scientific reports

Check for updates

OPEN

Spatio-temporal analysis and prediction of malaria cases using remote sensing meteorological data in Diébougou health district, Burkina Faso, 2016–2017

Cédric S. Bationo[1,2,6], Jean Gaudart[3,4]✉, Sokhna Dieng[1], Mady Cissoko[1,4], Paul Taconet[2,6], Boukary Ouedraogo[5], Anthony Somé[6], Issaka Zongo[6], Dieudonné D. Soma[2,6,7], Gauthier Tougri[8], Roch K. Dabiré[6], Alphonsine Koffi[9], Cédric Pennetier[2,6,9] & Nicolas Moiroux[2,6]

Malaria control and prevention programs are more efficient and cost-effective when they target hotspots or select the best periods of year to implement interventions. This study aimed to identify the spatial distribution of malaria hotspots at the village level in Diébougou health district, Burkina Faso, and to model the temporal dynamics of malaria cases as a function of meteorological conditions and of the distance between villages and health centres (HCs). Case data for 27 villages were collected in 13 HCs. Meteorological data were obtained through remote sensing. Two synthetic meteorological indicators (SMIs) were created to summarize meteorological variables. Spatial hotspots were detected using the Kulldorf scanning method. A General Additive Model was used to determine the time lag between cases and SMIs and to evaluate the effect of SMIs and distance to HC on the temporal evolution of malaria cases. The multivariate model was fitted with data from the epidemic year to predict the number of cases in the following outbreak. Overall, the incidence rate in the area was 429.13 cases per 1000 person-year with important spatial and temporal heterogeneities. Four spatial hotspots, involving 7 of the 27 villages, were detected, for an incidence rate of 854.02 cases per 1000 person-year. The hotspot with the highest risk (relative risk = 4.06) consisted of a single village, with an incidence rate of 1750.75 cases per 1000 person-years. The multivariate analysis found greater variability in incidence between HCs than between villages linked to the same HC. The time lag that generated the better predictions of cases was 9 weeks for SMI1 (positively correlated with precipitation variables) and 16 weeks for SMI2 (positively correlated with temperature variables. The prediction followed the overall pattern of the time series of reported cases and predicted the onset of the following outbreak with a precision of less than 3 weeks. This analysis of malaria cases in Diébougou health district, Burkina Faso, provides a powerful prospective method for identifying

[1]INSERM, IRD, SESSTIM, UMR1252, Institute of Public Health Sciences, ISSPAM, Aix Marseille Univ, 13005 Marseille, France. [2]CNRS, IRD, MIVEGEC, Univ. Montpellier, Montpellier, France. [3]INSERM, IRD, SESSTIM, UMR1252, Institute of Public Health Sciences, ISSPAM, APHM, Hop Timone, BioSTIC, Biostatistic & ICT, Aix Marseille Univ, 13005 Marseille, France. [4]Malaria Research and Training Center—Ogobara K. Doumbo (MRTC-OKD), FMOS-FAPH, Mali-NIAID-ICER, Université des Sciences, des Techniques et des Technologies de Bamako, Bamako 1805, Mali. [5]Direction des Systèmes d'information en Santé, Ministère de la Santé du Burkina Faso, Ouagadougou, Burkina Faso. [6]Institut de Recherche en Sciences de la Santé (IRSS), Bobo Dioulasso, Burkina Faso. [7]Institut Supérieur des Sciences de la Santé, Université Nazi Boni, Bobo-Dioulasso, Burkina Faso. [8]Programme National de Lutte Contre le Paludisme, Ministère de la Santé du Burkina Faso, Ouagadougou, Burkina Faso. [9]Institut Pierre Richet (IPR), Institut National de Santé Publique (INSP), Bouaké, Côte d'Ivoire. ✉email: jean.gaudart@univ-amu.fr

and predicting high-risk areas and high-transmission periods that could be targeted in future malaria control and prevention campaigns.

Abbreviations

WHO	World Health Organization
CI	Confidence interval
IPT	Intermittent preventive treatment
RDT	Rapid diagnostic test
SMC	Seasonal malaria chemoprevention
LLIN	Long-lasting insecticidal net
IRS	Indoor residual spraying
HC	Health Centre
SMI	Synthetic meteorological indicator
PCA	Principal component analysis
GAM	Generalized additive model
GLM	Generalized linear model
UBRE	Unbiased risk estimator
GAMM	Generalized additive mixed model
SIR	Standardized incidence ratio
RR	Relative risk
SD	Standard deviation

Malaria is one of the most life-threatening diseases and poses a great socio-economic burden worldwide[1]. According to World Health Organization (WHO) estimates, the global number of malaria cases was 229 million in 2019 compared to 251 million in 2010 and 214 million in 2015[1]. Although the estimated number of cases decreased by 23 million from 2010 to 2018, data for the period 2015–2018 highlight the lack of significant progress during this period. In 2018, the WHO African Region accounted for most cases (200 million or 93% of all cases), far ahead of the WHO South-East Asian region (3.4%) and the WHO Eastern Mediterranean Region (2.1%)[1]. At the time, nearly 80% of global malaria deaths were concentrated in 17 countries of the WHO African Region and in India. The WHO estimates that Burkina Faso carries about 6% of the global malaria burden[1]. Statistical data from the Ministry of Health of Burkina Faso for the year 2018 show that malaria is the second reason for consultation (31.7%), and that pregnant women and children under 5 years are the most at risk of contracting malaria[2]. According to those data, the average parasite prevalence in children under 5 years was 17% in 2017–2018[2,3]. In 2018, the number of confirmed cases reported in health facilities was 11,624,595 of which 4.14% were severe forms and 2.8% resulted in death.

The National Malaria Control Program in Burkina Faso recommends the following control strategies[4]: early case management in health facilities and at the community level, with a particular focus on children aged 3 to 59 months[5]; intermittent preventive treatment (IPT) for pregnant women; universal access to rapid diagnostic tests (RDTs) and artemisinin-based combination therapies; seasonal malaria chemoprevention (SMC) for children under 5 years; and vector control using long-lasting insecticidal nets (LLINs), indoor residual spraying (IRS), larval control, and environmental sanitation.

For strategic reasons or lack of resources, not all of these strategies are optimally implemented everywhere and all the time. Thus, in 2018, 25% of households reported not owning an LLIN (with coverage varying between 58 and 87% depending on the region) and 42% of pregnant women did not receive the recommended three doses of IPT, as reported by the Burkina Faso Malaria Indicator Survey[6].

At the same time, new tools and strategies are being developed, including administration of ivermectin, bi-impregnated nets, transmission-blocking vaccines, and conventional vaccines[4,7,8]. In Burkina Faso, the REACT project ("Insecticide resistance management in Burkina Faso and Côte d'Ivoire: A study on vector control strategies") conducted in 2016–2018 aimed to evaluate the efficacy of strategies designed to complement LLINs, namely pirimiphos methyl-based IRS, enhanced communication, and administration of ivermectin to domestic animals.

Malaria control and prevention programs are more efficient and cost-effective when they target high-risk spatial clusters (hotspots)[9] or when they select the best times of year[10] to initiate interventions (e.g. SMC or LLIN distribution). Indeed, as numerous studies have shown, malaria incidence at the local level is heterogeneous and associated with spatio-temporal clusters[11–13] that are likely to maintain transmission during low-risk periods and, consequently, to increase transmission during high-risk periods[14–16]. Identifying these clusters can therefore help to improve the fight against malaria and to anticipate future outbreaks.

