AJAS / SPAS

African Journal of Applied Statistics
Vol. 11 (2), 2024, pages 1629 - 1648.
DOI: https://dx.doi.org/10.16929/ajas/2024.1629.285

ISSN 2316-0861

Temporal Analysis and Prediction of Malaria Dynamics Using Meteorological Data in Southeastern of Senegal

Leontine Ndogou Bakhoum [1,*], **Mor Absa Loum** [2], **Mouhamad Mounirou Allaya** [3], **Medoune Ndiop** [4], **Mamadou Bousso** [5] **and Jean Louis Abdourahim Ndiaye** [6]

[1,*] Faculty of Science and Technology, Iba Der Thiam University of Thiès, Thiès, Senegal
[2] Faculty of Engineering Sciences, Iba Der Thiam University of Thiès, Thiès, Senegal
[3] Faculty of Economic and Social Sciences, Iba Der Thiam University of Thiès, Thiès, Senegal
[4] National Malaria Control Program, Dakar, Senegal
[5] Faculty of Economic and Social Sciences, Iba Der Thiam University of Thiès, Thiès, Senegal
[6] Parasitology and Mycology Department, Faculty of Health Sciences, Iba Der Thiam University of Thiès, Thiès, Senegal

Received on November 8, 2025; Accepted on December 09, 2025; published online on December 31, 2025

Abstract. Malaria continues to be a major public health challenge in Senegal. The intricate links between climate and malaria complicate the development of effective control strategies, despite ongoing efforts by the National Malaria Control Program and international partners. (See the full abstract at page 1630).

Key words: generalized additive model; malaria; principal component analysis; meteorological factors; prediction; time series.
AMS 2020 Mathematics Subject Classification: 62G08; 62P10; 62H25.

(*) Corresponding author: Leontine Ndogou Bakhoum (lndogou.bakhoum@univ-thies.sn)
Mor Absa Loum: morabsa.loum@univ-thies.sn
Mouhamad Mounirou Allaya : mouhamad.allaya@univ-thies.sn
Medoune Ndiop : mnzop5@gmail.com
Mamadou Bousso : mbousso@univ-thies.sn
Jean Louis Abdourahim Ndiaye : jlndiaye@univ-thies.sn

Full Abstract Malaria continues to be a major public health challenge in Senegal. The intricate links between climate and malaria complicate the development of effective control strategies, despite ongoing efforts by the National Malaria Control Program and international partners.This study investigates the relationship between weekly malaria cases and meteorological factors in four districts of Senegal (Kédougou, Salémata, Saraya, and Dianké Makha) using Generalized Additive Models. Data were sourced from *DHSI2* (malaria cases) and *NASA GIOVANNI* (meteorological data). Principal component analysis identified key explanatory variables, with climate regimes determined via Hierarchical Ascending Classification. Results show that the first component's two-week lag significantly reduces malaria risk at high values. The second component exhibits a nonlinear relationship: moderate levels are protective while very high values increase risk.The third component shows that a 15-week lag, moderate to high precipitation provides a protective effect (50-70% risk reduction), while excessive precipitation (80-100th percentiles) increases malaria risk by promoting mosquito proliferation.

Résumé (Abstract in French) Au Sénégal, le paludisme demeure un problème de santé publique. La complexité de la relation entre les facteurs climatiques et la transmission du paludisme rend difficile la mise en oeuvre efficace des stratégies de lutte contre le paludisme malgré les efforts du programme National de Lutte contre le paludisme du Sénégal et de ses partenaires internationaux. Cette étude examine la relation entre les cas hebdomadaires de paludisme et les facteurs météorologiques dans quatre districts du Sénégal (Kédougou, Salémata, Saraya et Dianké Makha) en utilisant des Modèles Additifs Généralisés. Les données proviennent de *DHSI2* (cas de paludisme) et de *NASA GIOVANNI* (données météorologiques). L'analyse en composantes principales a identifié les variables explicatives clées. Les régimes climatiques sont déterminés par Classification Ascendante Hiérarchique. Les résultats montrent qu'un décalage de deux semaines de la première composante réduit significativement le risque de paludisme à valeurs élevées. La deuxième composante présente une relation non linéaire : les niveaux faibles à modérés sont protecteurs tandis que les valeurs très élevées augmentent le risque de paludisme. La troisième composante montre qu'un décalage de 15 semaines révèle un effet protecteur des précipitations modérées (réduction du risque de 50-70%) mais que les précipitations excessives (déciles 80-100%) augmentent le risque de paludisme en favorisant la prolifération des moustiques.

Presentations of authors.

Leontine Ndogou Bakhoum, *M.Sc.*, is preparing a Ph.D. degree in Data Science, modeling and Applications at Iba Der Thiam University of Thies, SENEGAL.

Mor Absa LOUM, *Ph.D.*, is a Assistant Professor of biostatistics and modeling at the Iba Der Thiam University of Thiès,SENEGAL.

Mouhamad Mounirou Allaya, *Ph.D.*, is an Associate Professor of Applied Mathematics at Iba Der Thiam University of Thiès, SENEGAL.

Medoune NDIOP is a researcher at the National Malaria Control Program (*NMCP/PNLP*), SENEGAL.

Mamadou Bousso, *Ph.D.*, is an Associate Professor in Applied Computer Science and Artificial Intelligence at Iba Der Thiam University of Thiès, SENEGAL.

Jean-Louis Abdourahim NDIAYE, *Ph.D.*, is a full Professor of parasitology at the Iba Der Thiam University of Thiès, SENEGAL.

