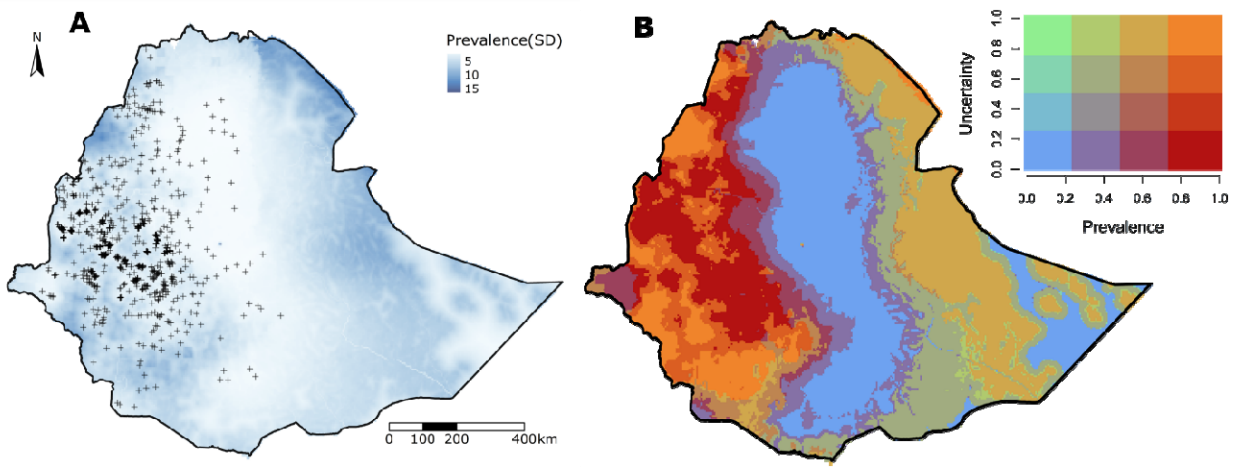


Fig 3. Uncertainty in the estimates of *O. volvulus* infection prevalence from the model. The standard deviation of the posterior distribution of prevalence (A) and the location of the observation are indicated by '+' on the map. The bivariate map (B) shows both prevalence and the uncertainty estimates rescaled from 0 to 1.

A district-level map was created by aggregating the mean prevalence from pixels within the respective districts which represent implementation units (IUs) for MDAi (Fig 4). The aggregated mean prevalence for the first dozen of the most endemic districts were greater than 40%. The difference between the highest and the lowest estimated prevalence pixels (range of mean prevalence) within the districts was as high as 50.72% for a district within the Kemashi zone of Ethiopia.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

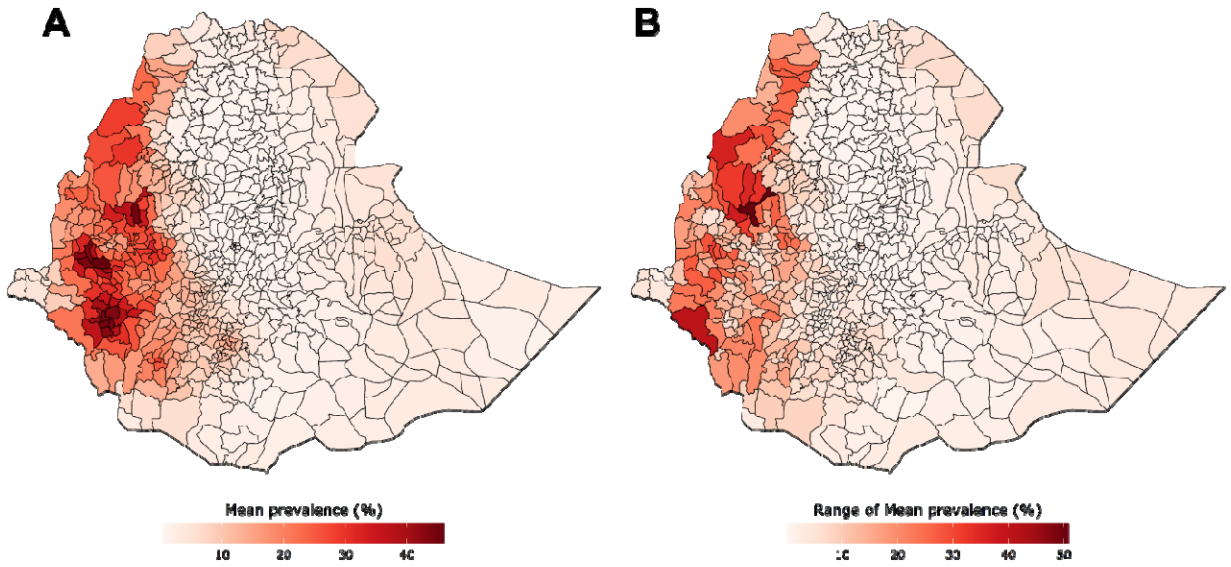


Fig 4. The aggregated mean prevalence and the range of the estimated mean prevalence within Ethiopian districts. The mean of the estimated prevalence of all the pixels within the district level border (**A**) and the range of the estimated prevalence within the district, i.e. the difference between the highest prevalence pixel and the lowest prevalence pixel (**B**), is shown.