This study aimed to identify the spatial distribution of malaria hotspots at the village level in Diébougou health district, Burkina Faso, and to model the temporal dynamics of malaria cases as a function of meteorological conditions and of the Euclidean distance between villages and their corresponding health centres (HCs). Data on malaria cases were obtained through HC-based passive case detection for the 27 villages included in the REACT project.

Materials and methods

Study area. The study was conducted in 27 villages of Diébougou health district that were included in the REACT project. All included villages met two criteria: a population between 200 and 500 inhabitants and a Euclidean distance of at least 2 km from the nearest village. A population census carried out in July 2016 by our

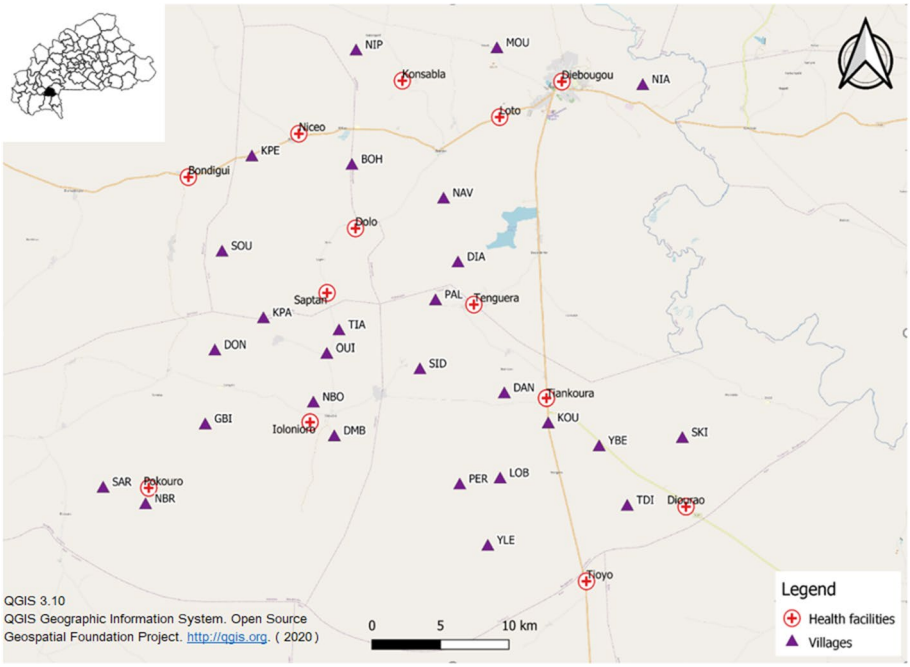

Figure 1. Map of the study area showing the location of villages (triangles) and health centres (red crosses). Background: OpenStreetMaps. *MOU* moule, *NIA* niaba, *NIP* nipodja, *BOH* bohero, *NBR* niombripo, *SAR* sarambour, *NBO* niombouna, *GBI* gongombiro, *KPA* kpalbalo, *DON* dontelo, *SID* sidmoukar; *DMB* dombouro, *OUI* ouidiaro, *TIA* tiakiro, *NAV* nouvielgane, *DIA* diagnon, *PAL* palembera, *KPE* kpedia, *SOU* soussoubro, *TDI* tordiero, *YLE* yellela, *YBE* yelbelela, *SKI* sinkiro, *DAN* dangbara, *KOU* kouloh, *LOB* lobignonao, *PER* perglembiro.

research team found that the 27 villages were home to 7408 inhabitants. The villages were linked to 13 HCs. Villages and HCs were geo-referenced using *Global Positioning System* (GPS) (Fig. 1).

Diébougou health district is located in South-Western Burkina Faso, a region characterized by a tropical climate with a dry season from October to April and a rainy season from May to September. The dry season is divided into a cold dry season lasting from December to February and a hot dry season lasting from March to April. Average daily minimum and maximum temperatures in the cold dry, hot dry, and rainy seasons are 18 and 36 °C, 25 and 39 °C, and 23 and 33 °C, respectively. Average annual rainfall is 1200 mm. The natural vegetation is dominated by wooded savannah dotted with clear forest gallery[17,18]. Let's remind that Burkina Faso is spread over 3 climatic zones: in the north, (Sahelian zone), rainfall is less than 600 mm/year. While the centre (or northern Sudanese) zone receives 600–900 mm/year, rainfall in the southern (or southern Sudanese, where Diebougou is located) zone exceeds 900 mm/year.

Passive case detection. Case data for the 27 villages included in the REACT project were collected using continuous HC-based passive case detection during 2016 and the first 36 weeks of 2017, which corresponded to the period preceding the implementation of the interventions studied (i.e. pirimiphos methyl-based IRS, enhanced communication, and administration of ivermectin to domestic animals). Specifically, consultation data for village residents were retrieved from each HC registries and recorded using tablets equipped with Open Data Kit collect forms. A malaria case was defined as a person who presented with fever and had a positive RDT result.

Study period. Of the 88 weeks of data collection, 52 weeks corresponding to an epidemic year (a complete malaria epidemic) were considered for spatio-temporal analysis. The epidemic year ran from week 20 (in May) of 2016 to week 19 (in May) of 2017 (Fig. 2).

Meteorological data. The meteorological data used in this study were drawn from the Era-5 dataset from 2016 to 2017[19] published by the European Centre for Medium-Range Weather Forecasts, which provides hourly estimates of several atmospheric and land parameters at a spatial resolution of 0.25°[20]. These data were aggregated into weekly counts. The meteorological variables included in the analysis were: Weekly rainfall (mm), number of rainy days per week, weekly mean of daily average temperature (°C), weekly mean of daily minimum temperature (°C), weekly mean of daily maximum temperature (°C), weekly mean of daily average wind speed (km/h), weekly mean of daily average relative humidity (%), weekly mean of daily average atmospheric pressure (hPa), weekly mean of daily average cloud cover (%), and weekly mean of daily thermal amplitude (°C) (Table 1).

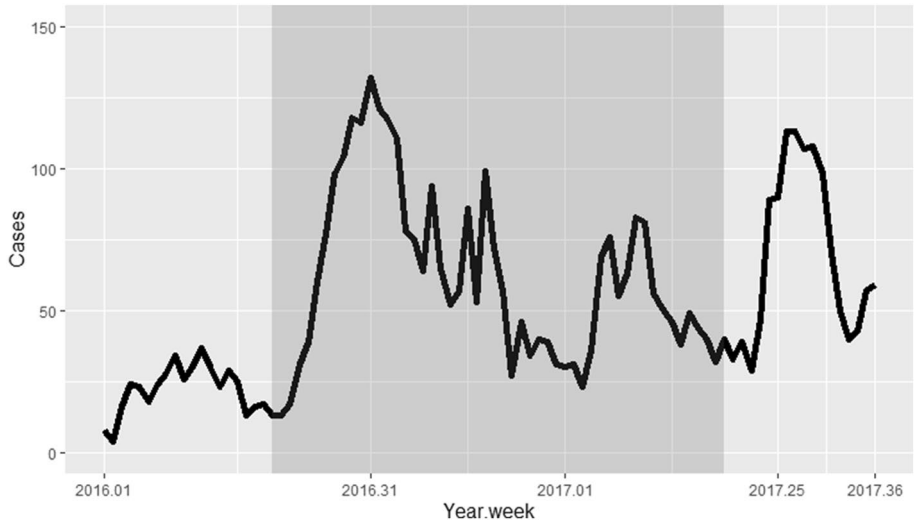

Figure 2. Time series of the number of malaria cases collected through passive case detection. The shaded (dark) area represents the epidemic year considered for analysis.

