


BMJ Open Exploring malaria prediction models in Togo: a time series forecasting by health district and target group

Anne Thomas ^{1,2,3,4,5} Tchaa Abalo Bakai,^{1,2,3,4,5,6} Tinah Atcha-Oubou,⁶ Tchassama Tchadjobo,⁶ Muriel Rabilloud,^{1,2,3,4} Nicolas Voirin⁵

To cite: Thomas A, Bakai TA, Atcha-Oubou T, *et al*. Exploring malaria prediction models in Togo: a time series forecasting by health district and target group. *BMJ Open* 2024;**14**:e066547. doi:10.1136/bmjopen-2022-066547

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2022-066547>).

Received 11 July 2022

Accepted 16 November 2023



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Université de Lyon, Lyon, France

²Université Lyon 1, Villeurbanne, France

³Service de Biostatistique et Bioinformatique, Pôle Santé Publique, Hospices Civils de Lyon, Lyon, France

⁴Équipe Biostatistiques Santé, Laboratoire de Biométrie et Biologie Évolutive, CNRS, UMR 5558, Villeurbanne, France

⁵Epidemiology and modelling of infectious diseases (EPIMOD), Lent, France

⁶Programme National de Lutte contre le Paludisme (PNLP), Lomé, Togo

Correspondence to

Dr Anne Thomas;
athomas@epimod.fr

ABSTRACT

Objectives Integrating malaria prediction models into malaria control strategies can help to anticipate the response to seasonal epidemics. This study aimed to explore the possibility of using routine malaria data and satellite-derived climate data to forecast malaria cases in Togo.

Methods Generalised additive (mixed) models were developed to forecast the monthly number of malaria cases in 40 health districts and three target groups. Routinely collected malaria data from 2013 to 2016 and meteorological and vegetation data with a time lag of 1 or 2 months were used for model training, while the year 2017 was used for model testing. Two methods for selecting lagged meteorological and environmental variables were compared: a first method based on statistical approach ('SA') and a second method based on biological reasoning ('BR'). Both methods were applied to obtain a model per target group and health district and a mixed model per target group and health region with the health district as a random effect. The predictive skills of the four models were compared for each health district and target group.

Results The most selected predictors in the models per district for the 'SA' method were the normalised difference vegetation index, minimum temperature and mean temperature. The 'SA' method provided the most accurate models for the training period, except for some health districts in children ≥ 5 years old and adults and in pregnant women. The most accurate models for the testing period varied by health district and target group, provided either by the 'SA' method or the 'BR' method. Despite the development of models with four different approaches, the number of malaria cases was inaccurately forecasted.

Conclusions These models cannot be used as such in malaria control activities in Togo. The use of finer spatial and temporal scales and non-environmental data could improve malaria prediction.

INTRODUCTION

Malaria is a vector-borne infectious disease that is widespread in the WHO African Region, with an estimated 228 million malaria cases in 2020 (95% of cases worldwide).¹ In Togo, despite large-scale interventions implemented for at least the last 15 years by

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Monthly data of confirmed malaria cases routinely collected by the National Malaria Control Programme of Togo were analysed in this study.
- ⇒ Malaria prediction models were developed with four different approaches for 40 health districts and 3 target groups.
- ⇒ Meteorological and environmental predictors derived from satellite data were used in the models.
- ⇒ The time series for the period 2013–2016 were used for model training, while the 2017 time series were used for model testing.
- ⇒ Non-environmental data such as intervention data or clinical data were not included in the models.

the National Malaria Control Programme (NMCP) to prevent malaria in target groups, malaria remains a major public health concern and socioeconomic burden. Indeed, in 2020, malaria accounted for 51% of care provided by community health workers, 36% of outpatient consultations, 27% of hospitalisations and 10% of deaths.²

Seasonal and geographical variations of malaria have been identified in Togo, with an increase in confirmed malaria cases during or shortly after a rainy season and greater seasonal variations of malaria in the northern health districts, particularly in children < 5 years old.³ Several studies in sub-Saharan Africa showed the association between meteorological and environmental factors and malaria cases.^{4–10} A systematic review published by Reiner *et al* showed that temperature and rainfall were often found to be significant predictors of malaria seasonality, along with vegetation indices such as the normalised difference vegetation index (NDVI).¹¹ Although these factors were the most frequently investigated, other factors such as relative humidity and wind speed were used in studies but more rarely.¹¹ Nevertheless, the association between these meteorological and environmental factors and

malaria cases or incidence, and the time lag, varied by location.^{5,7,11}

The development of malaria early warning systems (MEWSs), promoted by the WHO since 2001, can help malaria control programmes to anticipate the response to seasonal epidemics and reduce the morbidity and mortality of malaria.^{12–14} The lag of several weeks or months between meteorological data and variations of malaria cases gives hope in the interest of malaria forecasting. Furthermore, with the availability of and access to routine malaria data and quality satellite-derived climate data over several years, the conditions were assumed to be met for developing malaria prediction models.^{12,13} In sub-Saharan Africa, prediction models for malaria cases were developed in Burkina Faso,⁶ Kenya,¹⁵ Uganda,¹⁶ Mozambique,¹⁷ Burundi¹⁸ and South Africa.¹⁹ Except for Burundi, the analyses were conducted in a specific district or in health facilities representative of the diversity of malaria transmission. However, these models are difficult to generalise to the whole country.

This study aimed to explore the possibility of using routinely collected malaria data and meteorological and environmental predictors to forecast the monthly number of malaria cases in Togo.

METHODS

Setting

Located in West Africa, Togo is bordered by Ghana, Burkina Faso, Benin and the Bight of Benin and has a total area of 56 785 km².²⁰ Togo's population was estimated at 8.28 million in 2020.²¹ From 2013 to 2017, Togo was divided into 6 health regions (from north to south: Savanes, Kara, Centrale, Plateaux, Maritime and Lome-Commune) and 40 health districts (figure 1).