Relationship of environmental and socio-demographic covariates on the prevalence

A GAM curve was fitted between the predicted prevalence and the covariates used in the model to assess the relationship between them. The relationship profile of the predicted *O. volvulus* infection prevalence across the range of values of different covariates indicates which ecological conditions are suitable for onchocerciasis transmission (Fig 5). The relationship curve for the distance to the nearest river and the predicted prevalence shows the sharp decline in the *O. volvulus* infection prevalence to around 20-25 km, and the curve continues in the low prevalence region with increased uncertainty as distance increases from the nearest river (Fig 5). There was almost a linear increase in the predicted prevalence with an increase in soil moisture up to around 18 mm.

There was a negative association with the flow accumulation with a considerable increase in uncertainty in the areas with high flow accumulation, i.e., larger rivers. Nevertheless, the areas

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

with lower flow accumulation had a higher predicted prevalence than those with higher flow accumulation, suggesting the importance of intermediate-sized rivers to onchocerciasis epidemiology in Ethiopia. In addition, the relationship curve for the slope shows that a certain degree of slope is favorable for *O. volvulus* infection prevalence (S8 Fig). There is a similar response profile for population density where intermediate population density is favorable for onchocerciasis transmission. There was a steep decline in predicted prevalence with the initial increase in precipitation seasonality. However, there was a mixed non-linear response in the regions with precipitation seasonality from 60 to 130 mm.

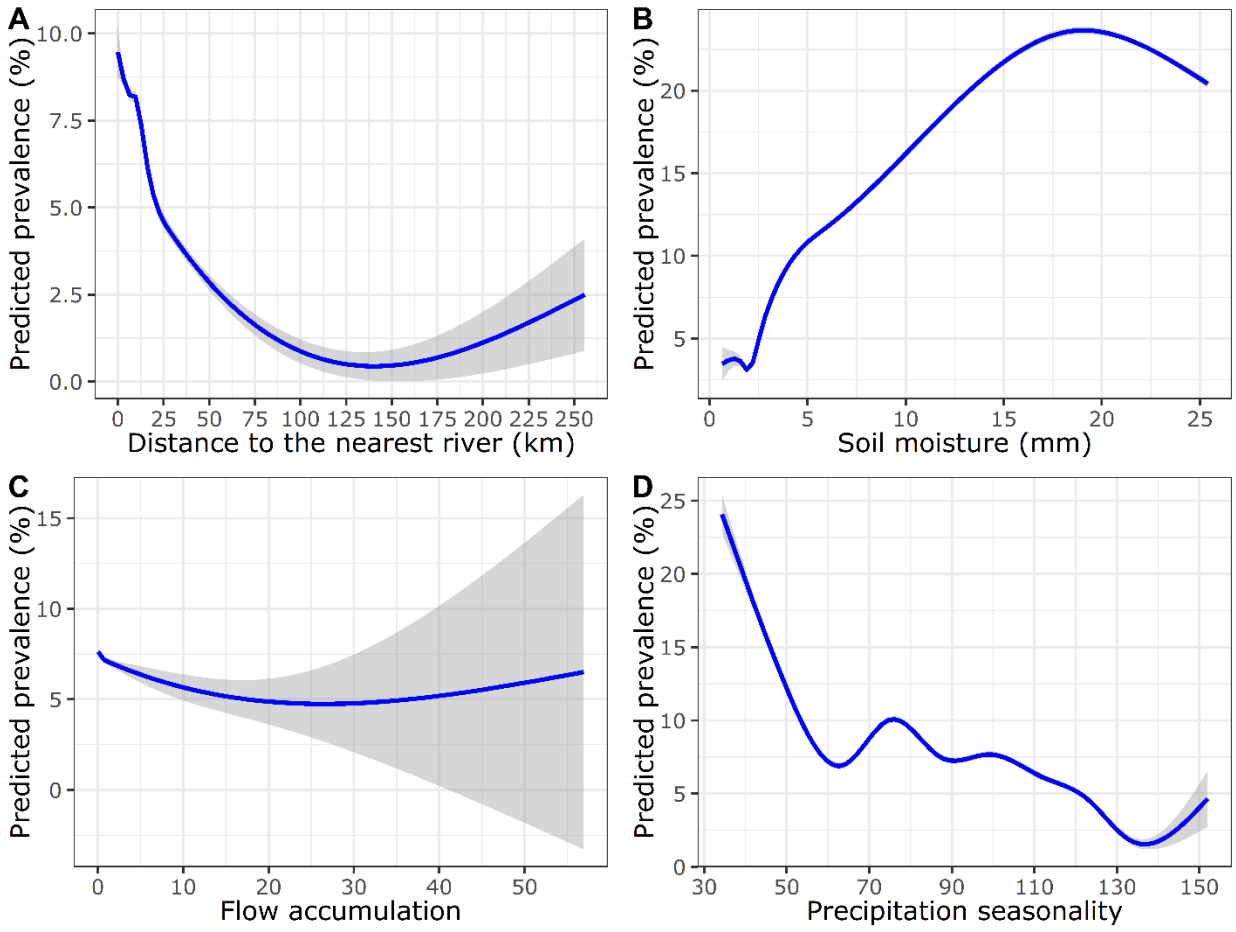


Fig 5. The relationship between the predicted mean prevalence with the significant environmental covariates in the regression model. The curve was fitted using a generalized additive model (GAM) using the smoothing function available in the *ggplot2* package. The

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

shaded region around the curve represents the 95% confidence interval. Flow accumulation had a range of high magnitude compared to other covariates (values ranged from 0 to 100418). Thus, this variable was rescaled from 0 to 100 to make its range comparable with other variables.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Discussion

We generated a country-level geospatial map of *O. volvulus* infection prevalence before the start of MDAi in Ethiopia, accounting for environmental and socio-demographic factors. The prevalence has been extrapolated to the country-level border of Ethiopia, including the eastern regions which were not mapped previously. Predicted prevalence in areas where people do not currently inhabit can indicate the risk of transmission should infected people establish communities. Prevalence was estimated using the pre-intervention nodule prevalence data and therefore represents the infection status before MDAi in Ethiopia. Thus, these predictions can act as a pre-control baseline map to on which decisions concerning new pre-MDAi mapping of likely hypoendemic areas that are not yet under MDAi can be prioritized and the effects of past interventions or of ecological changes at different locations can be assessed.