Var.	Mean	St. dev	Min	Pctl (25)	Median	Pctl (75)	Max
Tmean: weekly mean of daily average temperature (° C)	28.7	2.2	25.7	26.6	28.8	30.2	33.9
Re: number of rainy days per week	4.8	2.9	0	2	7	7	7
P: weekly mean of daily average atmospheric pressure (hPa)	975.3	1.6	972.3	974.0	975.5	976.7	978
Cl: weekly mean of daily average cloud cover (%)	0.6	0.2	0.1	0.4	0.6	0.7	0.9
W: weekly mean of daily average wind speed (km/h)	7.6	3.6	2.8	4.7	7.2	9.5	17.4
H: weekly mean of daily average relative humidity (%)	0.5	0.2	0.1	0.3	0.6	0.8	0.8
Tmax: weekly mean of daily maximum temperature (°C)	34.4	3.5	28.6	31.2	35.4	36.6	40.7
Tmin: weekly mean of daily minimum temperature (°C)	23.3	1.8	19.1	22.4	23.0	24.2	27.0
Rc: weekly rainfall (mm)	14.4	18.3	0.0	0.005	5.0	24.5	69.8
tvar: weekly mean of daily thermal amplitude (°C)	11.1	3.4	5.9	7.7	10.6	14.4	17.1

Table 1. List of meteorological variables with their abbreviations and descriptive statistics. *Var* variables, *St. Dev* standard deviation, *Min* minimum, *Pctl(25)* first quartile, *Pctl(75)* third quartile, *Max* maximum.

Data analysis and statistical methods. *Meteorological data.* To reduce the number of variables and avoid collinearity, we constructed synthetic meteorological indicators (SMIs) using a principal component analysis (PCA) of weekly meteorological variables. Different cofactors, affected by collinearity (such as rainfall, vegetation and temperature), are known to impact the parasitological cycle at different steps. Using PCA allowed keeping the main environmental characteristics without losing part of the environmental cofactors associated with malaria. Principal components that met Kaiser's criterion[21] were selected as SMIs and included in the temporal analysis.

Spatial analyses. Hotspots, i.e. high-risk clusters, were detected using the Kulldorf scanning method[22] with a Monte Carlo algorithm in a purely spatial analysis. The Kulldorf scanning method helps to identify spatial clusters based on geographical coordinates and to avoid the problem of multiple non-independent tests[22]. We defined clusters as aggregates of cases with observed values higher than expected (i.e. unlikely to have been obtained by chance). The p-value (i.e. the probability, under the null hypothesis, that the expected number of cases is the same or higher than the observed number of cases) was calculated for each cluster.

Scan parameters were: elliptical window, non-overlapping clusters, maximum cluster size set at 50% of the population at risk, Monte Carlo replication number set at 999.

Temporal analyses. Lagged SMI selection. Several studies have observed a lag between malaria time series and meteorological data time series[14–16]. In view of this, we decided to investigate the time lag (in weeks) between the time series of weekly malaria cases and the time series of SMIs. Using a generalized additive model (GAM) with a negative binomial distribution and a smoothing spline function, we modelled the time series of total malaria cases (for all villages) as a function of each SMI for time lags ranging from 1 to 30 weeks (thus generating 30 models per SMI). The GAM is an extension of the generalized linear model (GLM): while it includes random effects in the predictor like the GLM does, it can be used with nonparametric smoothing terms instead of

constant parameters[23–25]. The usefulness of the GAM lies in the fact that it provides a flexible method to identify the effects of non-linear covariates in exponential family distributions and in likelihood-based methods[26–28]. However, instead of estimating a single parameter, the GAM provides an unspecified (non-parametric) general function that compares predicted response values to predictor values.

We compared the 30 models generated for each SMI using the unbiased risk estimator (UBRE), i.e. an unbiased estimate of the mean square error of a non-linear biased estimator. For each SMI, the time lag associated with the best model (i.e. with the lowest UBRE) was selected for the multivariate analysis.

Multivariate time analysis. To account for the non-linearity of the relationship between the response and predictor variables, we analysed the time series of weekly cases reported in all villages during the epidemic year using a generalized additive mixed model (GAMM). To account for the non-independence of data from the same village or HC, we fitted this model with nested random intercepts for villages and HCs. The GAMM accounted for space and time processes, by using a Gaussian field and an auto-regressive model. The Gaussian field with a negative exponential variogram accounted for the spatial auto-correlation. The first-order auto-regressive temporal auto-correlation structure was introduced, within the random part of the mixed model, to account for the temporal auto-correlation of malaria incidence. We analysed the time series of cases using selected lagged SMIs (with a smoothing spline function) and of the Euclidean distance between villages and their corresponding HCs as predictors. For each predictor, the standardized incidence ratio (SIR) was estimated by modelling the log-transformed population as the offset.

To account for the non-linearity of the relationship between the response and predictor variables, we calculated SIRs according to the deciles of the distribution of values for each predictor. Indeed, SIRs cannot be calculated with GAMs as they are with GLMs, because when the relationship between the response and the predictor is non-linear, SIRs are not constant across the range of values of the predictor[24,28].

Lastly, we tested the multivariate model fitted with data from the epidemic year to predict the number of cases in both 2016 and 2017.

Software and packages. Statistical analyses were performed using R software (version 3.6.1)[29]. The PCA was performed using the PCA function in the *FactoMineR* package[30]. The GAMs and the GAMM were generated using the "gam" and "gamm" functions in the *mgcv* package, respectively[26–28]. Data overdispersion was tested using the "dispersiontest" function in the *AER* package[31]. The spatial analysis was performed using SatScan™ software (version 9.6). Maps were produced using QGIS software (version 3.10)[32].

Ethics approval and consent to participate. The protocol of this study was reviewed and approved by the Institutional Ethics Committee of the Institut de Recherche en Sciences de la Santé (IEC-IRSS) and registered as No A06/2016/CEIRES and all the methods were performed in accordance with the guidelines and regulations stated in the protocol. Informed consent was obtained from all subjects and/or their legal guardian(s).

Results

Descriptive analysis. A total of 3179 malaria cases were reported in HCs during the epidemic year, corresponding to an incidence of 429.13 cases per 1000 person-years. On average, 61.13 cases per week were reported, with a peak of 132 cases in week 31 of 2016 (week 1 of August; Fig. 2). The curve of cases over the epidemic year shows two peaks (Fig. 2): a very pronounced peak between weeks 27 and 45 of 2016 (August to November), which accounted for 60% of cases, and a less pronounced peak between weeks 7 and 11 of 2017 (mid-February to the end of March), which accounted for 12% of cases.

Synthetic meteorological indicators. The PCA conducted using Kaiser's criterion led us to construct and retain two SMIs that explained 85.4% of the total inertia (Fig. 3A).

The first SMI (i.e. the first principal component) explained 52.9% of the total inertia. The variables that most contributed to this SMI, henceforth called SMI1, were mainly correlated with precipitation variables: weekly mean of daily thermal amplitude (18.24%, correlation coefficient $r = -0.98$), weekly mean of daily average relative humidity (17.74%, $r = 0.96$), weekly rainfall (14. 5%, $r = 0.8$), weekly mean of daily average cloud cover (13.32%, $r = 0.83$), number of rainy days per week (12.47%, $r = 0.81$), and weekly mean of daily maximum temperature (12.03%; $r = -0.79$) (Fig. 3B). The second SMI (i.e. the second principal component) explained 32.5% of the total inertia. The variables that most contributed to this SMI, henceforth called SMI2, were mainly correlated with temperature variables: weekly mean of daily minimum temperature (25.83%; $r = 0.91$), weekly mean of daily average temperature (24.72%; $r = 0.89$), weekly mean of daily maximum temperature (10.51%; $r = 0.58$), and weekly mean of daily average atmospheric pressure (19.6%, $r = -0.79$) (Fig. 3C).