Malaria data

Monthly data of confirmed malaria cases were routinely collected by the NMCP of Togo and aggregated by health district and target group; that is, children <5 years old, children ≥5 years old and adults (without pregnant women) and pregnant women. A detailed description of the routine collection of these data was presented elsewhere.²² Data from January 2013 to December 2017 were analysed in this study. A confirmed malaria case was defined as a person with fever or history of fever (temperature ≥38.0°C) over the past 2 days and who had a positive malaria diagnostic test (microscopic diagnosis or rapid diagnostic test).

Meteorological and vegetation data

Meteorological and vegetation data were obtained from the National Aeronautics and Space Administration through the Prediction Of Worldwide Energy Resources (POWER) project and the MODerate resolution Imaging Spectroradiometer (MODIS) satellite. Monthly data of total precipitation, wet days, relative humidity, temperatures (mean, minimum and maximum), wind speeds

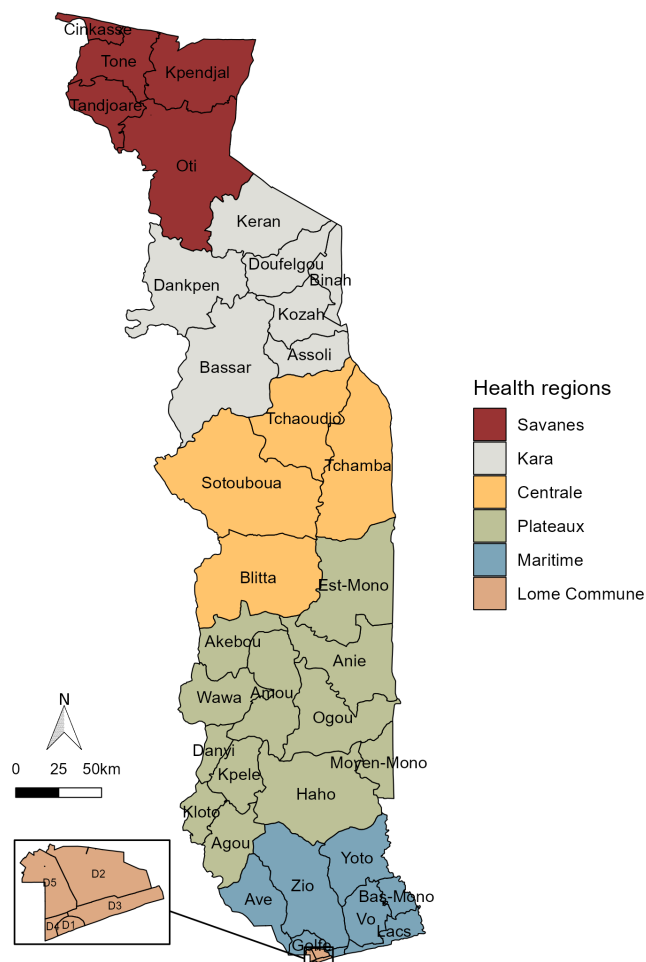


Figure 1 Health map of Togo during the study period, 2013–2017.

(mean, minimum and maximum) and vegetation from November 2012 to November 2017 covering the surface of Togo were extracted at a spatial resolution of 0.5°×0.5° for meteorological data and 1×1 km for vegetation data (table 1).

Statistical analysis

The time series of confirmed malaria cases showed the presence of an artificial peak in September 2016, as described elsewhere.³ A spline interpolation imputation was performed to replace the values of this peak. Two health districts (Kpendjal and Oti) had no monthly confirmed malaria cases reported in 2016, but the total number of cases for that year was not used for analysis. The zonal statistics were used to obtain monthly meteorological and vegetation data by health district from November 2012 to November 2017. The method provides the average of the values of the cells covering a given district, weighted by the proportion of area that a cell represents.²³

Descriptive statistics were used to explore malaria, meteorological and vegetation data by health district and target group. Generalised additive models (GAMs) and generalised additive mixed models (GAMMs) with

Table 1 Description of meteorological and environmental variables

Parameter	Variable name	Description	Unit	Spatial resolution	Source
Total precipitation	PRECTOTCORR_SUM	Monthly corrected sum of total precipitation on the earth's surface	mm	0.5×0.5°	POWER project
Number of rainy days	PREC_NUM	Number of days in a month with rainfall ≥1 mm*	Days	0.5×0.5°	POWER project
Relative humidity	RH2M	Monthly average relative humidity at 2 m above the earth's surface	%	0.5×0.5°	POWER project
Mean temperature	T2M	Monthly average air temperature at 2 m above the earth's surface	°C	0.5×0.5°	POWER project
Minimum temperature	T2M_MIN_AVG	Monthly average of daily minimum air temperature at 2 m above the earth's surface	°C	0.5×0.5°	POWER project
Maximum temperature	T2M_MAX_AVG	Monthly average of daily maximum air temperature at 2 m above the earth's surface	°C	0.5×0.5°	POWER project
Mean wind speed	WS2M	Monthly average wind speed at 2 m above the earth's surface	m/s	0.5×0.5°	POWER project
Minimum wind speed	WS2M_MIN_AVG	Monthly average of daily minimum wind speed at 2 m above the earth's surface	m/s	0.5×0.5°	POWER project
Maximum wind speed	WS2M_MAX_AVG	Monthly average of daily maximum wind speed at 2 m above the earth's surface	m/s	0.5×0.5°	POWER project
Normalised difference vegetation index	NDVI	Monthly normalised difference vegetation index	Scale from -1 to 1 (higher values indicate a higher density of green vegetation)	1×1 km	MODIS satellite

*This variable was not available as such in the POWER project and was calculated using total daily rainfall. MODIS, MODerate resolution Imaging Spectroradiometer; POWER, Prediction Of Worldwide Energy Resources.