The predicted infection prevalence was found to be relatively low in the central parts of Ethiopia. This can be attributed to the presence of a significant geographical feature, the Great Rift valley. The elevated highlands along the center and lowland to the east of the Great Rift valley are characterized by low predicted prevalence. The land east of the valley is dry with few rivers (16, 20). On the other hand, the high elevation and slopes in western Ethiopia experience much higher rainfall, resulting in fast-flowing rivers, a specific requirement for blackfly breeding and development. The response profile for slope indicates that there is an optimal slope for the prevalence of *O. volvulus* infection that may be related to the flow characteristics optimal for blackfly breeding.

The spatial pattern of *O. volvulus* infection prevalence predicted across Ethiopia by this geospatial model was consistent with previously published prevalence maps that were based on REMO and other data (17, 19). It is noteworthy that in those earlier maps and in this geospatially

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

435 predicted map there was a high level of spatial heterogeneity in infection prevalence, including
436 heterogeneity within health districts (which are the implementation units for MDAi in Ethiopia).
437 The difference between the highest and lowest prevalence pixel (range) within the districts was
438 as high as 50%. A study in Cameroon reported that hypoendemic areas could sustain low-grade
439 transmission and, therefore, might cause rapid recrudescence in neighboring meso- and
440 hyperendemic areas where the transmission has been successfully controlled (56). Given that
441 much of the unmapped onchocerciasis endemic areas of Ethiopia is hypoendemic, these areas
442 must be identified and treated for elimination of transmission to be reached. Hence, we need to
443 consider the spatial heterogeneity within and between the intervention units while planning the
444 elimination programs.

445 We used a bivariate map to visualize estimated prevalence and the associated uncertainty (Fig 4).
446 The presence/absence of data influences this uncertainty map; i.e., areas with ground truth data
447 have lower prediction errors. This is expected in geostatistical models as they depend on the
448 Euclidean distances between the reported observations (39). Thus, the uncertainty map can
449 indicate where additional data would reduce the overall prediction error of the prevalence map,
450 particularly in the areas with higher prevalence. This can be used to identify regions that might
451 benefit from targeted re-mapping or elimination mapping efforts (3). For example, there are
452 areas with higher prevalence in the west but varying uncertainty. The areas with high uncertainty
453 could be targeted for re-mapping. Similarly, there are areas in the east with both low prevalence
454 and lower uncertainty, i.e., with higher confidence, and thus, do not need to be re-mapped.

455 **Ecological features associated with *O. volvulus* infection prevalence**

456 The major environmental factors significantly associated with infection prevalence were distance
457 to the nearest river, soil moisture, precipitation seasonality, and flow accumulation. As expected,

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

458 there was a negative association between the distance to the nearest river and predicted
459 prevalence. Onchocerciasis has long been recognized as being higher in communities near rivers
460 and this correlation, which has been reported in prior geospatial modeling studies (19, 37, 39), is
461 driven by blackfly breeding and development requirements for fast-flowing rivers, such that
462 villages can be categorized epidemiologically as first, second, or third-line villages based on
463 their proximity to vector breeding sites (39, 57, 58).

464 The relationship curve between the predicted prevalence and the distance to the nearest river
465 shows that there is an initial rapid decline in prevalence followed by a less rapid decline as the
466 distance from the river increases, and the curve asymptotes to a very low prevalence with
467 increased uncertainty as the distance exceeds 100 km. A rapid decline in blackfly biting rate at
468 increasing distance from a river breeding site, based on vector biting rate data collected in
469 northern Cameroon over three years, has been reported previously (59). Similarly, a mark-
470 recapture study found a logarithmic decline in the proportional fly biting density as the distance
471 increased from the marking site (60), and a mark-release-recapture study in Ghana in West
472 Africa reported the average flight range of *S. damnosum* may be as high as 27 km (58). While
473 Ethiopia is host to several different competent blackfly vector species, the part of the curve
474 where the change in slope declines is consistent with this estimated flight range viz., ~20-25 km.
475 However, the curve does not reach its lowest point until 100 km, suggesting that the parasites
476 could be transmitted beyond the average dispersal range of an individual blackfly. This could be
477 because the dispersal range for gravid blood-seeking and ovipositing female blackflies has been
478 reported to be greater than the average dispersal range at around 60-100 km from the river (58).
479 In addition, wind-assisted long-distance migration of blackflies of hundreds of km and
480 transmission due to the human migration have also been reported (61-63). Thus, this study

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

supports that longer range migration is likely and that it also likely contributes to *O. volvulus* transmission.