1. Introduction

Malaria remains one of the leading causes of morbidity and mortality globally, particularly in Africa. In 2023, the number of malaria cases worldwide was estimated at 263 million, with an incidence of 60.4 cases per 1,000 population at risk and 597,000 deaths were estimated WHO(2023). In Senegal, malaria is endemic, and its distribution varies considerably across different regions NMCP(2023). The southeastern regions (Kolda, Tambacounda, and Kédougou) bear the highest burden, accounting for 78.5% of malaria cases and 43.6% of malaria-related deaths in 2023 Tairou *and al.*(2022). To mitigate the disease's burden, Senegal's National Malaria Control Program (*NMCP*) has implemented a range of strategies, including intermittent preventive treatment (*IPT*) for pregnant women, universal access to rapid diagnostic tests (*RDTs*) and artemisinin-based combination therapies (*ACTs*), seasonal malaria chemo-prevention (*SMC*) for children under 10, and vector control measures such as long-lasting insecticide-treated nets (*LLINs*) and indoor residual spraying (*IRS*). Despite the observed effectiveness of many interventions, malaria remains concentrated in the southeastern districts of Kédougou, Salémata, Saraya, and Dianké Makha Manga *and al.*(2023), Gadiaga *and al.*(2024), Olliaro *and al.*(2008), Diouf *and al.*(2018). Malaria is a vector-borne disease influenced by multiple factors, including the life cycle of the mosquito vector, the parasite, the human host, and environmental conditions Diouf *and al.*(2017), Ilunga-Ilunga *and al.*(2016), Meibalan and Marti (2017). There is a close relationship between temperature, rainfall, humidity, and malaria transmission Diao *and al.*(2023), Fall *and al.*(2023b), Seck *and al.*(2017), Dieng *and al.*(2020b), Fall *and al.*(2022), Fall *and al.*(2023a), Taye *and al.*(2015), Midekisa *and al.*(2015), Siraj *and al.*(2015), Ugwu and Zewotir (2020). Indeed, in Senegal, research studies have demonstrated a clear seasonal pattern in malaria transmission, with higher incidence during and shortly after the rainy season Diao *and al.*(2023), Fall *and al.*(2023b), Seck *and al.*(2017). The temporal and spatial variations in malaria incidence have been attributed to the intensity and frequency of rainfall, with peaks in malaria cases typically lagging one month behind peak rainfall Dieng *and al.*(2020b), Fall *and al.*(2022), Fall *and al.*(2023a). Furthermore, a study conducted in three different endemic regions of Senegal Diao *and al.*(2023) demonstrated that incorporating appropriate time lags for meteorological variables significantly improves the accuracy of malaria incidence forecasts. However, the influence of meteorological factors on malaria transmission varies by geographic area, study unit, and time period Taye *and al.*(2015), Midekisa *and al.*(2015), Siraj *and al.*(2015), Ugwu and Zewotir (2020). Additionally, meteorological factors play a crucial role in the formation and persistence of mosquito breeding sites, directly influencing malaria transmission risk. For instance, higher temperatures, increased rainfall, and humidity can lead to the proliferation of malaria-carrying mosquitoes The President's Malaria Initiative (PMI) (2020). Moreover, temperature affects the mosquito life cycle as well as the extrinsic incubation period of Plasmodium parasites, while humidity influences mosquito survival and biting behavior. It is therefore essential to understand the specific effects of these climatic factors in each area and at fine temporal resolutions in order to adapt interventions and optimize malaria control strategies. Yet, few studies have thoroughly explored the non-linear effects of meteorological factors on malaria incidence at fine temporal scales, particularly in southeastern Senegal, where this understanding could guide more targeted and effective interventions. Therefore, this study aims to analyze the nonlinear effects of environmental factors on malaria incidence, accounting for temporal lags and exploring variable selection to develop a predictive model of malaria dynamics. In this study, we use a methodological approach based on Generalized Additive Models (*GAM*), combining weekly malaria case surveillance data and meteorological data. We focus on four districts in southeastern Senegal (Kédougou, Salémata, Saraya, and Dianké Makha), which have consistently recorded the highest incidence rates since 2016. This observation stems from our preliminary descriptive analysis of monthly surveillance data, which revealed a persistent concentration of malaria cases in these areas over the years, highlighting a crucial challenge for National Malaria Control Program (*NMCP*) in determining the optimal start date for Seasonal Malaria Chemo-prevention (*SMC*).

2. Material and Methods

2.1. Study Area

The study was conducted in four health districts located in the extreme southeast of Senegal: Kédougou, Salémata, and Saraya, which are predominantly part of the Kédougou region, and Dianké Makha, located in the Tambacounda region (Figure 1). Throughout the study period, at least one malaria case was reported each week in all four districts. A total of 590,230 rapid diagnostic tests (*RDTs*) were performed across the four districts, resulting in 298,072 positive malaria cases. On average, 1,146.4 cases were reported weekly, with considerable variability observed (standard deviation: 1,564.2 cases). This area of Senegal had an estimated population of 210,502 in 2022. This region is one of the wettest in the country, with a rainy season lasting approximately six months, from May to October. Rainfall in this region exhibits significant temporal variability, peaking during August and September. The periods of highest temperatures are recorded from March to May. Relative humidity is very high during the rainy season, exceeding 80% in August and October (Figure 2). The time series of malaria cases revealed a clear seasonal pattern, with incidence peaks occurring around early October each year (Figure 3). These peaks correspond closely to the end of the rainy season, suggesting a potential lagged relationship between rainfall and malaria transmission.

2.2. Data Collection

This study utilized datasets covering the period from January 2018 to December 2022, including:

(i) Weekly malaria cases confirmed by rapid diagnostic tests (*RDTs*) for each district, extracted from the *DHIS2* platform DHIS2(2024);

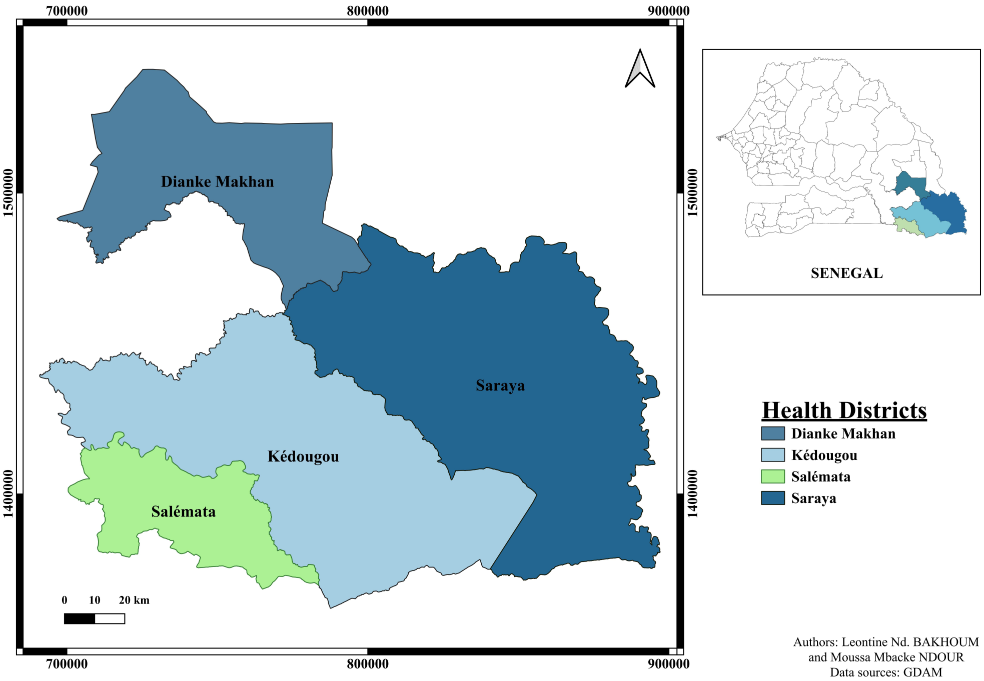


Fig. 1. Map of the study area showing the location of the four health districts: Kédougou,Salémata, Saraya and Dianké Makha