a negative binomial error distribution were developed to forecast the monthly number of malaria cases using meteorological and vegetation data with a time lag of 1 or 2 months. A cross-correlation analysis was performed to determine the best predictor lags of the meteorological and environmental variables. Spearman's rank correlation coefficients between malaria data and each predictor variable with a time lag of 1, 2 and 3 months were calculated. The two time lags with the highest correlation coefficient were selected. GAM is a non-parametric regression technique using smooth functions of predictor variables to account for non-linear relationships between the covariates and the variable of interest.²⁴ Choosing the 'best' combination of lagged predictor variables can be very complex. Therefore, two methods for selecting lagged meteorological and environmental variables were compared: a first method based on statistical approach and a second method based on biological reasoning. Both methods were applied to obtain a model per target group and health district and a mixed model per target group and health region with the health district as a random effect. The 'statistical approach' method used in this study was described by Fisher *et al.*²⁵ This method

allows the construction of a complete model set based on a range of potential predictors using GA(M)Ms. Models with correlated predictors above the defined correlation cut-off value are automatically deleted. All models in this set are compared using a model selection criterion. In this study, a list of 20 potential predictors was identified, that is, the ten meteorological and environmental variables with a time lag of 1 and 2 months. The maximum number of predictors included in a model and the correlation cut-off value were set at 5 and 0.28 (default values), respectively. The model with the smallest corrected Akaike information criterion (AICc) was defined as the most accurate model for this first method. The 'biological reasoning' method is an a priori selection of the meteorological variables with a proven effect on the development of *Anopheles* mosquitoes and malaria parasites. Total precipitation with a time lag of 2 months and mean temperature with a time lag of 1 month were included as predictor variables using smoothers. This choice was based on the theoretical vector-parasite-host cycle, which requires a minimum of 38 days from rainfall and deposition of mosquito eggs to the onset of malaria symptoms in the human host under optimal conditions.⁸ Indeed, rainfall is a factor in

the development of larval sites, and temperature affects the longevity of mosquitoes and the development of the parasite.²⁶ For both methods, the dimension of the basis used to represent the smooth term was set at five in order to reduce overfitting, a cubic spline regression for the smooth term was used and the year was included as a centred variable to take into account the trend over time. The models are written as follows:

$$y_t \sim NB(\mu_t, \phi\mu_t)$$

with y_t the number of confirmed malaria cases in month t , μ_t the mean number of cases and $\phi > 1$ the over-dispersion parameter.

For models per district, $\log(\mu_t) = \beta_0 + \beta_1 \text{year} + \sum_{i=1}^k f_i(x_{i(t-1)} \text{ or } (t-2))$ with β_0 the intercept of the model, β_1 the regression coefficient of the year and f_i the cubic spline function for each meteorological and environmental variable x_i included in the models with a time lag equal to $t-1$ or $t-2$.

For mixed models per region, $\log(\mu_{ij}) = \beta_0 + u_{0j} + \beta_1 \text{year} + \sum_{i=1}^k f_i(x_{i(t-1)} \text{ or } (t-2))$ with μ_{ij} the mean number of cases in month t and district j and $u_{0j} \sim N(0, \sigma_0^2)$ the intercept random effect of district j .

The time series for the period 2013–2016 were used for model training, while the 2017 time series were used for model testing. The predictive values and their 95% prediction interval were superimposed on the time series of confirmed malaria cases for graphical analysis. Thus, the predictive skills of the four models were compared for each health district and target group. The AICc and explained deviance were used as measures of accuracy for the training period, and the root mean squared error and mean absolute error (MAE) were used as measures of accuracy for the testing period. The most accurate model for the training period was defined as the model with the highest explained deviance, while the most accurate model for the testing period was defined as the model with the lowest MAE. The prediction errors per month and over the year 2017 (in a number of malaria cases and percentage error) were calculated from the most accurate models for the testing period.

Software and packages

All statistical analyses were performed using R software V.4.1.2.²⁷ The imputation by spline interpolation was performed using the *imputeTS* package.²⁸ Meteorological and vegetation data were extracted using the *nasapower*²⁹ and *MODISsp*³⁰ packages, respectively. The zonal statistics were performed using the *raster* package.³¹ The ‘statistical approach’ method was performed using the *FSSgam* package.²⁵ The GA(M)Ms were generated using the *mgcv* package.²⁴

Patient and public involvement

Neither patients nor the public were involved in the design, conduct, reporting or dissemination plans of this research.

RESULTS

Confirmed malaria cases

From 2013 to 2017, 5 522 650 confirmed malaria cases were reported in Togo (online supplemental table S1). Children <5 years old, children ≥5 years old and adults, and pregnant women represented 36.6%, 58.5% and 4.9% of the total number of confirmed malaria cases, respectively. The Plateaux region had the highest number of confirmed malaria cases (1 512 129, 27.4%), and the Lome-Commune region had the lowest number of confirmed malaria cases (330 410, 6.0%).

Meteorological and vegetation data

Descriptive statistics of monthly meteorological and vegetation data from November 2012 to November 2017 in Togo are presented in online supplemental table S2. One rainy season was observed in the northern health districts from April to October, and two rainy seasons were observed in the southern health districts from March to June and September–October, approximatively. Total precipitation was higher in the north than in the south. The median precipitation during the wettest month ranged from 129 mm in June in the Ave district to 283 mm in August in the Oti district (online supplemental figure S1). From the Cinkasse to the Kpele districts, the number of rainy days peaked in August with a median number of rainy days ranging from 28 to 31. From the Haho district to District 4, the number of rainy days peaked in May, June or October with a median number of rainy days ranging from 26 to 28 (online supplemental figure S2). In the northern health districts, relative humidity was lower during the Harmattan period (from December to February) and increased during the rainy season. For example, the Cinkasse district had a median relative humidity of 23% in February and 84% in August. In the southern health districts, the median relative humidity was above 60% in all months of the year and increased during the rainy season (online supplemental figure S3). The widest temperature variation was observed in the northern health districts during the Harmattan period, ranging from a median minimum temperature of 16–19°C to a median maximum temperature of 34–36°C in January, and the temperature increased in February and March just before the rainy season (median maximum temperature of 35–39°C). In the southern health districts, the temperature varied from a median minimum temperature of 19–23°C to a median maximum temperature of 31–35°C throughout the year. Temperature decreased in all health districts of Togo during the rainy season (online supplemental figure S4). Wind speed was higher in the northern health districts during the Harmattan period and reached up to 5.2 m/s in median in February

in the Cinkasse district. The centre of Togo was the least windy area while in the coastal areas had a median value of mean wind speed above 2.1 m/s in all months of the year (online supplemental figure S5). Northern Togo had very little vegetation during the dry season with a median NDVI of 0.2 in March in the Cinkasse, Tone and Kpendjal districts. NDVI increased in all health districts of Togo during the rainy season, reaching a median of 0.78 in October in the Wawa district. Vegetation was densest in central Togo. A low vegetation index was observed in the health district of the Lome-Commune region throughout the year (<0.41) (online supplemental figure S6).