We observed a positive association between soil moisture and *O. volvulus* infection prevalence. Soil moisture is high in areas with high precipitation or near water bodies, including rivers—i.e., where there are suitable blackfly breeding sites. Soil moisture is an indicator of agricultural suitability, and agricultural areas have historically known to have high prevalence of *O. volvulus* infection (19, 37, 64). Agricultural lands and farms in these areas tend to be near rivers for easy irrigation. Therefore, the increased prevalence of *O. volvulus* infection among people involved in agriculture and farming (37, 65, 66) is presumably because these workers are generally outdoors, often in proximity to rivers, and thus experience increased exposure to blackflies (7, 9, 67).

Flow accumulation is used in hydrogeology as a proxy for river grades and represents the cumulative number of cells in a raster object that flow into a given cell: high flow accumulation represents large rivers, and lower flow accumulation represents secondary rivers and their tributaries. It has been used to map onchocerciasis hotspots in hypoendemic settings of the Democratic Republic of Congo (68). In this study, flow accumulation was negatively associated with *O. volvulus* infection prevalence, meaning that the infection was more common in the communities near the secondary rivers and tributaries than the large rivers. The primary vectors of onchocerciasis in Ethiopia are *S. damnosum s.l.* and *S. neavei* (69). In a study describing the ecological study of West African *Simulium* spp., *S. damnosum s.l.* were found in rivers of medium width with a lower flow accumulation than the large size rivers (6). Furthermore, the important characteristic of *S. neavei* is the obligatory phoretic association of larvae with the freshwater crabs which are more common in sheltered smaller river streams in the forests than in larger rivers (70, 71). The forested riverine areas, where *S. neavei* are found, is a dominant

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

ecotype in southwestern midland and highlands of Ethiopia, where onchocerciasis is highly endemic and from where most of the data were collected for this study (9, 72).

Similarly, precipitation seasonality was also negatively associated with the predicted prevalence, i.e., the prevalence was high in the areas with lower precipitation seasonality. Areas with high precipitation seasonality might have ephemeral rather than perennial streams. If the breeding sites are ephemeral, blood-feeding by blackflies would only happen during some part of the year, lowering the annual biting rate, a key parameter in *O. volvulus* transmission (73-75). Southeast Ethiopia, characterized by low prevalence of *O. volvulus* infection, has high seasonality in precipitation that is characterized by two short wet seasons with a dry period in between (76). However, the southwestern areas where the disease is most endemic have low precipitation seasonality and high annual precipitation (77).

As one would expect, the environmental factors that were significantly associated with *O. volvulus* infection prevalence are all exert strong influence on vector breeding and thus impact blackfly density and biting rates. The implication of this strong association between determinants of vector breeding and infection prevalence implies that spatial variation in vector breeding drives the spatial variation of *O. volvulus* infection prevalence in Ethiopia and that the geospatial model we present here, based on nodule prevalence data, is also predictive of vector distribution. This suggests that ongoing climate change, which is affecting the pattern of precipitation (78, 79), and other anthropogenic environmental changes such as changing river flow with the construction of river dams for hydroelectricity or irrigation might significantly change vector distribution and thus the spatial occurrence of the disease (80, 81). The impacts of these changes could be modeled using this approach.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Model limitations and recommendations

The geospatial model we report here incorporates different environmental and socio-demographic variables that are known to influence the transmission and prevalence of *O. volvulus* infection and the distribution of blackflies. However, the data incorporated in the model do not include all factors that may be epidemiologically relevant, such as direct/indirect interventions affecting infection prevalence and human behaviors that may increase or decrease the risk of infection. The non-uniform mean spatial field across the triangulation mesh shows that there might be some effects that are unaccounted for by the model (S6 Fig), and the possibility remains that an unidentified covariate that closely resembles the spatial field might aid in explaining the spatial variation in prevalence. In addition, the inclusion of blackfly distribution maps based on the identification of breeding sites and their productivity might improve the model fit. Unfortunately, such data are not available for Ethiopia.

Some variables that we expected may correlate with prevalence, such as human host population density and vegetation, were not shown to be significantly associated in these analyses. Blackflies are not usually reported to be found in dense urban environments and, similarly, vegetation cover is essential for blackfly breeding and thus *O. volvulus* transmission (37, 64, 82). We suggest that the lack of association might be because the country-wide spatial scale neutralizes factors that impact prevalence at a smaller geographic scale. Therefore, targeted spatial analysis in regions with differences in vegetation near rivers or with differences in rural-urban indices (83) might be helpful to explore the effects of these variables on infection prevalence. Furthermore, *O. volvulus* transmission is highly dynamic, not just spatially but also temporally. Extending the current spatial model to a spatio-temporal model might improve the

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

548 model fit, which requires both the prevalence and the covariate data at sufficient temporal
549 resolution.

550 We could not include prevalence measures based on other diagnostic methods for Ethiopia
551 because fine-scaled prevalence data based on mf counts from skin snips or antibody tests (Ov16)
552 were not available. However, these data could be used as an alternative to, or in addition to,
553 nodule prevalence. Combining data across methods is challenging, as correlations between mf
554 counts and nodule counts can be highly variable (however, see (18)) and the correlation between
555 these measures and Ov16 seropositivity prevalence is unclear. Nevertheless, the map presented
556 here could be used by onchocerciasis elimination programs to direct resources for elimination
557 mapping because elimination mapping of any disease can be expensive (3), and the method
558 described here may be an inexpensive first step that can extrapolate country-wide prevalence
559 from existing data and thus better target re-mapping efforts.