(ii) Meteorological data collected from the *NASA GIOVANNI* platform NASA (2024), including:
   - Weekly sum of daily rainfall intensity (mm) and frequency;
   - Weekly mean of daily mean, minimum, and maximum temperatures ($°C$);
   - Weekly mean of daily maximum, mean, and minimum wind speeds (m/s) at 10 meters;
   - Weekly mean of daily wind direction;
   - Weekly mean of daily relative humidity (%);
   - Weekly mean of daily specific humidity (%);
   - Weekly mean of daily atmospheric pressure (hPa).

(iii) Population size data collected from National Agency for Statistics and Demography of Senegal National Agency for Statistics and Demography of Senegal.

In this study, the response variable (or dependent variable) is the number of malaria cases confirmed by *RDTs* (Figure 3).

2.3. Modeling Approach

To examine the relationship between meteorological factors and malaria incidence, we used a Generalized Additive Model (*GAM*). *GAM* is chosen for their flexibility in capturing non-linear relationships between the response variable (weekly malaria

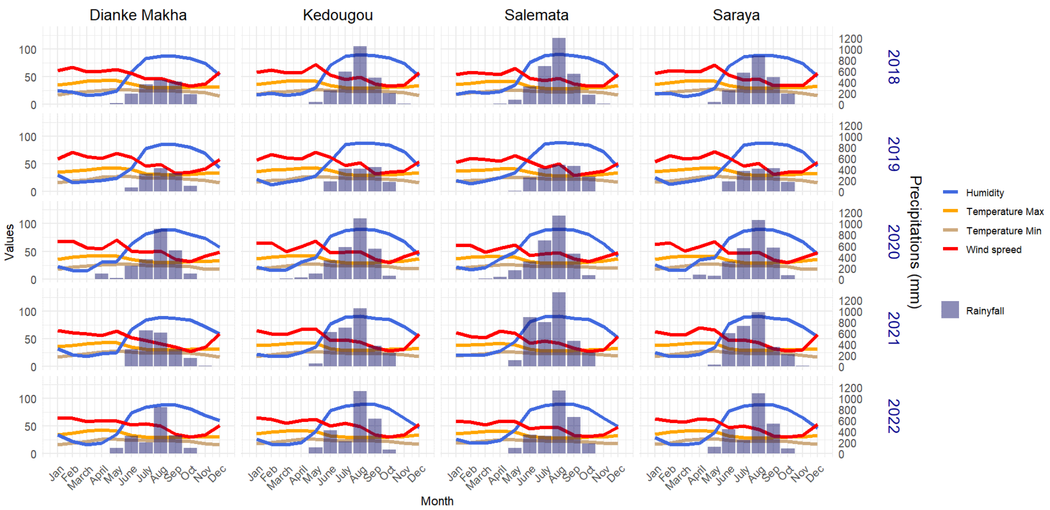


Fig. 2. Monthly variations of (minimum temperature light orange line) and (maximum temperature orange line), relative humidity (blue line), (wind speed red line) and precipitation intensity (dark blue bar graph). On the horizontal axis we have the dates corresponding to the date of the last day of the week, on the first Y axis on the left, we have the values of temperatures, humidity percentage, wind speed expressed in millimeters per second while the precipitation expressed in millimeters is represented on the second Y axis on the right.

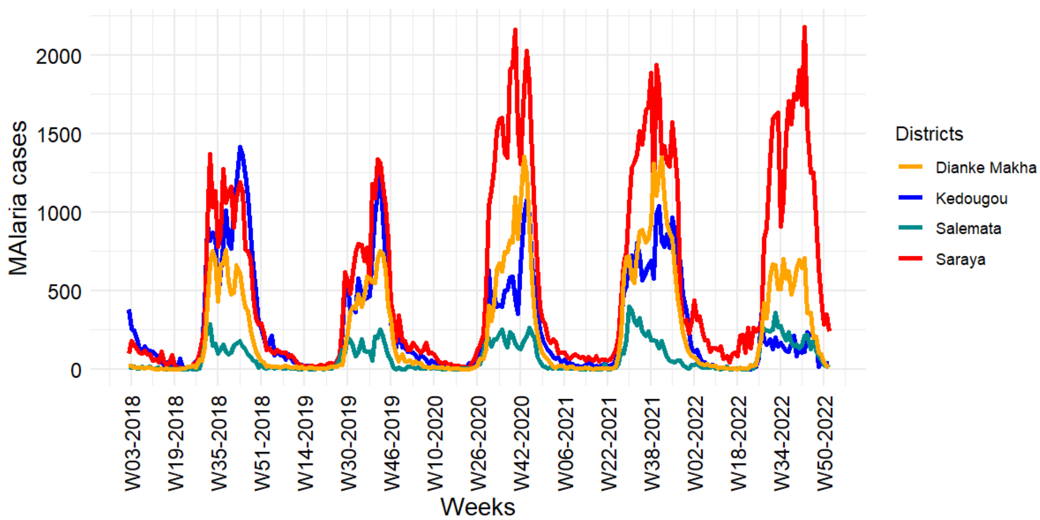


Fig. 3. Weekly evolution of malaria cases confirmed by rapid diagnostic tests in the districts of Kédougou, Saraya, Salémata, and Dianké Makha (2018-2022). The curves show seasonal variations, with X-axis labels indicating the weeks of the year.

cases) and explanatory variables. The general form of the *GAM* used in this study is:

$$g(\mu_i) = \beta_0 + \beta_1 * Pop + \sum_{j=1}^{p} f_j(X_{ij}) + s(Month) + s(year) + log(test), \qquad (1)$$

where

- $g(\mu_i)$ is the link function;
- f_j represents smooth functions of the explanatory variables X_{ij} modeled using regression spline;
- $s(Month)$ is a cyclic spline function used to capture seasonal effect;
- $s(year)$ is a cubic spline function used to capture trend effect;
- $log(test)$ is an offset term representing the logarithm of the number of diagnostic tests performed each week.