The values of SMI1 were positive between late June and early October, which corresponds to the rainy season (Fig. 4). The values of SMI2 were positive between mid-February and mid-June, which corresponds to the hot dry season (March–June), and between October and November, which corresponds to the transition period between the rainy season and the dry season. Both SMIs were negative throughout the cold dry season (December–mid-February) (Fig. 4).

Spatial analysis. The spatial analysis allowed us to identify and map malaria hotspots for the epidemic year from week 20 (in May) of 2016 to week 19 (in May) of 2017. Four hotspots were detected that accounted for 1685 cases in 1973 inhabitants, i.e. an average incidence rate of 854.02 cases per 1000 person-years (Fig. 5). These hotspots were mainly located in the southern and central parts of the study area. The hotspot with the highest

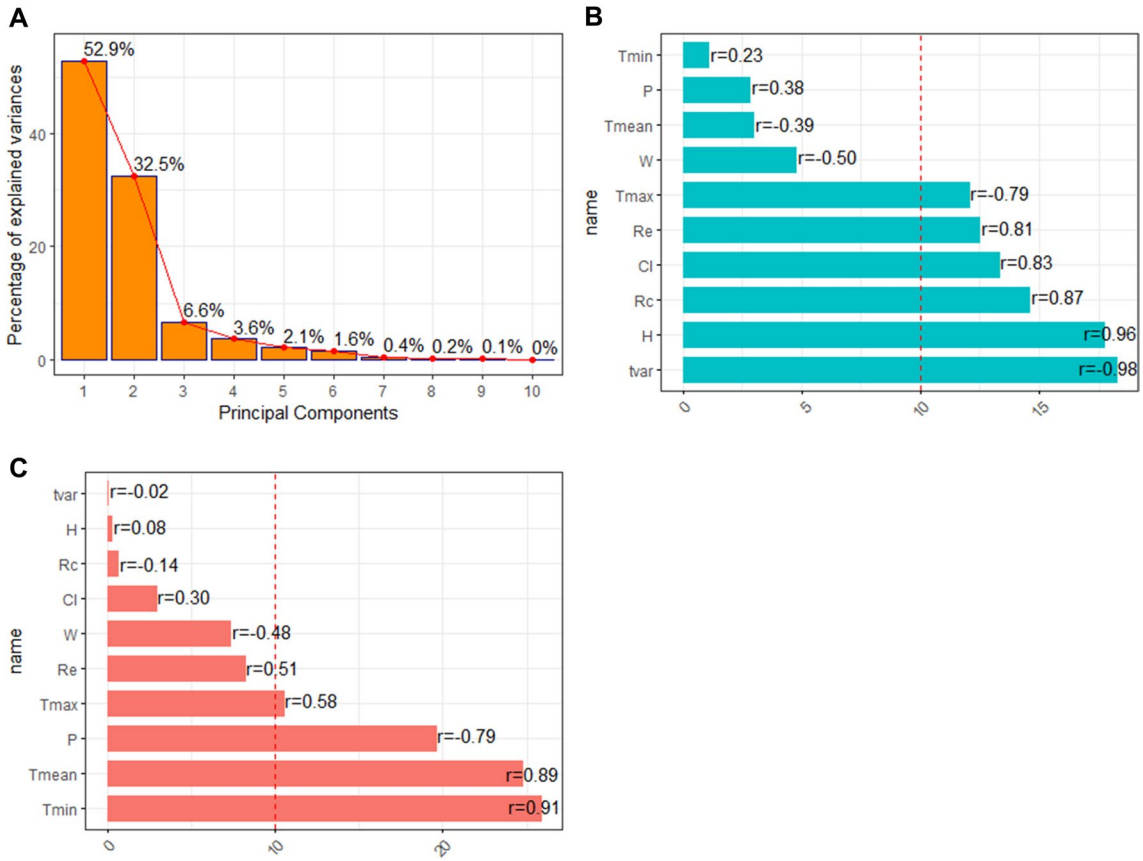

Figure 3. Principal Component Analysis of meteorological variables. Percentage of inertia explained by each principal component (**A**). Contribution of meteorological variables to the first principal component (SMI1; (**B**)) and the second principal component (SMI2; (**C**)). *SMI* synthetic meteorological indicator, *r* correlation coefficient between the meteorological variable and the SMI. Abbreviations of variable names are detailed in Table 1. The dashed line represents the contribution that would have been expected if all variables had contributed equally to the SMI.

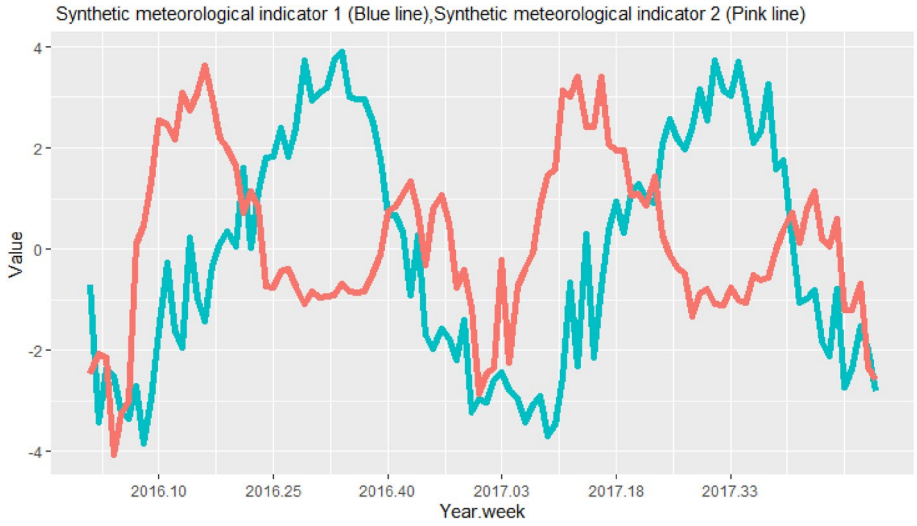

Figure 4. Time series of synthetic meteorological indicator 1 (mainly correlated with precipitation variables) and synthetic meteorological indicator 2 (mainly correlated with temperature variables) from 2016 to 2017.

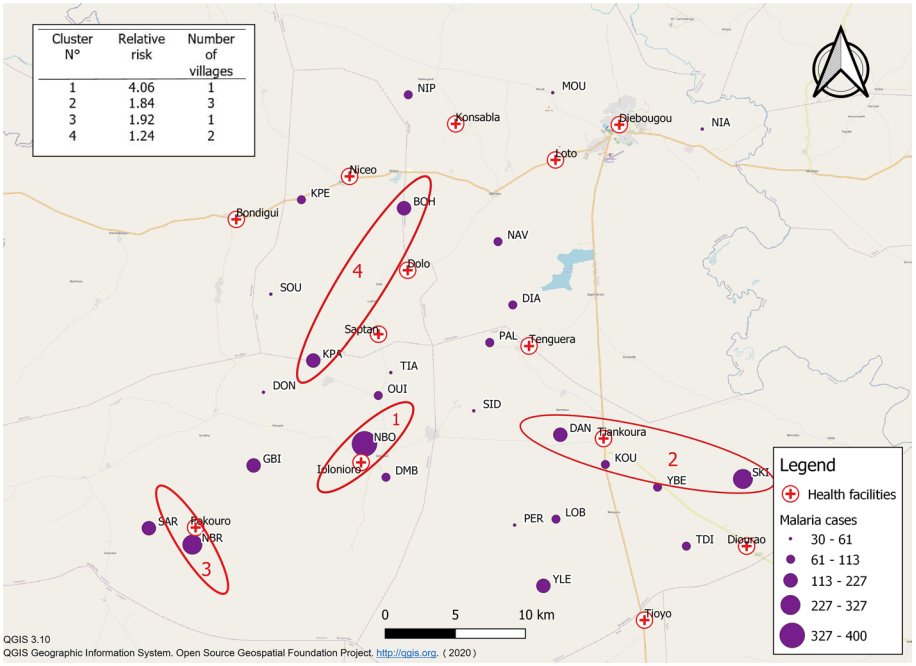

Figure 5. Map of malaria cases detected in 27 villages of Diébougou health district, Burkina Faso, for the epidemic year 2016–2017; hotspots identified with the Kulldorf scanning method. Background: OpenStreetMaps.