Comparison of the predictive skills of the models

In the three target groups, the selection of meteorological and environmental variables using the 'statistical approach' method did not follow any particular pattern (online supplemental figure S7). In children <5 years old,

the most selected predictor in the models per district was NDVI, included in 35.0% of the models. In children ≥ 5 years old and adults, the most selected predictor in the models per district was minimum temperature, included in 42.5% of the models. In pregnant women, the most selected predictor in the models per district was mean temperature, included in 30.0% of the models. Minimum wind speed and maximum temperature were the least selected predictors in the models for all three groups ($\leq 5.0\%$ and $\leq 7.5\%$, respectively) (online supplemental figure S7). The 'statistical approach' method provided the most accurate models for the training period, except for the Est mono and Danyi districts in children ≥ 5 years old and adults and for the health districts of the Plateaux region in pregnant women (figure 2). The explained deviance of the most accurate models for the training period ranged from 41.4 to 81.9% in children <5 years old, from

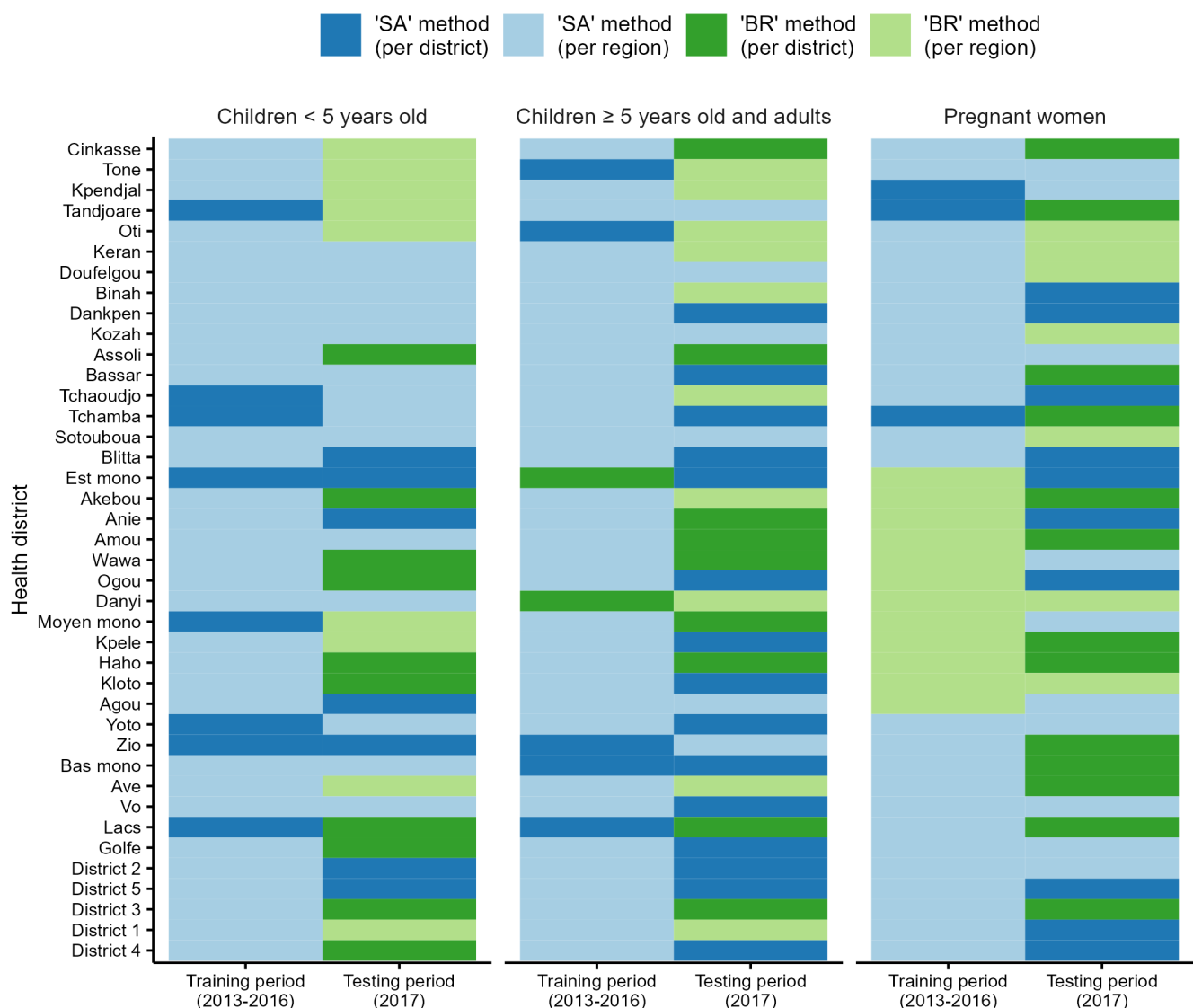


Figure 2 Most accurate models for the training (2013–2016) and testing (2017) periods for each health district and target group. Health districts were classified according to their latitude. BR, biological reasoning; SA, statistical approach.