To reduce the number of explanatory variables and avoid multicollinearity between meteorological variables, we constructed synthetic meteorological indicators using principal component analysis (*PCA*) of weekly meteorological variables. *PCA* helps preserve the main meteorological characteristics while summarizing covariates into fewer uncorrelated synthetic indicators called principal components (*PCs*) Jolliffe (2002). *PCs* were selected using the Kaiser criterion, retaining components with eigenvalues greater than one RAKOTOMALALA (2023). In order to find the best predictors of weekly malaria cases, we explored three different variable selection methods. Firstly, we estimated a Generalized Additive Model using the selected principal components as predictors. This approach helped reduce dimensionality and account for correlations among meteorological variables while preserving most of the variance in the original data. Secondly, for each principal component, we selected the variable that contributed the most and estimated another *GAM* model using these selected variables. This allowed us to test the impact of specific meteorological factors while reducing the complexity of the model. Finally, we tested the effect of interactions, assuming that the effects of the main components on malaria may change depending on the global climate regime. The climate regimes (or clusters) were determined by applying Hierarchical Ascending Classification to the *PCA* results Husson *and al.*(2010). A third *GAM* model was estimated within this framework. *HAC* segments time into clusters of climate regimes, which can help understand how transitions between clusters (*i.e.*, climate regimes) influence malaria incidence. The existence of significant differences of the components according to the climatic regimes was tested using the *ANOVA* test and Tukey's honestly significant difference test Aaron Sc (2025). To account for the delayed impact of meteorological factors on malaria transmission, we introduced time lags ranging from 0 to 15 weeks for each explanatory variable, following the approach of Guo et al. Guo *et al.*(2015). This range was chosen based on previous studies and the biological life cycle of malaria, which suggest that environmental conditions influence mosquito abundance and malaria transmission with varying temporal delays Diao *and al.*(2023), Fall *and al.*(2023b), Cissoko *and al.*(2020). In total, 4,096 models were constructed by testing different combinations of lagged variables to capture potential delayed effects.The models were evaluated using the Akaike Information Criterion (*AIC*) score to balance model fit and complexity. The best-fitting

model was selected based on the lowest *AIC* value. Odds Ratios (*ORs*) were calculated for the best-fitting model, accounting for the non-linear relationships between the response and predictor variables. To capture these non-linear effects, we estimated *ORs* across the deciles of each predictor's value distribution. This approach allowed us to better understand how malaria risk varied across different exposure levels while reflecting the complexity of the underlying associations R Core Team (2020). Our model was fitted using data from 2018 to 2021 for each district and tested on 2022 data to assess its predictive performance. The quality of the predictions was evaluated using criteria such as the Root Mean Square Error (*RMSE*) and the adjusted coefficient of determination. These metrics helped quantify the gap between predicted and observed values, as well as the model's ability to capture the temporal dynamics of malaria incidence.
All statistical analyses were conducted using **R** software (version 4.5.1) R Core Team (2022).

3. Results

3.1. Synthetic meteorological indicators

The *PCA* carried out using Kaiser's criterion led us to retain three synthetic meteorological indicators (*PC*) that explained 86.2% of the total inertia. The first *PC* explained 48.8% of total inertia and was primarily associated with relative humidity, specific humidity, and rainfall frequency (negatively correlated), as well as maximum temperature and wind speed (positively correlated). The second *PC* (24.8% of total inertia) represented minimum and average temperature, wind direction, and atmospheric pressure. The third component (12.6% of total inertia) was mainly influenced by rainfall (Figure 4(a)). These synthetic indicators were used as explanatory variables in the first *GAM* model to describe the variation in malaria cases.For the first principal component, we selected the variable *RH2M* (relative humidity), which shows a strong contribution (12.9%) with a correlation coefficient of -0.95 (Figure 4(b)). Regarding the second PC, the variable T2_MIN (minimum temperature) shows the most significant contribution (29.2%) with a correlation coefficient of 0.92 (Figure 4(c)).For the third principal component, the variable PRECTOCORR (rainfall intensity) contributes the most (19%) with a correlation coefficient of 0.59 (Figure 4(d)). *RH2M*, T2_MIN and PRECTOCORR were used to estimate a second *GAM* model.

3.2. Climatic regimes

Following the Principal Component Analysis, a Hierarchical Clustering Analysis (*HCA*) was applied to group the observations based on their main characteristics. This approach led to the identification of three climatic regimes (clusters), each reflecting specific profiles of variation in weather conditions (Figure 5(a)). The first cluster is observed during the months of June to November (Figure 5(b)) and is characterized by high humidity, lower temperatures, decreased wind speed, and high rainfall. Cluster 2 is observed during the months of November to February and in March, and is characterized by variable humidity levels. Cluster 3 is ob-