risk (hotspot 1; relative risk (RR) = 4.06, $p < 0.0001$) consisted of a single village (Niombouna) and accounted for 400 cases for 228 inhabitants, i.e. an incidence rate of 1,750.75 cases per 1000 person-years. The second hotspot (hotspot 2; RR = 1.84, $p < 0.0001$) was made up of three villages (Sinkiro, Yelbelela, and Dangbara) and accounted for 604 cases for 753 inhabitants, i.e. an incidence rate of 802.12 cases per 1000 person-years. The third hotspot (hotspot 3; RR = 1.92, $p < 0.0001$) was made up of a single village (Niombripo) and accounted for 326 cases for 376 inhabitants, i.e. an incidence rate of 867.02 cases per 1000 person-years. The fourth hotspot (hotspot 4; RR = 1.24, $p = 0.04$) consisted of two villages (Bohero and Kpalbalo) and accounted for 355 cases for 616 inhabitants, i.e. an incidence rate of 576.2 cases per 1000 person-years.

Temporal analysis. The time lag that generated the model with the lowest UBRE was 9 weeks for SMI1 and 16 weeks for SMI2.

The multivariate analysis found greater variability in incidence between HCs (standard deviation (SD) = 5.74) than between villages linked to the same HC (SD = 0.69). The coefficient of the temporal autocorrelation structure indicated the presence of temporal autocorrelation between cases (Phi = 0.32, 95% CI [0.20, 0.38]).

In the multivariate model, lagged SMI1 and lagged SMI2 were significantly associated with the number of malaria cases at the village level ($p < 0.001$ and $p < 0.001$, respectively). The relationship between the number of cases and SMI1 (consisting mainly of precipitation variables) was positive and almost linear (Fig. 6A) across the range of values. A positive non-linear relationship was observed for SMI2 (consisting mainly of temperature variables), with the number of cases increasing for SMI2 values above zero (Fig. 6C). Below zero, changes in SMI2 values did not influence the number of cases (Fig. 6C). The Euclidean distance between villages and their corresponding HCs was not correlated to the recorded malaria incidence ($p = 0.78$).

The evolution of SIRs as a function of SMI values is presented in Fig. 6. For SMI1, risk was constant over deciles 1 to 3 (SIR = 1.07, 95% CI [1.03, 1.10], [1.05, 1.08], and [1.06, 1.08], respectively), increased from decile 4 to 7, and then reached a plateau from decile 8 to 10 (SIR = 1.14 [1.14, 1.14]) (Fig. 6B). For SMI2, risk was constant over deciles 1 to 2 (SIR = 0.99, 95% CI [0.93, 1.05], and [0.98, 1.00], respectively), increased from decile 4 to 8 (SIR = 1.37 [1.36, 1.37]), and then decreased from decile 9 to 10 (Fig. 6D). For the distance, no evolution in the risk was displayed (Fig. 6F).

Prediction. The multivariate model generated for the epidemiological year was used to predict the number of cases in the 27 villages for all of 2016 and for the first 36 weeks of 2017. The resulting prediction was superimposed on the time series of reported cases for graphical analysis (Fig. 7). The prediction followed the overall pattern of the time series of reported cases but with a tendency for underestimation, especially during the second peak in early 2017. In addition, the model predicted the onset of the malaria outbreak for the 2017–2018 epidemic year with a delay of three weeks.

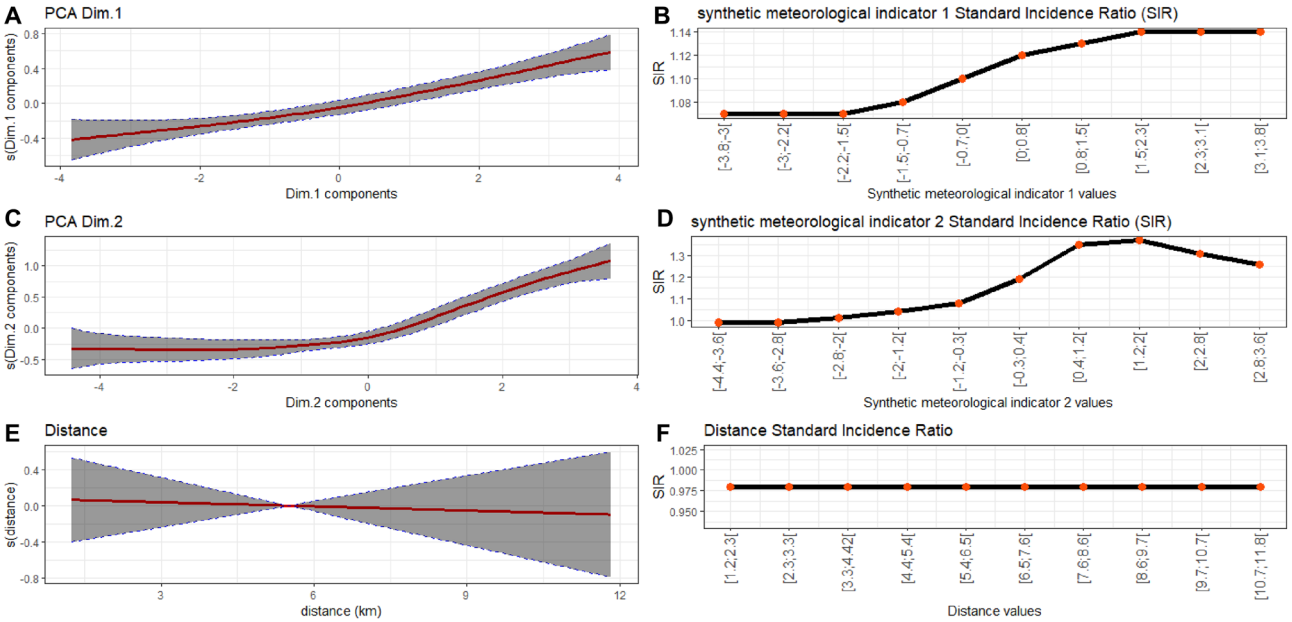

Figure 6. Relationship (red curve) between malaria cases and SMI1 (**A**), SMI2 (**C**), and Euclidean distance to health centre (**E**) with 95% confidence intervals (shaded area). Evolution of the standard incidence ratio (SIR) as a function of SMI1 (mainly correlated with precipitation variables) (**B**), SMI2 (mainly correlated with temperature variables) (**D**), and Euclidean distance to health centre (**F**).