45.8 to 78.3% in children ≥ 5 years old and adults and from 50.3 to 81.3% in pregnant women (online supplemental figure S3). The ‘statistical approach’ method provided the most accurate models for the testing period in 21 health districts for children < 5 years old, in 21 health districts for children ≥ 5 years old and adults and in 20 health districts for pregnant women (figure 2). The graphical analysis of the observed and predicted malaria time series revealed that the prediction in 2017 was not adequate despite a satisfactory model construction, in particular with the models per district for the ‘statistical approach’ method (figure 3 for children < 5 years old, online supplemental figure S8 for children ≥ 5 years old and adults and online supplemental figure S9 for pregnant women). The difference between the numbers of predicted and observed cases during the 2017 testing period varied between health districts and target groups (figure 4, online supplemental figure S10). In children < 5 years old, the difference between the numbers of predicted and observed cases ranged from -2709 to 2421 cases and the percentage error ranged from -22.2 to 43.4% . In children ≥ 5 years old and adults, this difference ranged from -2717 to 8859 cases and the percentage error ranged from -15.4 to 74.7% . In pregnant women, this difference ranged from -666 to 767 cases and the percentage error ranged from -33.5 to 112.1% .

DISCUSSION

This study aimed to develop malaria prediction models by health district and target group in order to use them as an additional tool in malaria control activities in Togo. This is the first study in Togo that explored meteorological and environmental predictors for forecasting malaria cases all over the country. Despite the development of models with four different approaches and their application on 120 time series, the number of malaria cases was inaccurately forecasted.

The graphical analysis showed that the values predicted during the training period (2013–2016) by the models per district for the ‘statistical approach’ method followed the pattern of the time series of confirmed malaria cases in most health districts and target groups (figure 3, online supplemental figure S8 and S9). This confirms a strong association between meteorological and environmental data and malaria cases in Togo, as shown in other studies in sub-Saharan Africa.^{4–10} NDVI, minimum temperature and mean temperature were the most selected predictors in the models per district. Reiner *et al* reported that minimum temperature, rainfall and vegetation indices were often considered as significant predictors of malaria seasonality in the scientific literature.¹¹ In addition, the selected meteorological and environmental variables and their time lag differed between health districts and target groups (online supplemental figure S7). Other studies showed spatial and temporal variations.^{5 7 11} For example, in Uganda, time lags observed between environmental factors (rainfall, temperature and vegetation) and the

number of malaria cases varied by transmission setting (low, moderate or high malaria transmission).⁷ In Mali, Ateba *et al* showed that the associations between environmental factors and malaria incidence differed between the two ecological zones studied.⁵

In this study, two methods of variable selection were compared based on the following reasoning. Can we predict in the same way using an a priori variable selection method (the ‘biological reasoning’ method) or a more complex variable selection method (the ‘statistical approach’ method)? In the light of the results, the answer to this question is not obvious. On one hand, the ‘statistical approach’ method seemed to provide more accurate models for the training period but perhaps with an overfitting that led to inaccurate forecasts for the testing period in some health districts.³² On the other hand, the ‘biological reasoning’ method did not appear to provide better prediction results for all health districts and target groups.

Different statistical methods can be used to forecast malaria incidence such as generalised linear models (GLMs), autoregressive integrated moving average (ARIMA) models or Holt-Winters models.³³ The GAM is an extension of the conventional GLM and is used for modelling non-linear relationships between the covariates and the variable of interest.²⁴ In this study, GAMs were chosen for their flexibility of use compared with ARIMA models, given the large number of models to be generated. In addition, the prediction method had to be easily reproducible and implementable in a decision support software. Two studies conducted in Burkina Faso and Kenya also used GAMs for malaria forecasting.^{6 15} The use of GAMs gave the possibility to apply another method of predictor variable selection based on the full-subset information theoretic approaches using the *FSSgam* package in R. This method described by Fisher *et al* is presented as an alternative to data reduction techniques and stepwise regression and avoids the problems of multicollinearity.²⁵

Climate-related predictors are frequently used in malaria prediction models.^{6 15–19 33} Temperature, precipitation and humidity are often cited in these studies, but other predictors such as vegetation index, wind speed, evapotranspiration, atmospheric pressure, cloud cover, visibility or altitude are also used. The availability of and access to quality climate data in Africa are a real challenge. The weather stations are not evenly distributed across the countries and data are of poor quality or missing.³⁴ Real-time access to climate data must also be guaranteed for prediction. Open access to historical and current satellite-derived climate data, such as that made available by the POWER project and the MODIS satellite, is a major advantage for the development of MEWSs in Africa. This study showed that it was possible to use meteorological and environmental data in malaria research in Togo and should encourage other researchers working on malaria or other vector-borne diseases to use them.

It was decided to develop malaria prediction models according to the three target groups, given the different