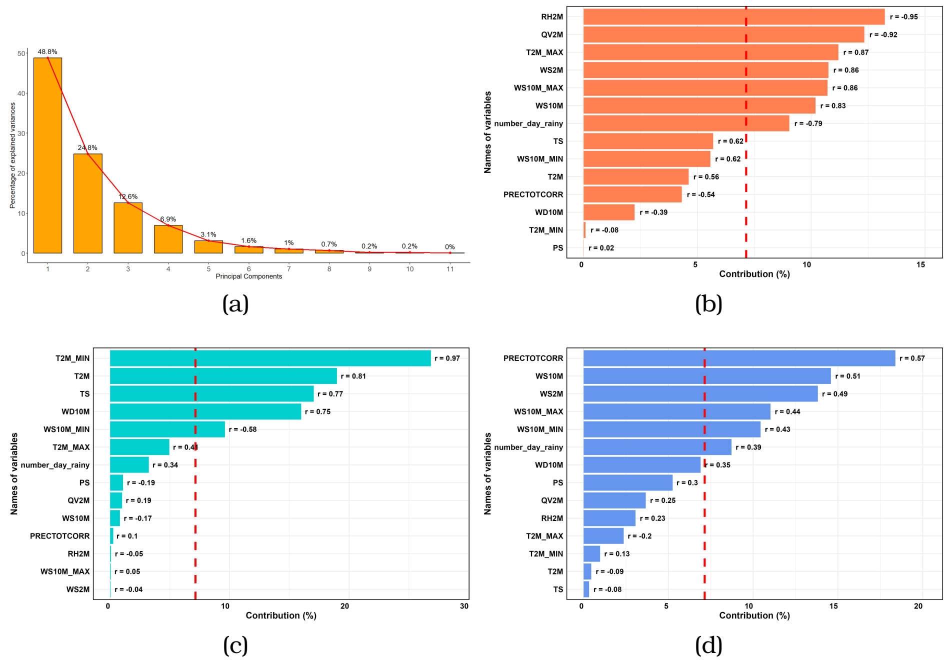


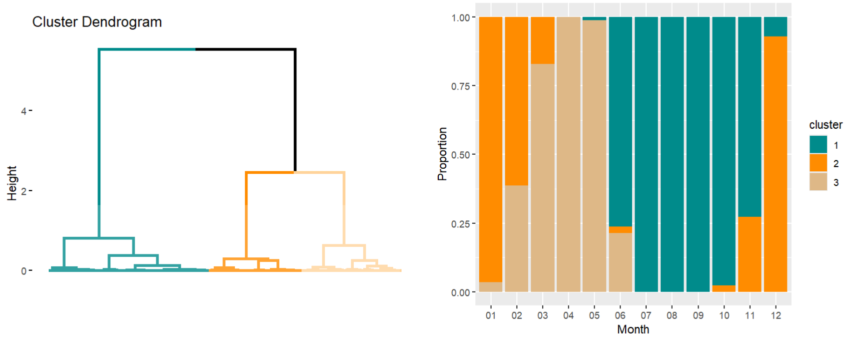
Fig. 4. Principal component analysis of meteorological variables. The percentage of inertia explained by each principal component (a).The horizontal red line indicates the Kaiser criterion of a normalized *PCA* according to which the components associated with a value greater than one were retained. The contribution (in %) of the meteorological variables on the first principal component (b), the second principal component (c) and the third (d). The r indicates the correlation coefficient between the meteorological variable and the synthetic indicator. The dotted red line represents the contribution that would have been expected if all the variables had contributed equally to the synthetic variable.

served from March to May; it represents a dry period with very high temperatures ranging from 35°C to 45°C. Tukey's test shows that the identified groups have significantly different climatic profiles. We can examine the variation of the principal components (*PCs*) in relation to the identified climatic regimes in Figure 6. Cluster 1 has negative coordinates on the first principal component (Dim1), cluster 2 has negative values on the second *PC* (Dim2), and cluster 3 has positive coordinates on the third dimension. Although the statistical test showed a significant difference between Dim 3 and the clusters, it is not possible to clearly identify this difference in the graphs. It is essential to test, using a third model, the interaction effect between component 3 (Dim 3) and the climatic regimes (clusters) in order to better understand the complex relationships that influence malaria incidence.

3.3. Time Lag Analysis and Model Selection

The first optimal model included:

- The first and second *PCs* lagged by 2 weeks;
- The third *PC* lagged by 15 weeks;
- A seasonal term;
- A trend term;



(a) (b)

Fig. 5. Dendrogram illustrating the classification of meteorological conditions based on *PCA* analysis of meteorological variables. This representation helps to visualize the relationships and similarities among different weather conditions (e) and a bar graph for locating each meteorological period(cluster) in time on a monthly scale, providing a temporal context for the identified weather patterns(f).

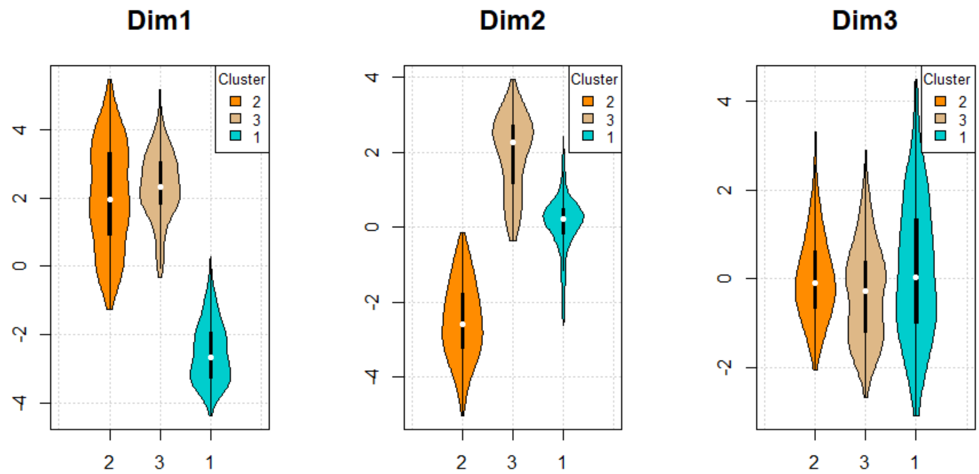


Fig. 6. Variation of the principal components according to the three meteorological periods represented by the clusters. The first (left) represents the variation of the first principal components, the second (center) represents the variation of the second principal component, the third (right) represents the variation of the third principal component.

- Population size;
- Random effect of health districts;
- Offset of the number of rapid diagnostic tests.

This model explained 92.2% of the variance in malaria incidence, with an explained deviance of 83.8%. Both the first and second *PCs* had significant non-linear effects on malaria dynamics ($p < 2 \times 10^{-16}$), while the third *PC* did not show a significant direct effect ($p = 0.81$) but improved the estimation of seasonality. The seasonal and trend terms were significant ($p < 2 \times 10^{-16}$). A second optimal model constructed using the meteorological variables that contributed most to the variation in each *PC* included:

- A 4-week lag for relative humidity (*RH2M*);
- A 14-week lag for minimum temperature (T2M_MIN);
- A 15-week lag for rainfall intensity (PRECTOTCORR);
- A seasonal term;
- A trend term;
- Population size;
- Random effect of health districts;
- Offset of the number of rapid diagnostic tests.