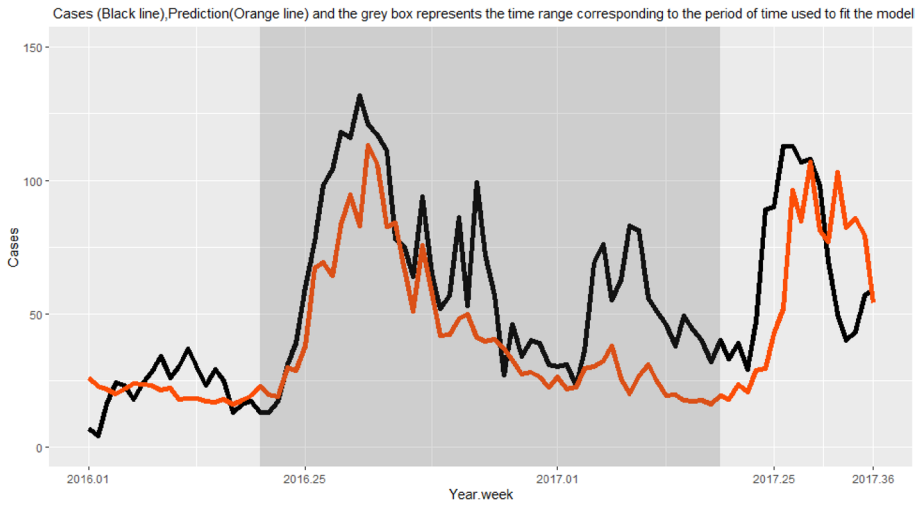

Figure 7. Cumulative number of reported cases (black line) and predicted cases (orange line) in 27 villages of South-Western Burkina Faso using a meteorological model.

Discussion

In this study, we analysed the spatio-temporal distribution of malaria cases in 27 villages of South-Western Burkina Faso.

The spatial analysis conducted using the Kulldorf scanning method helped to identify four malaria hotspots. The first three hotspots were located in the southern part of the study area and the last one was located in the central part, reflecting spatial heterogeneity in the distribution of cases. A comparison of the spatial distribution of these hotspots with that of mosquito vector density[33] showed no correlation between the two, leading us to conclude that the spatial heterogeneity of vector density does not explain the distribution of hotspots in our study area.

A number of studies have found an association between spatial inequalities in access to care and spatial heterogeneity of malaria incidence[34,35]. Yet, contrary to what has been reported elsewhere[36], we failed to found a correlation between the number of malaria cases and the Euclidean distance between villages and their corresponding HC. We used Euclidean distance because it is considered the simplest proxy for travel time, which is

considered a good measure of access to care. However, Euclidean distance may not have been the best option, as roads in our study area are in highly variable condition and some become impassable during the rainy season, with some villages left completely isolated. Future studies in the region should use better proxies for travel time in trying to explain the detected hotspots[36].

Since entomological factors and spatial inequalities in access to care failed to explain the distribution of hotspots in our study area, other potential explanatory factors should be investigated in the future, including socio-economic factors (level of education, income, professional activity, individual and societal behaviour, etc.)[37–40] and factors linked to LLIN usage[41–44]. Such investigations could help to explain in particular why the two hotspots composed of a single village (Niombripo and Niombouna) had a much higher incidence than neighbouring villages. Nevertheless, hotspot analyses like ours make it possible to identify, in a simple and cost-efficient manner, villages that can constitute priority areas for intervention. Indeed, studies conducted elsewhere have shown that targeting hotspots helps to reduce malaria transmission[45,46]. This strategy is appropriate in resource-limited countries like Burkina Faso as it allows for efficient allocation of prevention resources[14–16].

Our analysis of the temporal dynamics of malaria cases found a strong correlation between malaria incidence and two SMIs with specific time lags. These SMIs were constructed through a PCA of meteorological data derived from readily and rapidly available satellite imagery. The first SMI (SMI1: positively correlated with cumulative rainfall, humidity, cloud cover, and number of rainy days, and negatively correlated with thermal amplitude) corresponded to the rainy season, while the second (SMI2: positively correlated with temperature and negatively correlated with atmospheric pressure) corresponded to the warm periods preceding and following the rainy season. We found that SMI1 and SMI2 predicted the number of cases with a time lag of 9 and 16 weeks, respectively, which is consistent with studies carried out in Burkina Faso, Mali, and Ethiopia[14,47,48].

In our study, the relationship between rainfall (SMI1) and the number of cases was quasi-linear, as was the case in a study performed in the Ouagadougou area of Burkina Faso[14]. By contrast, two studies conducted in the Sahel region—one in Mali (Niger River Valley, Timbuktu region) and the other in Senegal (Bambey and Fatick Health Districts)—found a monotonic non-linear relationship between rainfall and malaria incidence[15,16]. The drop in the number of cases above a certain level of cumulative rainfall observed in Mali and Senegal may be explained by the flushing out of larval breeding sites, which can lead to high mortality in Anopheles larval populations[49,50] and can reduce the human biting rate[50]. Vector populations are almost monospecific in these two countries: They are largely dominated by *An. arabiensis* in Senegal[51,52] and by *An. coluzzii* in Mali[53]. These two species are also present in our study area and in the Ouagadougou area of Burkina Faso[54]. However, in both these areas, they live in sympatry with both *An. gambiae s.s.* and *An. funestus*[33,55–57]. The quasi-linear relationship observed in our study between rainfall and the number of malaria cases may be explained by the fact that these species are not very susceptible to flushing out, due to rapid larval development in the case of *An. gambiae s.s.*[58,59] and to a preference for deeper environments in the case of *An. funestus*[60]. These species may therefore relay *An. coluzzii* when abundances of this later fall due to excessive rainfall.

In our study, the relationship between the number of malaria cases and temperature (SMI2) was non-linear. This is consistent with findings from two other studies conducted in the Sahel region (in Mali and in the Ouagadougou area of Burkina Faso)[14,15]. However, unlike these studies, we found no negative relationship between the number of malaria cases and temperature at higher temperature values. This discrepancy may be explained by the fact that temperatures can reach higher values in Mali and in the Ouagadougou area (> 34 °C) than in the Diébougou region, which is sufficient to inhibit the development of Anopheles larvae[61] and to reduce the survival of adult Anopheles[62,63]. In addition, we found that below a certain temperature, an increase in temperature had no effect on the number of cases (a finding also observed by Cissoko et al.[15]). Our hypothesis is that the increase in temperature, which should favour the development of *Anopheles*, is compensated by another phenomenon at low SMI2 values. While this phenomenon has yet to be clearly identified, high levels of LLIN usage during cooler periods may be a contributing factor[41].

Our spatio-temporal model fitted with two lagged SMIs and case data for a single epidemiological year helped to predict the start of the next outbreak nine weeks in advance, but with an error of three weeks (i.e. the actual outbreak began three weeks before the prediction). The prediction was good enough to make it possible to issue early warnings and to organize local prevention campaigns ahead of time. Our model could probably be improved with routine inclusion of new data and regular updated predictions. This study can be generalized to determine optimal periods and zones for prevention campaigns or interventions campaigns related to weather-related diseases such as dengue fever, malaria, etc.… Indeed, prioritizing a few numbers of areas and periods is helpful in strengthening malaria control programme in the context of lack of resources. This is even more important when countries will reach the pre-elimination phase: resources should then be concentrated on the most effective areas and periods. This is the principle of the "bottle neck" approach[11]. On the other hand, recent papers have shown that there is no conclusive evidence that targeted hotspot interventions accelerate malaria elimination. Therefore, a targeted approach to high-risk individuals that allows for a precise delineation of parasite transmission networks within and between households may be investigated[45,64]. For this purpose, data from HC consultations should be made available quickly, ideally at the same pace as ERA5 meteorological data (i.e. within 5 days). This can easily be achieved by using connected tablets for data entry.