560 ## Conclusion

561 Onchocerciasis programs have transitioned from control of onchocerciasis as a public health
562 problem to elimination of *O. volvulus* transmission, triggering the need to develop new tools to
563 more efficiently prioritize decisions concerning elimination mapping and interventions in
564 hypoendemic foci that were not previously targeted for intervention. To this end, we have
565 generated a baseline pre-intervention prevalence map for the whole of Ethiopia using geospatial
566 modelling that is based on pre-intervention nodule prevalence data and spatial variation in
567 different environmental and socio-demographic factors. We extrapolated existing historical
568 nodule prevalence measures to previously unmapped regions of Ethiopia and quantified
569 uncertainty in predicted prevalence. This map could be used as an aid to decision making on
570 where and how to (a) extend elimination mapping into areas identified as likely hypoendemic

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

571 foci and (b) prioritize the allocation of scarce health system resources to areas most likely to
572 benefit from that allocation. Furthermore, this study found that hydrological variables such as
573 distance to the nearest river, soil moisture, precipitation seasonality, and flow accumulation were
574 significant in describing the spatial heterogeneity of *O. volvulus* infection in Ethiopia. All these
575 ecological features are related to the suitability of an area for vector breeding, movement, biting
576 behavior, and density, leading to the conclusion that vector suitability and movement are the
577 primary determinants of the spatial distribution of *O. volvulus* infection in Ethiopia.
578 Consequently, changes in these ecological features due to anthropomorphic changes in climate,
579 agriculture, vegetation type (e.g., slash-and-clear), or construction of hydroelectric or irrigation
580 dams might significantly alter the occurrence of the disease. We suggest, therefore, that the
581 importance of these vector-related ecological factors in determining infection distribution and
582 intensity reaffirms that inclusion of vector control could augment current interventions based
583 primarily on prophylactic chemotherapy.

584 **Acknowledgements**

585 The authors wish to acknowledge Mr Sindew M. Feleke at the Ethiopian Public Health Institute
586 (EPHI), Dr Moses Katabarwa at the Carter Center for assistance in navigating the data sources
587 and providing insights into Ethiopian control programs, and Dr Katie Crawford and Mr Haylo
588 Roberts for helpful discussion during the analysis.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Diggle P, Ribeiro PJ. Model-based geostatistics. New York, NY: Springer; 2007.
2. Moraga P. Geospatial health data: modeling and visualization with R-INLA and Shiny. Boca Raton: CRC Press; 2020.
3. Rebollo MP, Zoure H, Ogoussan K, Sodahlon Y, Ottesen EA, Cantey PT. Onchocerciasis: shifting the target from control to elimination requires a new first-step—elimination mapping. International Health. 2018;10(suppl_1):i14-i9.
4. Diggle P. Model-based geostatistics for global public health: methods and applications. Boca Raton: Taylor & Francis; 2019.
5. Eneanya OA, Fronterre C, Anagbogu I, Okoronkwo C, Garske T, Cano J, et al. Mapping the baseline prevalence of lymphatic filariasis across Nigeria. Parasites & Vectors. 2019;12(1):440.
6. Cheke RA, Young S, Garms R. Ecological characteristics of *Simulium* breeding sites in West Africa. Acta Tropica. 2017;167:148-56.
7. Adeleke MA, Mafiana CF, Sam-Wobo SO, Olatunde GO, Ekpo UF, Akinwale OP, et al. Biting behaviour of *Simulium damnosum* complex and *Onchocerca volvulus* infection along the Osun River, Southwest Nigeria. Parasites & vectors. 2010;3(1):1–7.
8. Lamberton PH, Cheke RA, Walker M, Winskill P, Osei-Atweneboana MY, Tirados I, et al. Onchocerciasis transmission in Ghana: biting and parous rates of host-seeking sibling species of the *Simulium damnosum* complex. Parasites & Vectors. 2014;7(1):511.
9. Mose AD, Mamo BT, Alamirew SY. Monthly dynamics and biting behavior of principal onchocerciasis vector (*Simulium damnosum sl*) in endemic area of Southwest Ethiopia. International Journal of One Health. 2020;6(1):23–8.
10. Basáñez M-G, Pion SDS, Churcher TS, Breitling LP, Little MP, Boussinesq M. River Blindness: A Success Story under Threat? PLoS Medicine. 2006;3(9):e371.
11. Chesnais CB, Nana-Djeunga HC, Njamnshi AK, Lenou-Nanga CG, Boullé C, Bissek A-CZ-K, et al. The temporal relationship between onchocerciasis and epilepsy: a population-based cohort study. The Lancet Infectious Diseases. 2018;18(11):1278-86.
12. Colebunders R, Njamnshi AK, van Oijen M, Mukendi D, Kashama JM, Mandro M, et al. Onchocerciasis-associated epilepsy: From recent epidemiological and clinical findings to policy implications. Epilepsia Open. 2017;2(2):145–52.
13. Katabarwa MN, Zarroug IMA, Negussu N, Aziz NM, Tadesse Z, Elmubark WA, et al. The Galabat-Metema cross-border onchocerciasis focus: The first coordinated interruption of onchocerciasis transmission in Africa. PLOS Neglected Tropical Diseases. 2020;14(2):e0007830.
14. Vlaminck J, Fischer PU, Weil GJ. Diagnostic Tools for Onchocerciasis Elimination Programs. Trends in Parasitology. 2015;31(11):571-82.
15. Noma M, Nwoke BEB, Nutall I, Tambala PA, Enyong P, Namsenmo A, et al. Rapid epidemiological mapping of onchocerciasis (REMO): its application by the African Programme for Onchocerciasis Control (APOC). Annals of Tropical Medicine & Parasitology. 2002;96(sup1):S29-S39.
16. Noma M, Zouré HG, Tekle AH, Enyong PA, Nwoke BE, Remme JH. The geographic distribution of onchocerciasis in the 20 participating countries of the African Programme for Onchocerciasis Control: (1) priority areas for ivermectin treatment. Parasites & Vectors. 2014;7(1):325.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