This model explained 91.9% of the variance, with an explained deviance of 84.4%, but failed to capture the trend variation in malaria incidence.Variable Month of the year shows a strong and highly significant nonlinear effect ($p < 2 \times 10^{-16}$), confirming the seasonal nature of malaria transmission in Senegal. Relative humidity at a 4-week lag (s(lag4_RH2M)) have a significant and nonlinear influence ($p < 2 \times 10^{-16}$). Minimum temperature at a 14-week lag (s(lag14_T2M_MIN)) is also significant ($p < 2 \times 10^{-16}$), suggesting that lower temperatures several weeks prior can impact malaria transmission with a delayed effect. In contrast, precipitation at a 15-week lag (s(lag15_PRECTOTCORR)) does not show a significant effect in this model ($p = 0.509$). The smooth term is nearly flat, indicating no clear relationship detected between rainfall at this lag and malaria incidence.Variable Year shows a statistically significant effect in this model ($p = 0.003$), reflecting a temporal trend in malaria incidence. However, this effect is more pronounced in the first approach, where the p-value ($p < 2 \times 10^{-16}$) is even smaller, suggesting that the temporal trend is more strongly captured when the model includes synthetic variables (*PCs*). The last model was constructed using *PCs* and an interaction effect between the third PC and clusters. The best model for this approach included:

- The first synthetic *PC* (Dim1);
- The second synthetic *PC* (Dim2);
- An interaction effect of the third synthetic *PC* (Dim3) by each cluster;
- A seasonal term;
- A trend term;
- Population size;
- Random effect of health districts;
- Offset of the number of rapid diagnostic tests.

This model explained a large portion of the variation ($R^2_{adj} = 0.921$, deviance explained $= 83.9\%$). Both the first and second *PCs* with 2-week lags had significant nonlinear effects on malaria dynamics ($p < 2 \times 10^{-16}$) (Figures 7(a) and 7(c)). Regarding the interaction between the 15-week lag of the third component and the clusters, the results suggest a possible association. There is a trend towards a significant relationship between lag15_dim3 and cluster 1 (Figure 7(e)). Based on comparison criteria, this last model is chosen for the continuation of the analysis as it better explains the variation in the data.

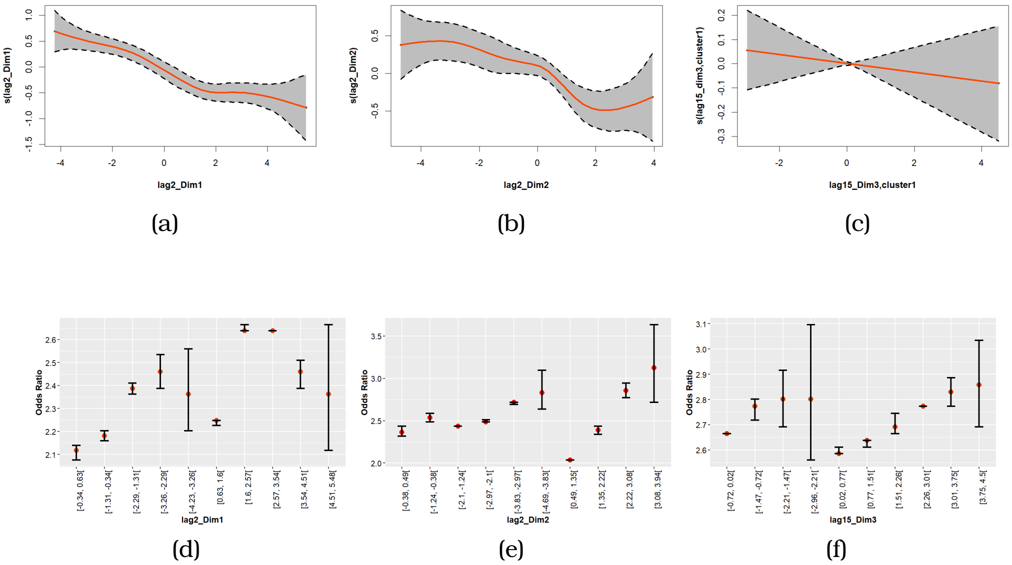


Fig. 7. Relationship between the number of malaria cases and the first synthetic indicator (a), the second (b) and the third synthetic indicator in cluster 1(c), with confidence intervals (shaded area). On the X axis, have the values of the predictor and on the Y axis, we have the effects of prediction on response. Odds ratios with their corresponding 95% confidence intervals represented as error bars. Each odds ratio indicates the strength of association between the first component(d), the second(e) and the third(f) and the malaria cases. The error bars illustrate the uncertainty around the odds ratios, with the endpoints representing the lower and upper bounds of the confidence intervals. This visualization aids in understanding the variability and reliability of the estimated odds ratios across different conditions. Bar chart illustrating the odds ratios associated with changes over different periods. The X axis represents the various periods of change, while the Y axis displays the corresponding odds ratios. Each bar is annotated with the percentage increase or decrease in malaria cases associated with the respective change

3.4. Odds ratio analysis

Figure 7(b) presents odds ratios (*OR*) for the variable lag2_dim1 across different intervals, indicating the relationship between lag2_dim1 and malaria incidence. The lag2_dim1 is negatively associated with malaria incidence in most intervals, with odds ratios consistently below 1. This suggests that higher values of lag2_dim1 (which likely represent certain meteorological variables, perhaps related to temperature or humidity) are associated with a decreased likelihood of malaria incidence. In the interval $[-4.23, -3.26]$, the odds ratio is 0.86, with a confidence interval ranging from 0.79 to 0.94, indicating a significant reduction in malaria risk (since the odds ratio is less than 1). In intervals such as $[-0.34, 0.63]$ and $[0.63, 1.6]$, the odds ratios are slightly higher but still less than 1, indicating a relatively reduced risk, but not as strongly as in the lower ranges of lag2_dim1. However, as lag2_dim1 increases in the higher positive intervals, the odds ratios hover around 0.86 and even show a wide confidence interval in the last interval, suggesting a decreasing protective effect or a shift towards a less significant protective impact. In Figure 7(d), the variable lag2_Dim2, which represents a principal component of meteorological variables lagged by two weeks, shows a nonlinear association with malaria risk. In the middle range of its distribution (deciles 3 to 8), the odds ratios are consistently below 1, indicating a protective effect—malaria risk tends to decrease as lag2_dim2 increases within this range. This suggests that moderate values of the underlying climatic conditions captured by this component may help reduce transmission. However, in the highest two deciles, specifically from values above 2.22, the odds ratios rise above 1, reaching 1.14 in the last decile, with confidence intervals indicating a statistically significant increase in risk. This suggests that very high values of lag2_Dim2 are associated with elevated malaria risk, possibly reflecting extreme climatic conditions that favor transmission. At the lowest end (first decile), the odds ratio is slightly above 1, but the wide confidence interval ($0.97 - 1.13$) suggests no clear effect. Understanding the meteorological variables composing this principal component is essential to interpret the mechanisms behind this relationship. Moreover, moderate to high Dim3 values 15 weeks before the observation period appear to have a moderate protective effect on malaria risk, with significant risk reductions observed in the first deciles (50-70%). However, very high rainfall at the 15-week lag increases malaria risk in higher deciles (80-100%), suggesting that excessive rainfall may be detrimental, promoting mosquito proliferation and increasing malaria transmission (Figure 7(f)).