Conclusion

In this study, a spatial analysis was conducted that highlighted the spatial heterogeneity of malaria cases and helped to identify four malaria hotspots in South-Western Burkina Faso. In the temporal analysis, an effective predictive model was built with data obtained through passive case detection and with simple and accessible meteorological data. Future studies should further investigate the detected hotspots to identify the local determinants of transmission. Our spatio-temporal analysis provides a powerful prospective method to identify high-risk

areas that may constitute priority areas during malaria prevention campaigns. Since our approach allowed us to determine the hotspots and to predict the start of the next annual epidemic, this approach should be cost-effective. Because of the scarcity of the resources in developing countries such as Burkina Faso, implementing the same policy at the same time through the whole health districts should be less cost-effective. Indeed, malaria epidemiology, at least onset, peak and length of malaria annual epidemic, is variable, as well as the environmental characteristics, through the country. Being able to determine areas and predict malaria onset dates can help policies makers to target actions at the right time and at the right places.

Data availability

The datasets analysed in this study may be available from the last author on reasonable request.

Received: 9 June 2021; Accepted: 27 September 2021
Published online: 08 October 2021

References

1. OMS. *Rapport sur le Paludisme dans le Monde* (2019). https://www.who.int/malaria/publications/world-malaria-report-2019/report/fr. Accessed on March 6, 2020.
2. Ministère de la santé du Burkina Faso. *Annuaire Statistique 2015* (2015). http://cns.bf/IMG/pdf/annuaire_ms_2015_signe.pdf. Accessed on March 6, 2020.
3. Institut National de la Statistique et de la Démographie, Burkina Faso. *Enquête sur les Indicateurs du Paludisme 2014* (2014). https://dhsprogram.com/pubs/pdf/MIS19/MIS19.pdf. Accessed on January 4, 2020.
4. Ministère de la santé. *Directives Nationales Pour la Prise en Charge du Paludisme dans les Formations Sanitaires du Burkina Faso* (2010). http://pdf.usaid.gov/pdf_docs/PA00JPHB.pdf. Accessed on March 13, 2020..
5. Ministère de la Sante. *Plan Stratégique National de lutte Contre le Paludisme 2016–2020* (2017). http://onspsante.bf/sites/default/files/publications/166/PSN%20%20%20%202016-2020_Paludisme_20_02_2017.pdf. Accessed on March 13, 2020.
6. INSD. *Enquête sur les Indicateurs du Paludisme 2017–2018* (2018). http://www.insd.bf/n/contenu/enquetes_recensements/enquete_palu/EIPBF_2018.pdf. Accessed on January 6, 2020.
7. Ouédraogo, A. L. *et al.* Efficacy and safety of the mosquitocidal drug ivermectin to prevent malaria transmission after treatment: A double-blind, randomized, clinical trial. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **60**, 357–365 (2015).
8. Stoute, J. A. *et al.* Long-term efficacy and immune responses following immunization with the RTS, S malaria vaccine. *J. Infect. Dis.* **178**(4), 1139–1144 (1998).
9. Selvaraj, P., Wenger, E. A. & Gerardin, J. Seasonality and heterogeneity of malaria transmission determine success of interventions in high-endemic settings: A modeling study. *BMC Infect. Dis.* **18**, 413 (2018).
10. Mogeni, P. *et al.* Effect of transmission intensity on hotspots and micro-epidemiology of malaria in sub-Saharan Africa. *BMC Med.* **15**, 121 (2017).
11. Landier, J., Rebaudet, S., Piarroux, R. & Gaudart, J. Spatiotemporal analysis of malaria for new sustainable control strategies. *BMC Med.* **16**, 226 (2018).
12. Ribeiro, J. M. C., Seulu, F., Abose, T., Kidane, G. & Teklehaimanot, A. Temporal and spatial distribution of anopheline mosquitos in an Ethiopian village: Implications for malaria control strategies. *Bull. World Health Organ.* **1996**(743), 299–305 (1996).
13. Bousema, T. *et al.* Hitting hotspots: Spatial targeting of malaria for control and elimination. *PLoS Med.* **9**, e1001165 (2012).
14. Ouedraogo, B. *et al.* Spatio-temporal dynamic of malaria in Ouagadougou, Burkina Faso, 2011–2015. *Malar. J.* **17**, 138 (2018).
15. Cissoko, M. *et al.* Geo-epidemiology of malaria at the health area level, dire health district, Mali, 2013–2017. *Int. J. Environ. Res. Public. Health* **17**, 3982 (2020).
16. Dieng, S. *et al.* Spatio-temporal variation of malaria hotspots in Central Senegal, 2008–2012. *BMC Infect. Dis.* **20**, 424 (2020).
17. INSD. *Tableau de bord économique et social 2014 de la région du Sud-Ouest* (2015). https://www.insd.bf/contenu/pub_periodiques/tableaux_de_bord/TBES/TBES_SO_2014.pdf. Accessed on March 15, 2020.
18. INSD. *Enquête Nationale sur le Secteur de l'orpaillage* (2017). http://www.insd.bf/contenu/enquetes_recensements/ENSO/Principaux_Resultats_ENSO.pdf. Accessed on March 15, 2020.
19. Copernicus Climate Change Service. *ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate* (Copernicus Climate Change Service, 2017).
20. Belmonte Rivas, M. & Stoffelen, A. Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT. *Ocean Sci.* **15**, 831–852 (2019).
21. Yeomans, K. & Golder, P. The Guttman-Kaiser criterion as a predictor of the number of common factors. *The Statistician.* https://doi.org/10.2307/2987988 (1982).
22. SaTScan_Users_Guide.pdf (2021). https://www.satscan.org/cgi-bin/satscan/register.pl/SaTScan_Users_Guide.pdf?todo=process_userguide_download. Accessed on February 1, 2020.
23. Wood, S. N. *Generalized Additive Models: An Introduction with R* (CRC Press, 2006).
24. Guisan, A., Edwards, T. C. Jr. & Hastie, T. Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecol. Model.* **157**, 12 (2002).
25. Hastie, T. & Tibshirani, R. Generalized additive models. *Stat. Sci.* **1**, 297–310 (1986).
26. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, 3–36 (2011).
27. Wood, S. N. *Generalized Additive Models: An Introduction with R* 2nd edn. (Routledge and CRC Press, 2017).
28. Wood, S. N. Thin plate regression splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65**, 95–114 (2003).
29. R Development Core Team. *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, 2005). http://www.R-project.org. Accessed on February 1, 2020.
30. Lê, S., Josse, J. & Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
31. Kleiber, C. & Zeileis, A. Applied Econometrics with R: Package Vignette and Errata. https://doi.org/10.1007/978-0-387-77318-6 (2015).
32. QGIS.org QGIS Geographic Information System. *Open Source Geospatial Foundation Project.* (2020). http://qgis.org. Accessed on February 1, 2020.
33. Soma, D. D. *et al.* Anopheles bionomics, insecticide resistance and malaria transmission in southwest Burkina Faso: A pre-intervention study. *PLoS ONE* **15**, e0236920 (2020).
34. Lombrail, P. & Pascal, J. Inégalités sociales de santé et accès aux soins. *Trib. Santé* **8**, 31–39 (2005).
35. Kadio, K., Ridde, V. & Mallé Samb, O. Les difficultés d'accès aux soins de santé des indigents vivant dans des ménages non pauvres. *Santé Publique* **26**, 89–97 (2014).
36. Ilboudo, S. D. O., Sombié, I., Soubeiga, A. K. & Dræbel, T. Facteurs influençant le refus de consulter au centre de santé dans la région rurale Ouest du Burkina Faso. *Santé Publique* **28**, 391–397 (2016).