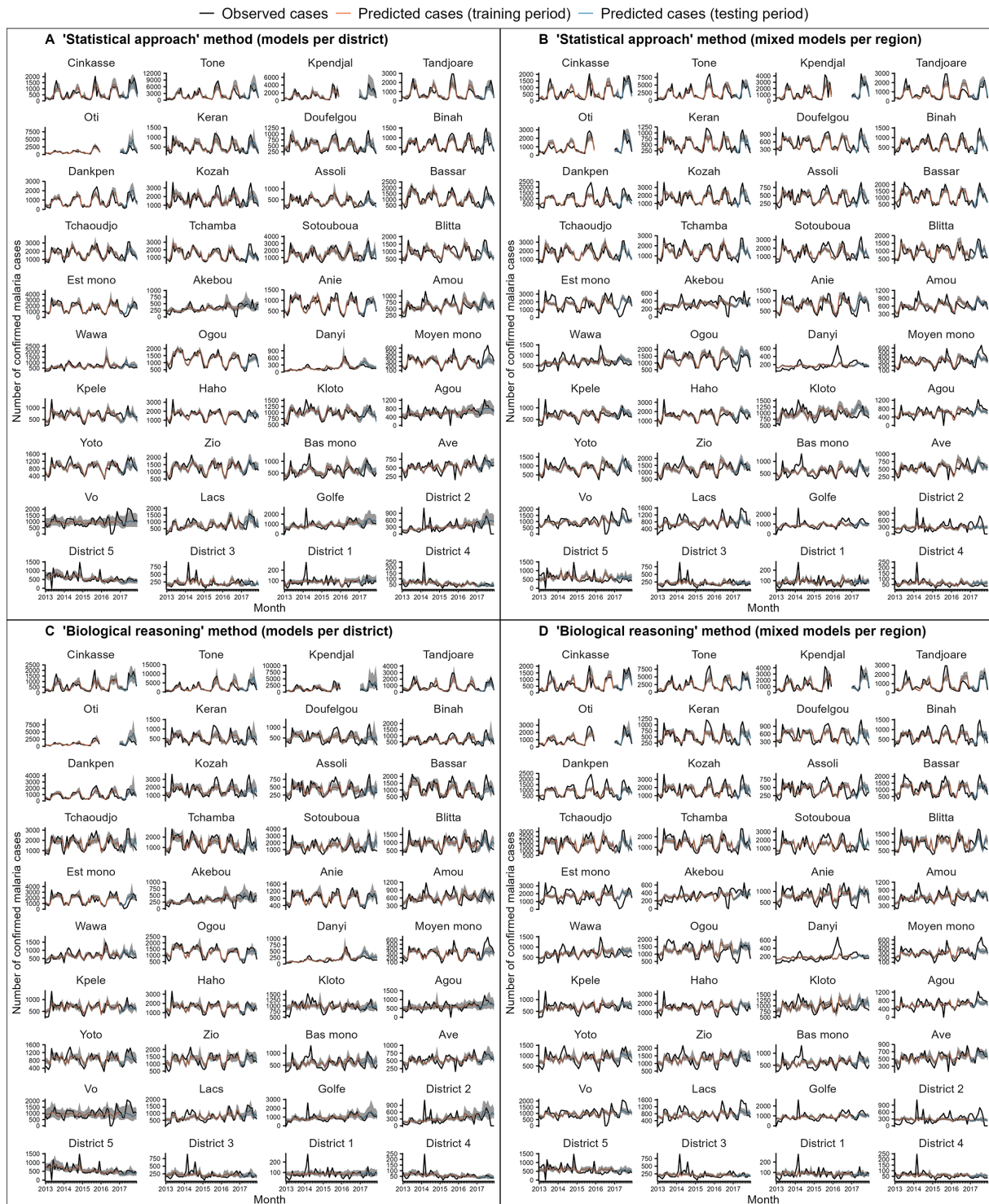


Figure 3 Time series of observed (black lines) and predicted (red lines for the training period and blue lines for the testing period) malaria cases and their 95% prediction interval in children <5 years old by health district using the models per district (A) and mixed models per region (B) for the ‘statistical approach’ method and the models per district (C) and mixed models per region (D) for the ‘biological reasoning’ method. For each panel, at the top left is the northernmost district and at the bottom right is the southernmost district. Time series have different Y-axis scales.

levels of immunity, risks of complication, specific prevention and treatment responses in children, pregnant women or adults.^{35 36} Mfisimana *et al* also adopted this strategy in their study in Burundi by developing malaria prediction models in four subgroups (children between

0 and 11 months, children between 12 and 59 months, pregnant women, and pregnant women and children under 5 years) and in the overall general population.¹⁸

This analysis has some methodological limitations. First, the performance of the models was tested over a