3.5. Prediction of Malaria Cases

The predictive capability of the models was assessed by generating forecasts for 2022 based on data from 2018 to 2021. The predictions were overlaid on the observed time series of malaria incidence (Figure 8). The clustering *PCA*-based model demonstrated higher predictive performance, accurately capturing the timing and magnitude of peak malaria incidence.

Discussion

This study examined the relationship between meteorological factors and malaria incidence in four districts of southeastern Senegal using a generalized additive model and time-lagged explanatory variables derived from principal component analysis (*PCA*). The findings revealed significant associations between malaria incidence and key meteorological variables, particularly humidity and temperature according to the three periods obtained from clustering.

Our analysis identified a strong correlation between malaria incidence and the first two principal components, which primarily represent humidity, rainfall frequency, and temperature. The lag of 2 weeks observed for these components suggests that meteorological conditions during a preceding two-week period are critical in driving malaria transmission. This lag period aligns with the biological life cycle of the malaria vector (Anopheles mosquitoes) and the Plasmodium parasite, from mosquito development to the onset of clinical symptoms in humans. The lack of a direct significant effect of the third component, which primarily represents rainfall intensity, may indicate that while rainfall intensity influences mosquito breeding conditions, its effect on malaria incidence is indirect and likely mediated through other factors such as humidity and standing water availability. However, including the third component improved the model's estimation of seasonality, highlighting the importance of cumulative rainfall in

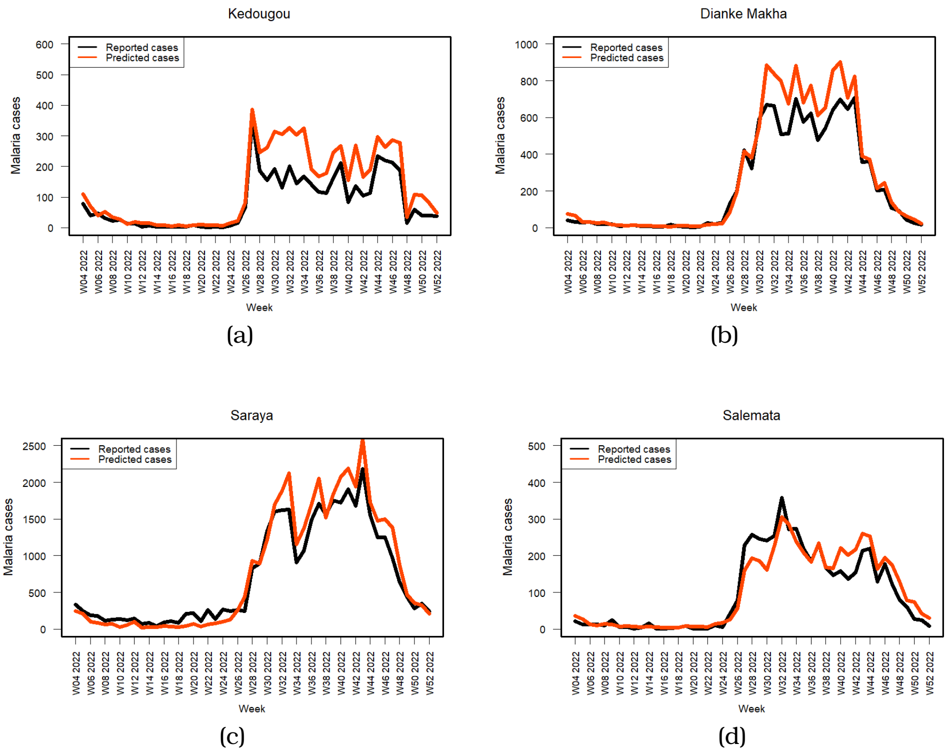


Fig. 8. Overlay of malaria cases time series (black) overlaid with model-predicted data with principal components and clusters(orange). we compared malaria cases time series in 2022 and malaria cases time series predicted by the model in Kédougou (o), Dianké Makha (p), Saraya (q) and Salémata (r) districts.

sustaining transmission over time. These findings are consistent with previous studies conducted in Senegal and other West African countries, which have reported lagged relationships between rainfall, humidity, and malaria incidence Cissoko *and al.*(2020), Nigussie *and al.*(2023), Bationo *and al.*(2021).

Our results further demonstrate that high temperatures, when combined with increased humidity, are conducive to mosquito survival and parasite development, which corroborates findings from similar ecological settings in neighboring regions Dieng *and al.*(2020b). Using clustering on *PCA*, we observed a direct effect of meteorological variables according to the three periods.
We compared three modeling approaches: the first using *PCA*-derived synthetic indicators, the second using individual meteorological variables selected based on their contribution to the principal components, and the third using cluster *PCA*-derived synthetic indicators. While both models showed good explanatory power, the cluster *PCA*-based model was superior in capturing the seasonal and trend patterns of malaria incidence. This suggests that clustering on *PCA*, by summarizing correlated meteorological variables into fewer components, may

enhance model robustness and reduce the risk of multicollinearity affecting parameter estimates. Despite the slightly lower explained variance (91.9%) of the alternative model, its performance in predicting malaria incidence highlights the potential utility of simpler models that use individual meteorological variables. These models may be preferable in settings where computational efficiency or interpretability is a priority.

The identification of specific meteorological factors influencing malaria transmission provides valuable insights for improving malaria surveillance and early warning systems. By incorporating meteorological forecasts into predictive models, health authorities can better anticipate malaria peaks and implement targeted interventions, such as seasonal malaria chemoprevention (*SMC*) or indoor residual spraying (*IRS*), at optimal periods. Our findings also highlight the importance of maintaining long-term meteorological and health surveillance data. Continuously updated datasets would allow for the refinement of predictive models and improve the accuracy of malaria forecasting in the context of climate variability and change.