17. Zouré HG, Noma M, Tekle AH, Amazigo UV, Diggle PJ, Giorgi E, et al. The geographic distribution of onchocerciasis in the 20 participating countries of the African Programme for Onchocerciasis Control: (2) pre-control endemicity levels and estimated number infected. Parasites & Vectors. 2014;7(1):326.
18. Coffeng LE, Pion SDS, O'Hanlon S, Cousens S, Abiose AO, Fischer PU, et al. Onchocerciasis: The Pre-control Association between Prevalence of Palpable Nodules and Skin Microfilariae. PLoS Neglected Tropical Diseases. 2013;7(4):e2168.
19. Cromwell EA, Osborne JCP, Unnasch TR, Basáñez M-G, Gass KM, Barbre KA, et al. Predicting the environmental suitability for onchocerciasis in Africa as an aid to elimination planning. PLOS Neglected Tropical Diseases. 2021;15(7):e0008824.
20. Meribo K, Kebede B, Feleke SM, Mengistu B, Mulugeta A, Sileshi M, et al. Review of Ethiopian onchocerciasis elimination programme. Ethiopian medical journal. 2017;55(Suppl 1):55.
21. Feleke SM, Tadesse G, Mekete K, Tekle AH, Kebede A. Epidemiological Mapping of Human Onchocerciasis in Transmission Suspected Districts of Bale, Borena, and West Arsi Zones of Eastern Ethiopia. Interdisciplinary Perspectives on Infectious Diseases. 2016;2016:1-5.
22. Mengitsu B, Shafi O, Kebede B, Kebede F, Worku DT, Herero M, et al. Ethiopia and its steps to mobilize resources to achieve 2020 elimination and control goals for neglected tropical diseases: Spider webs joined can tie a lion. International Health. 2016;8(suppl 1):i34-i52.
23. Katabarwa MN, Endeshaw T, Taye A, Tadesse Z, Frank RO. The disappearance of onchocerciasis without intervention in Tigray Region in Northwest Ethiopia. Pathogens and Global Health. 2014;108(3):123-.
24. Ayalew F, Atnafu DD, Bedimo M, Mulatu K. Determinants of community-led ivermectin treatment adherence for onchocerciasis control in Western Ethiopia: a case-control study. Tropical Medicine and Health. 2020;48(1):22.
25. Kifle B, Nigatu M. Compliance to a Five-Year Biannual Ivermectin Treatment for Onchocerciasis Elimination and Its Determinants among Adults in the Bench Maji Zone, Southwest Ethiopia: A Community-Based Cross-Sectional Study. Journal of Parasitology Research. 2021;2021:e8866639.
26. Yirga D, Deribe K, Woldemichael K, Wondafrash M, Kassahun W. Factors associated with compliance with community directed treatment with ivermectin for onchocerciasis control in Southwestern Ethiopia. Parasites & Vectors. 2010;3(1):48.
27. Gebrezgabiher G, Mekonnen Z, Yewhalaw D, Hailu A. Reaching the last mile: main challenges relating to and recommendations to accelerate onchocerciasis elimination in Africa. Infectious Diseases of Poverty. 2019;8(1):60.
28. Lakwo T, Oguttu D, Ukety T, Post R, Bakajika D. Onchocerciasis Elimination: Progress and Challenges. Research and Reports in Tropical Medicine. 2020;11:81-95.
29. Hamley JID, Blok DJ, Walker M, Milton P, Hopkins AD, Hamill LC, et al. What does the COVID-19 pandemic mean for the next decade of onchocerciasis control and elimination? Transactions of The Royal Society of Tropical Medicine and Hygiene. 2021;115(3):269-80.
30. Plaisier AP, Alley ES, van Oortmarssen GJ, Boatin BA, Habbema JD. Required duration of combined annual ivermectin treatment and vector control in the Onchocerciasis Control Programme in west Africa. Bulletin of the World Health Organization. 1997;75(3):237-45.
31. Remme J, Dadzie KY, Rolland A, Thylefors B. Ocular onchocerciasis and intensity of infection in the community. I. West African savanna. Tropical medicine and parasitology. 1989;40(3):340-7.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