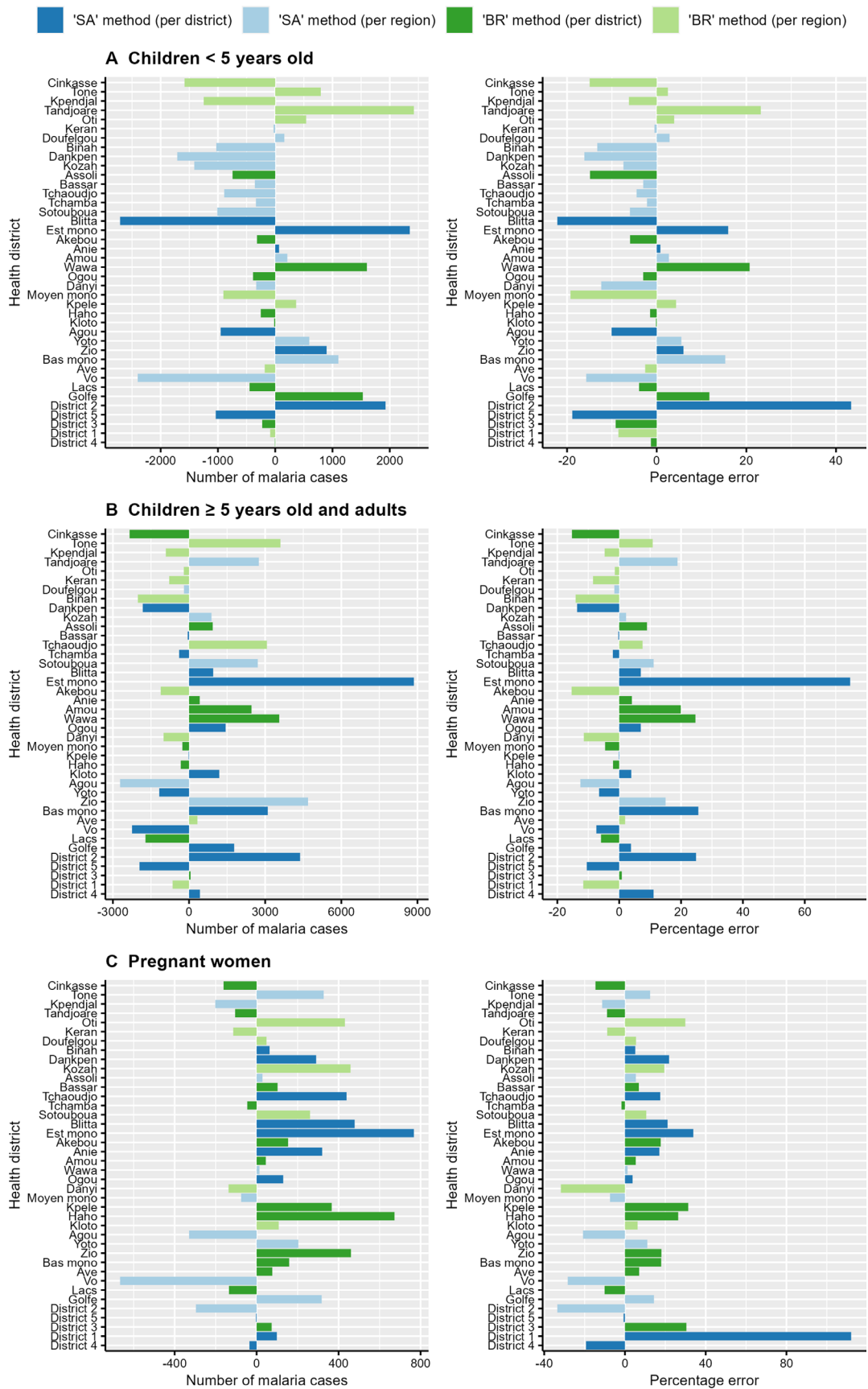


Figure 4 Difference between predicted and observed malaria cases during the 2017 testing period for each health district in children <5 years old (A), in children ≥5 years old and adults (B) and in pregnant women (C). The number of cases is shown on the left and the percentage error is shown on the right. Health districts were classified according to their latitude. BR, biological reasoning; SA, statistical approach.

single year. However, data after 2017 were not available. Second, the population size was not included as an offset term in the malaria prediction models because the distribution of the population by target group was not known. This may have an impact on the quality of predictions.

What could be done to improve malaria forecasting? First, the use of finer spatial and temporal scales, such as the weekly number of malaria cases at facility level, could improve malaria prediction. Cottrell *et al* even showed spatial variations in *Anopheles* density at the scale of villages and houses, explained by the house's immediate surroundings such as proximity to a watercourse, soil type and vegetation index (NDVI).³⁷ However, routine data collection in Togo is not organised in this way. It might be interesting to use the malaria sentinel surveillance data, collected weekly in 16 health facilities in Togo,³⁸ to develop climate-driven malaria models in a future study. Second, meteorological and environmental factors do not fully explain the seasonality of malaria. For example, socioeconomic factors play also an important role in malaria incidence. In a systematic review and meta-analysis published in 2013, Tusting *et al* showed that economic status such as asset ownership, household wealth, socioeconomic index and parents' occupations was associated with the risk of malaria in children aged 0–15 years.³⁹ Some authors suggest exploring the predictive capacity of other geographic or environmental data (eg, land cover and altitude) and non-environmental data such as intervention data (eg, indoor residual spraying and insecticide-treated nets) or clinical data (eg, diagnostic and treatment) to improve the accuracy of prediction models.^{8 16 33} The challenge remains the availability and access to this information.

This study shows that malaria forecasting is a complex exercise. Seasonal variations of malaria are not only related to meteorological and environmental conditions but are a more complex process. Care should be taken when using routine information system data to anticipate malaria cases. Many challenges remain to integrate malaria prediction models into malaria control strategies to support decision-making in Togo. We encourage the scientific community and national stakeholders to engage in this research.

CONCLUSIONS

In this study, malaria prediction models were developed for each health district and target group in Togo using routinely collected malaria data and meteorological and environmental predictors. Despite the different approaches used for building these models, the predictions were inaccurate, suggesting that these models should be avoided in malaria control activities in Togo.

Acknowledgements Meteorological data used in this study were obtained from the NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science/Applied Science Program.

Contributors AT, TAB, MR and NV designed the study. TAB, TA-O and TT collected the data. AT analysed the data. AT, MR and NV interpreted the results. AT, MR and

NV wrote the manuscript. All authors read and approved the final version of the manuscript. NV is responsible for the overall content as guarantor.

Funding This work was supported by the EPIMOD company, the French Association Nationale Recherche Technologie (CIFRE grant number 2017/1805) and the Auvergne-Rhône-Alpes region (grant number 19 003721 01).

Map disclaimer The inclusion of any map (including the depiction of any boundaries there), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The study was approved and authorised by the Ministry of Health of Togo (Memorandum N°280/17/MSPS/CAB/SG/DGEPIS/DEPP/DER). Individual data were not collected.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. The data analysed here are the property of the Togolese Ministry of Health and cannot be made available by the authors. To access the data, interested parties can get information from the Ministry of Health of Togo.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Anne Thomas <http://orcid.org/0000-0002-3084-6192>

REFERENCES

- 1 WHO. *World malaria report 2021*. Geneva: World Health Organization, 2021.
- 2 Programme National de Lutte contre le Paludisme. *Rapport annuel 2020*. Lomé: PNLP, 2021.
- 3 Thomas A, Bakai TA, Atcha-Oubou T, *et al*. Seasonality of confirmed malaria cases from 2008 to 2017 in Togo: a time series analysis by health district and target group. *BMC Infect Dis* 2021;21:1189.
- 4 Adeola A, Ncongwane K, Abiodun G, *et al*. Rainfall trends and malaria occurrences in Limpopo Province, South Africa. *Int J Environ Res Public Health* 2019;16:5156.
- 5 Ateba FF, Sagara I, Sogoba N, *et al*. Spatio-temporal dynamic of malaria incidence: a comparison of two ecological zones in Mali. *Int J Environ Res Public Health* 2020;17:4698.
- 6 Bationo CS, Gaudart J, Dieng S, *et al*. Spatio-temporal analysis and prediction of malaria cases using remote sensing meteorological data in Diébougou health district, Burkina Faso, 2016-2017. *Sci Rep* 2021;11:20027.
- 7 Kigozi R, Zinszer K, Mpimbaza A, *et al*. Assessing temporal associations between environmental factors and malaria morbidity at varying transmission settings in Uganda. *Malar J* 2016;15:511.
- 8 Krefis AC, Schwarz NG, Krüger A, *et al*. Modeling the relationship between precipitation and malaria incidence in children from a holoendemic area in Ghana. *Am J Trop Med Hyg* 2011;84:285–91.
- 9 Ouedraogo B, Inoue Y, Kambiré A, *et al*. Spatio-temporal dynamic of malaria in Ouagadougou, Burkina Faso, 2011-2015. *Malar J* 2018;17:138.