37. Pierrat, C. Risque palustre: Appréhender la vulnérabilité des individus à l'échelle locale (Sud du Bénin). *VertigO Rev. Électron. Sci. Environ.* https://doi.org/10.4000/vertigo.11549 (2012).
38. Yonkeu, S., Maïga, A. H., Wethé, J., Mampouya, M. & Maga, G. P. Conditions socio-économiques des populations et risques de maladies: Le bassin versant du barrage de Yitenga au Burkina Faso. *VertigO Rev. Électron. En Sci. Environ.* https://doi.org/10.4000/vertigo.4778 (2003).
39. Baragatti, M. *et al.* Social and environmental malaria risk factors in urban areas of Ouagadougou, Burkina Faso. *Malar. J.* **8**, 13 (2009).
40. Berthélemy, J.-C. & Seban, J. Dépenses de santé et équité dans l'accès aux services de santé dans les pays en développement. *Rev. Déconomie Dév.* **17**, 33–71 (2009).
41. Kreuels, B. *et al.* Spatial variation of malaria incidence in young children from a geographically homogeneous area with high endemicity. *J. Infect. Dis.* **197**, 85–93 (2008).
42. Brooker, S. *et al.* Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya. *Trop. Med. Int. Health* **9**, 757–766 (2004).
43. Van Der Hoek, W. *et al.* Towards a risk map of malaria for Sri Lanka: The importance of house location relative to vector breeding sites. *Int. J. Epidemiol.* **32**, 280–285 (2003).
44. el Samani, F. Z., Willett, W. C. & Ware, J. H. Nutritional and socio-demographic risk indicators of malaria in children under five: A cross-sectional study in a Sudanese rural community. *J. Trop. Med. Hyg.* **90**, 69–78 (1987).
45. Bousema, T. *et al.* The impact of hotspot-targeted interventions on malaria transmission in Rachuonyo South District in the Western Kenyan highlands: A cluster-randomized controlled trial. *PLoS Med.* **13**, e1001993 (2016).
46. Nesbitt, R. C. *et al.* Methods to measure potential spatial access to delivery care in low- and middle-income countries: A case study in rural Ghana. *Int. J. Health Geogr.* **13**, 25 (2014).
47. Sissoko, M. S. *et al.* Temporal dynamic of malaria in a suburban area along the Niger River. *Malar. J.* **16**, 420 (2017).
48. Kibret, S., Glenn Wilson, G., Ryder, D., Tekie, H. & Petros, B. Environmental and meteorological factors linked to malaria transmission around large dams at three ecological settings in Ethiopia. *Malar. J.* **18**, 54 (2019).
49. Paaijmans, K. P., Wandago, M. O., Githeko, A. K. & Takken, W. Unexpected high Losses of *Anopheles gambiae* larvae due to rainfall. *PLoS ONE* **2**, e1146 (2007).
50. Moiroux, N. *et al.* Modelling the risk of being bitten by malaria vectors in a vector control area in southern Benin, west Africa. *Parasit. Vectors* **6**, 71 (2013).
51. Robert, V. *et al.* Malaria transmission in the rural zone of Niakhar, Senegal. *Trop. Med. Int. Health* **3**, 667–677 (1998).
52. Sy, O. *et al.* Entomological impact of indoor residual spraying with pirimiphos-methyl: A pilot study in an area of low malaria transmission in Senegal. *Malar. J.* **17**, 64 (2018).
53. Sogoba, N. *et al.* Spatial distribution of the chromosomal forms of *Anopheles gambiae* in Mali. *Malar. J.* **7**, 205 (2008).
54. Ouédraogo, M. *et al.* Spatial distribution and determinants of asymptomatic malaria risk among children under 5 years in 24 districts in Burkina Faso. *Malar. J.* **17**, 460 (2018).
55. Costantini, C. *et al.* Living at the edge: Biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*. *BMC Ecol.* **9**, 16 (2009).
56. Fournet, F. *et al.* Diversity in anopheline larval habitats and adult composition during the dry and wet seasons in Ouagadougou (Burkina Faso). *Malar. J.* **9**, 78 (2010).
57. Guelbeogo, W. M. *et al.* Seasonal distribution of *Anopheles funestus* chromosomal forms from Burkina Faso. *Malar. J.* **8**, 239 (2009).
58. Diabate, A. *et al.* Larval development of the molecular forms of *Anopheles gambiae* (Diptera: Culicidae) in different habitats: A transplantation experiment. *J. Med. Entomol.* **42**, 548–553 (2005).
59. Paaijmans, K. P., Huijben, S., Githeko, A. K. & Takken, W. Competitive interactions between larvae of the malaria mosquitoes *Anopheles arabiensis* and *Anopheles gambiae* under semi-field conditions in western Kenya. *Acta Trop.* **109**, 124–130 (2009).
60. Hamon, J. *Biologie d'Anopheles funestus. 6 p. multigr.* (1955). http://www.documentation.ird.fr/hor/fdi:28874. Accessed on June 23, 2020
61. Bayoh, M. N. & Lindsay, S. W. Effect of temperature on the development of the aquatic stages of *Anopheles gambiae* sensu stricto (Diptera: Culicidae). *Bull. Entomol. Res.* **93**, 375–381 (2003).
62. Lyons, C. L., Coetzee, M. & Chown, S. L. Stable and fluctuating temperature effects on the development rate and survival of two malaria vectors, *Anopheles arabiensis* and *Anopheles funestus*. *Parasit. Vectors* **6**, 104 (2013).
63. Lyons, C. L., Coetzee, M., Terblanche, J. S. & Chown, S. L. Desiccation tolerance as a function of age, sex, humidity and temperature in adults of the African malaria vectors *Anopheles arabiensis* and *Anopheles funestus*. *J. Exp. Biol.* **217**, 3823–3833 (2014).
64. Stresman, G., Bousema, T. & Cook, J. Malaria hotspots: Is there epidemiological evidence for fine-scale spatial targeting of interventions? *Trends Parasitol.* **35**, 822–834 (2019).

Acknowledgements

The authors would like to thank the Ministry of Health of Burkina Faso, in particular Dr. Dembélé Henri, Médecin–chef de poste in Diebougou, and the local medical team, for facilitating data collection. Special thanks are due to Mr. Maiga Issouf for his strong involvement in data collection. We are also grateful to the "Laboratoire Mixte International sur les Maladies à Vecteurs" (LAMIVECT) for providing technical support. Lastly, we would like to thank Mr. Ouattara Adama and Mr. Félix Zoumènou for providing administrative support.

Author contributions

C.S.B., J.G., C.P., A.K., R.K.D. and N.M. designed the study; A.S. and D.D.S. conducted the field study; C.S.B., J.G. and N.M. designed the statistical analysis plan; C.S.B. conducted the statistical analysis under the supervision of J.G. and N.M.; C.S.B. performed the cartographic analysis with the participation of M.C. and P.T.; S.D. participated in the statistical analysis; C.S.B., J.G. and N.M. validated and interpreted the results; C.S.B., J.G. and N.M. wrote the manuscript; all authors read and approved the final manuscript.

Funding

This work was part of the REACT project, funded by the French Initiative 5%—Expertise France (No. 15SANIN213). C.S.B. received a grant from the *Revivre Développement* Endowment Fund through the NGO *Prospective et Coopération*, as well as a French Government Grant through the French Embassy in Burkina Faso. The funding was also provided by Excellence Initiative of Aix–Marseille University –A*MIDEX.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.