32. Winnen M, Plaisier AP, Alley ES, Nagelkerke NJD, van Oortmarssen G, Boatin BA, et al. Can ivermectin mass treatments eliminate onchocerciasis in Africa? Bulletin of the World Health Organization. 2002;80(5):384-91.
33. Kang SY, Battle KE, Gibson HS, Ratsimbasoa A, Randrianarivelojosia M, Ramboarina S, et al. Spatio-temporal mapping of Madagascar's Malaria Indicator Survey results to assess *Plasmodium falciparum* endemicity trends between 2011 and 2016. BMC Medicine. 2018;16(1):71.
34. ESPEN. Site level onchocerciasis prevalence data. Data. ESPEN; 2020 2020.
35. Ngoumou P, Walsh JF, others. A manual for rapid epidemiological mapping of onchocerciasis. World Health Organization; 1993 1993.
36. Barro AS, Oyana TJ. Predictive and epidemiologic modeling of the spatial risk of human onchocerciasis using biophysical factors: A case study of Ghana and Burundi. Spatial and Spatio-temporal Epidemiology. 2012;3(4):273-85.
37. Eneanya OA, Koudou BG, Aboulaye M, Elvis AA, Souleymane Y, Kouakou M-M, et al. Progress towards onchocerciasis elimination in Côte d'Ivoire: A geospatial modelling study. PLOS Neglected Tropical Diseases. 2021;15(2):e0009091.
38. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment. 2017;202:18-27.
39. O'Hanlon SJ, Slater HC, Cheke RA, Boatin BA, Coffeng LE, Pion SDS, et al. Model-Based Geostatistical Mapping of the Prevalence of *Onchocerca volvulus* in West Africa. PLOS Neglected Tropical Diseases. 2016;10(1):e0004328.
40. van den Hoogen J, Geisen S, Routh D, Ferris H, Traunspurger W, Wardle DA, et al. Soil nematode abundance and functional group composition at a global scale. Nature. 2019;572(7768):194-8.
41. Hijmans RJ, Van Etten J, Cheng J, Mattiuzzi M, Sumner M, Greenberg JA, et al. Package 'raster'. R package. 2015;734.
42. Team R. RStudio: Integrated Development Environment for R. Boston, MA: RStudio, PBC; 2021 2021.
43. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021 2021.
44. Fox J, Weisberg S. An R Companion to Applied Regression. Third ed. Thousand Oaks CA: Sage; 2019 2019.
45. Schloerke B, Crowley J, Cook D. Package 'GGally'. Extension to 'ggplot2.'See https://cran. r-project. org/web/packages/GGally …; 2018.
46. Moraga P, Cano J, Baggaley RF, Gyapong JO, Njenga SM, Nikolay B, et al. 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. 2015;8(1):560.
47. Lezama-Ochoa N, Pennino MG, Hall MA, Lopez J, Murua H. Using a Bayesian modelling approach (INLA-SPDE) to predict the occurrence of the Spinetail Devil Ray (*Mobular mobular*). Scientific Reports. 2020;10(1):18822.
48. Blangiardo M, Cameletti M. Spatial and spatio-temporal Bayesian models with R-INLA. Chichester, West Sussex: John Wiley and Sons, Inc; 2015 2015. 1 p.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

724 49. Karagiannis-Voules D-A, Scholte RGC, Guimarães LH, Utzinger J, Vounatsou P.
725 Bayesian Geostatistical Modeling of Leishmaniasis Incidence in Brazil. PLoS Neglected
726 Tropical Diseases. 2013;7(5):e2213.
727 50. Osgood-Zimmerman A, Millear AI, Stubbs RW, Shields C, Pickering BV, Earl L, et al.
728 Mapping child growth failure in Africa between 2000 and 2015. Nature. 2018;555(7694):41-7.
729 51. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use
730 in disease mapping. Biometrics. 1987:671–81.
731 52. Asmarian N, Ayatollahi SMT, Sharafi Z, Zare N. Bayesian Spatial Joint Model for
732 Disease Mapping of Zero-Inflated Data with R-INLA: A Simulation Study and an Application to
733 Male Breast Cancer in Iran. International Journal of Environmental Research and Public Health.
734 2019;16(22):4460.
735 53. Lindgren F, Rue Ha, Lindström J. An explicit link between Gaussian fields and Gaussian
736 Markov random fields: the stochastic partial differential equation approach. Journal of the Royal
737 Statistical Society: Series B (Statistical Methodology). 2011;73(4):423–98.
738 54. Cameletti M, Lindgren F, Simpson D, Rue H. Spatio-temporal modeling of particulate
739 matter concentration through the SPDE approach. AStA Advances in Statistical Analysis.
740 2013;97(2):109-31.
741 55. Moraga P, Dean C, Inoue J, Morawiecki P, Noureen SR, Wang F. Bayesian spatial
742 modelling of geostatistical data using INLA and SPDE methods: A case study predicting malaria
743 risk in Mozambique. Spatial and Spatio-temporal Epidemiology. 2021;39:100440.
744 56. Katabarwa MN, Eyamba A, Chouaibou M, Enyong P, Kuété T, Yaya S, et al. Does
745 onchocerciasis transmission take place in hypoendemic areas? a study from the North Region of
746 Cameroon: Does onchocerciasis transmission take place in hypoendemic areas? Tropical
747 Medicine & International Health. 2010;15(5):645-52.
748 57. Ngoumou P, Walsh JF, Mace JM. A rapid mapping technique for the prevalence and
749 distribution of onchocerciasis: a Cameroon case study. Annals of Tropical Medicine and
750 Parasitology. 1994;88(5):463-74.
751 58. WHO. Report of the Third Meeting of the WHO Onchocerciasis Technical Advisory
752 Subgroup Geneva, Switzerland, 26□28 February 2019. 2020 2020.
753 59. Renz A, Wenk P. Studies on the dynamics of transmission of onchocerciasis in a Sudan-
754 savanna area of North Cameroon I: Prevailing *Simulium* vectors, their biting rates and age-
755 composition at different distances from their breeding sites. Annals of Tropical Medicine &
756 Parasitology. 1987;81(3):215-28.
757 60. Thompson BH. Studies on the flight range and dispersal of *Simulium damnosum*
758 (Diptera: Simuliidae) in the rain-forest of Cameroon. Annals of Tropical Medicine &
759 Parasitology. 1976;70(3):343-54.
760 61. Baker RHA, Guillet P, Sékétéli A, Poudiougo P, Boakye D, Wilson MD, et al. Progress
761 in controlling the reinvasion of windborne vectors into the western area of the Onchocerciasis
762 Control Programme in West Africa. Philosophical Transactions of the Royal Society of London
763 B, Biological Sciences. 1990;328(1251):731-50.
764 62. Garms R, Walsh JF, Davies JB. Studies on the reinvasion of the Onchocerciasis Control
765 Programme in the Volta River Basin by *Simulium damnosum s.I.* with emphasis on the south-
766 western areas. Tropenmedizin Und Parasitologie. 1979;30(3):345-62.
767 63. Hedtke SM, Kuesel AC, Crawford KE, Graves PM, Boussinesq M, Lau CL, et al.
768 Genomic Epidemiology in Filarial Nematodes: Transforming the Basis for Elimination Program
769 Decisions. Frontiers in Genetics. 2020;10:1282.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