- 10 Amadi JA, Olago DO, Ong'amo GO, *et al.* Sensitivity of vegetation to climate variability and its implications for malaria risk in Baringo, Kenya. *PLoS One* 2018;13:e0199357.
- 11 Reiner RC, Geary M, Atkinson PM, *et al.* Seasonality of plasmodium falciparum transmission: a systematic review. *Malar J* 2015;14:343.
- 12 Global partnership to roll back malaria. A framework for field research in Africa: malaria early warning systems: concepts, indicators and partners [World Health Organization]. 2001. Available: <https://apps.who.int/iris/handle/10665/66848>
- 13 Cox J, Abeku TA. Early warning systems for malaria in Africa: from blueprint to practice. *Trends Parasitol* 2007;23:243–6.
- 14 Landier J, Rebaudet S, Piarroux R, *et al.* Spatiotemporal analysis of malaria for new sustainable control strategies. *BMC Med* 2018;16:226.
- 15 Sewe MO, Tozan Y, Ahlm C, *et al.* Using remote sensing environmental data to forecast malaria incidence at a rural district hospital in Western Kenya. *Sci Rep* 2017;7:2589.
- 16 Zinszer K, Kigozi R, Charland K, *et al.* Forecasting malaria in a highly endemic country using environmental and clinical predictors. *Malar J* 2015;14:245.
- 17 Ferrão JL, Mendes JM, Painho M. Modelling the influence of climate on malaria occurrence in Chimoio municipality, Mozambique. *Parasit Vectors* 2017;10:260.
- 18 Mfisismana LD, Nibayisabe E, Badu K, *et al.* Exploring predictive frameworks for malaria in Burundi. *Infect Dis Model* 2022;7:33–44.
- 19 Kim Y, Ratnam JV, Doi T, *et al.* Malaria predictions based on seasonal climate forecasts in South Africa: a time series distributed lag nonlinear model. *Sci Rep* 2019;9:17882.
- 20 Central Intelligence Agency. Togo. The World Factbook. 2022. Available: <https://www.cia.gov/the-world-factbook/countries/togo/#geography> [Accessed 02 Mar 2022].
- 21 World Bank. Population, total - Togo. 2022. Available: <https://data.worldbank.org/indicator/SP.POP.TOTL?locations=TG> [Accessed 02 Mar 2022].
- 22 Bakai TA, Thomas A, Iwaz J, *et al.* Changes in registered malaria cases and deaths in Togo from 2008 to 2017. *Int J Infect Dis* 2020;101:298–305.
- 23 RDocumentation. extract: Extract values from Raster objects. Available: <https://www.rdocumentation.org/packages/raster/versions/3.5-15/topics/extract> [Accessed 06 Jun 2022].
- 24 Wood SN. *Generalized additive models: an introduction with R, 2nd.* New York: Chapman and Hall/CRC, 2017.
- 25 Fisher R, Wilson SK, Sin TM, *et al.* A simple function for full-subsets multiple regression in ecology with R. *Ecol Evol* 2018;8:6104–13.
- 26 Centers for Disease Control and Prevention. About malaria | biology. 2020. Available: <https://www.cdc.gov/malaria/about/biology/index.html> [Accessed 07 Mar 2022].
- 27 R Core Team. R: a language and environment for statistical computing. Vienna, Austria R Foundation for Statistical Computing; 2021. Available: <https://www.r-project.org/> [Accessed 03 Mar 2022].
- 28 Moritz S, Bartz-Beielstein T. imputeTS: Time Series Missing Value Imputation in R. *R Journal* 2017;9:207.
- 29 Sparks AH. nasapower: A NASA power global meteorology, surface solar energy and climatology data client for R. *JOSS* 2018;3:1035.
- 30 Busetto L, Ranghetti L. MODISr: An R package for automatic preprocessing of MODIS Land Products time series. *Comput Geosci* 2016;97:40–8.
- 31 Hijmans RJ. Raster: geographic data analysis and modeling. Available: <https://CRAN.R-project.org/package=raster> [Accessed 23 Jan 2023].
- 32 Shmueli G. To explain or to predict? *Statist Sci* 2010;25:289–310.
- 33 Zinszer K, Verma AD, Charland K, *et al.* A scoping review of malaria forecasting: past work and future directions. *BMJ Open* 2012;2:e001992.
- 34 Dinku T. Chapter 7 - challenges with availability and quality of climate data in Africa. In: Melesse AM, Abtew W, Senay G, eds. *Extreme Hydrology and Climate Variability*. Elsevier, 2019: 71–80.
- 35 Ashley EA, Pyae Phyo A, Woodrow CJ. Malaria. *Lancet* 2018;391:1608–21.
- 36 Nkumama IN, O'Meara WP, Osier FHA. Changes in malaria epidemiology in Africa and new challenges for elimination. *Trends Parasitol* 2017;33:128–40.
- 37 Cottrell G, Kouwaye B, Pierrat C, *et al.* Modeling the influence of local environmental factors on malaria transmission in Benin and its implications for cohort study. *PLoS One* 2012;7:e28812.
- 38 Thomas A, Bakai TA, Atcha-Oubou T, *et al.* Implementation of a malaria sentinel surveillance system in Togo: a pilot study. *Malar J* 2020;19:330.
- 39 Tusting LS, Willey B, Lucas H, *et al.* Socioeconomic development as an intervention against malaria: a systematic review and meta-analysis. *Lancet* 2013;382:963–72.