64. De Sole G, Kloos H. Transmission patterns of onchocerciasis in southwest Ethiopia. Parassitologia. 1976;18(1-3):53-65.
65. Dunn C, Callahan K, Katabarwa M, Richards F, Hopkins D, Jr PCW, et al. The Contributions of Onchocerciasis Control and Elimination Programs toward the Achievement of the Millennium Development Goals. PLOS Neglected Tropical Diseases. 2015;9(5):e0003703.
66. Oladepo O, Brieger WR, Otusanya S, Kale OO, Offiong S, Titiloye M. Farm land size and onchocerciasis status of peasant farmers in south-western Nigeria. Tropical Medicine & International Health. 1997;2(4):334-40.
67. Opoku AA. The ecology and biting activity of blackflies (Simuliidae) and the prevalence of onchocerciasis in an agricultural community in Ghana. West African Journal of Applied Ecology. 2006;9(1).
68. Kelly-Hope LA, Unnasch TR, Stanton MC, Molyneux DH. Hypo-endemic onchocerciasis hotspots: defining areas of high risk through micro-mapping and environmental delineation. Infectious Diseases of Poverty. 2015;4:36.
69. Ethiopia FMoH. Guidelines for onchocerciasis elimination in Ethiopia 2015 [updated 2015.
70. Lewis DJ, Hanney PW. On the *Simulium neavei* complex (Diptera: Simuliidae). Proceedings of the Royal Entomological Society of London Series B, Taxonomy. 1965;34(1-2):12-6.
71. McMahon JP. The Discovery of the early Stages of *Simulium neavei* in phoretic Association with Crabs and a Description of the Pupa and the Male. Bulletin of Entomological Research. 1951;42(2):419-26.
72. Taye A, Gebre-Michael T, Taticheff S. Onchocerciasis in Gilgel Ghibe River valley Southwest Ethiopia. East African Medical Journal. 2000;77(2).
73. Dietz K. The population dynamics of onchocerciasis. In: Anderson RM, editor. The Population Dynamics of Infectious Diseases: Theory and Applications. Population and Community Biology. Boston, MA: Springer US; 1982. p. 209-41.
74. Duerr HP, Eichner M. Epidemiology and control of onchocerciasis: The threshold biting rate of savannah onchocerciasis in Africa. International Journal for Parasitology. 2010;40(6):641-50.
75. Duerr HP, Raddatz G, Eichner M. Control of onchocerciasis in Africa: Threshold shifts, breakpoints and rules for elimination. International Journal for Parasitology. 2011;41(5):581-9.
76. Liebmann B, Bladé I, Kiladis GN, Carvalho LMV, B. Senay G, Allured D, et al. Seasonality of African Precipitation from 1996 to 2009. Journal of Climate. 2012;25(12):4304-22.
77. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology. 2017;37(12):4302-15.
78. Cheke RA, Basáñez M-G, Perry M, White MT, Garms R, Obuobie E, et al. Potential effects of warmer worms and vectors on onchocerciasis transmission in West Africa. Philosophical Transactions of the Royal Society B: Biological Sciences. 2015;370(1665):20130559.
79. Viste E, Korecha D, Sorteberg A. Recent drought and precipitation tendencies in Ethiopia. Theoretical and Applied Climatology. 2013;112(3-4):535-51.
80. Abong RA, Amambo GN, Hamid AA, Enow BA, Beng AA, Nietcho FN, et al. The Mbam drainage system and onchocerciasis transmission post ivermectin mass drug

medRxiv preprint doi: https://doi.org/10.1101/2022.01.10.22269016; this version posted January 11, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

administration (MDA) campaign, Cameroon. PLOS Neglected Tropical Diseases. 2021;15(1):e0008926.

81. Zarroug IMA, Elaagip A, Gumaa SG, Ali AK, Ahmed A, Siam HAM, et al. Notes on distribution of *Simulium damnosum s. l.* along Atbara River in Galabat sub-focus, eastern Sudan. BMC Infectious Diseases. 2019;19(1):477.

82. Macé JM, Boussinesq M, Ngoumou P, Enyegue Oye J, Koéranga A, Godin C. Country-wide rapid epidemiological mapping of onchocerciasis (REMO) in Cameroon. Annals of Tropical Medicine & Parasitology. 1997;91(4):379-91.

83. Pesaresi M, Freire S. GHS settlement grid, following the REGIO model 2014 in application to GHSL Landsat and CIESIN GPW v4-multitemporal (1975-1990-2000-2015). 2016.