While this study provides important insights into the relationship between meteorological factors and malaria incidence, several limitations should be noted. The study focuses on four districts in southeastern Senegal, which limits the generalizability of the findings to other regions with different ecological and climatic conditions. Future studies should extend this analysis to other malaria-endemic regions in Senegal and neighboring countries. Furthermore, although *PCA* reduced the multicollinearity between explanatory variables, it may have obscured potential interaction effects between individual meteorological factors, such as rainfall and temperature. Future research could explore more complex models that incorporate interaction terms to better capture the combined effects of key environmental drivers on malaria transmission. Incomplete data for 2022 may also affect the accuracy of predictions, limiting the model's ability to capture recent trends in malaria incidence. While the *GAM* approach allowed for flexible modeling of non-linear relationships, alternative machine learning approaches, such as random forests or neural networks, could be explored to further improve predictive accuracy and identify complex patterns in the data.

A major constraint is the lack of entomological data, including information on Anopheles species composition and their biting behavior. Since malaria transmission is strongly influenced by mosquito biting habits, the absence of such data limits our understanding of transmission dynamics. Furthermore, the study does not incorporate detailed cartographic analyses, which would have allowed for a more precise spatial representation of malaria risk areas and potential hotspots. Moreover, hidden ecological factors that can influence climatic variables, such as changes in land use, deforestation, and urbanization, were not explicitly considered, which could affect the precision of the associations observed between climate and malaria incidence.

Conclusion

This study highlights the significant influence of meteorological factors on malaria incidence in the southeastern region of Senegal, using generalized additive models with time-lagged explanatory variables derived from principal component analysis. Humidity and temperature were identified as key determinants, with a two-week lag corresponding to the life cycle of the vector and parasite. The *PCA*-based approach helped to better capture the seasonality of malaria and improved the robustness of the model. These results emphasize the importance of integrating weather forecasts into early warning systems to optimize malaria prevention and control strategies. However, future research is needed to refine the predictions by incorporating other environmental and socio-economic factors.

Acknowledgments

We would like to thank the Bill & Melinda Gates Foundation for funding, as well as the Ministry of Health and the National Malaria Control Program (*NMCP*) for facilitating data collection. The authors gratefully acknowledge the support of the *WAMCAD* project. The authors are also grateful to Dr Maryam Diarra, Dr Ousmane Sy, Dr Fatoumata Bintou Sall, Dr Lucien Diegane Gning, Khady Ndiaye and Almamy Youssef Ly for their valuable insights and constructive feedback throughout this research. We would like to express our sincere gratitude to the editorial board, the associate editor, and the anonymous reviewer for their valuable time, constructive feedback, and insightful comments that have greatly improved the quality of our manuscript.

Funding

This study was funded by a grant from the Gates Foundation (INV-047051). The sponsors had no role and no influence on the design and interpretation of the data collected, as well as on the writing of the manuscript.

Competing interests

The authors declared that they had no potential conflicts of interest with respect to the investigation, authorship or publication of this article.

Author contribution

L.N.B., M.B., and J.L.N. designed research; L.N.B., M.A.L, M.D, M.A, L.N., K.N, and A.Y.L conducted research; J.L.N., and N.D Provision of study materials; L.N.B. analyzed data; L.N.B., M.A.L, M.M.A, M.D, O.S, and M.B. wrote the paper. L.N.B. and M.A.L had the primary responsibility for the final content. All authors read and approved the final manuscript.

References

Aaron Sc, 2025. Post Hoc Analysis: Tukey. https://rpubs.com/aaronsc32/post-hoc-analysis-tukey

Bationo, C.S., Gaudart, J., Dieng, S., Cissoko, M., Taconet, P., Ouedraogo, B., Some, A., Zongo, I., Soma, D.D., Tougri, G., and others, 2021. Spatio-temporal analysis and prediction of malaria cases using remote sensing meteorological data in Diebougou health district, Burkina Faso, 2016–2017. *Scientific Reports*, 11(1), 20027.

Cissoko, M., Sagara, I., Sankare, M.H., Dieng, S., Guindo, A., Doumbia, Z., Allasseini, B., Traore, D., Fomba, S., Bendiane, M.K., and others, 2020. Geo-epidemiology of malaria at the health area level, dire health district, Mali, 2013–2017. *International Journal of Environmental Research and Public Health*, 17(11), 3982.

Diao, O., Absil, P.A., and Diallo, M., 2023. Generalized Linear Models to Forecast Malaria Incidence in Three Endemic Regions of Senegal. *International Journal of Environmental Research and Public Health*, 20(13), 6303.

Dieng, S., Ba, E.H., Cisse, B., Sallah, K., Guindo, A., Ouedraogo, B., Piarroux, M., Rebaudet, S., Piarroux, R., Landier, J., and others, 2020. Spatio-temporal variation of malaria hotspots in Central Senegal, 2008–2012. *BMC Infectious Diseases*, 20, 1–13.

Dieng, S., Ba, E., Sallah, K., Guindo, A., Piarroux, M., Rebaudet, S., Sokhna, C., and Gaudart, J., 2020. Geo-epidemiology of malaria in the central region of Senegal: spatio-temporal instability of high-risk areas and associated factors. *Review of Epidemiology and Public Health*, 68, S99.

Diouf, M., Diouf, E.H., Niang, E.H.A., Diagne, C.T., Konate, L., and Faye, O., 2018.Evaluation of the physical condition and biological efficacy of two types of long-lasting insecticidal nets used for 5 to 36 months and collected from 11 districts in Senegal. *Bulletin of the Society of Exotic Pathology*, 111(2), 126.

Diouf, I., Rodriguez-Fonseca, B., Deme, A., Caminade, C., Morse, A.P., Cisse, M., Sy, I., Dia, I., Ermert, V., Ndione, J.A., and others, 2017. Comparison of malaria simulations driven by meteorological observations and reanalysis products in Senegal. *International Journal of Environmental Research and Public Health*, 14(10), 1119.

Fall, P., Diouf, I., Deme, A., and Sene, D., 2022. Assessment of climate-driven variations in malaria transmission in Senegal using the VECTRI model. *Atmosphere*, 13(3), 418.

Fall, P., Diouf, I., Deme, A., Diouf, S., Sene, D., Sultan, B., Famien, A.M., and Janicot, S., 2023. Bias-corrected CMIP5 projections for climate change and assessments of impact on Malaria in Senegal under the VECTRI Model. *Tropical Medicine and Infectious Disease*, 8(6), 310.

Fall, P., Diouf, I., Deme, A., Diouf, S., Sene, D., Sultan, B., and Janicot, S., 2023. Enhancing Understanding of the Impact of Climate Change on Malaria in West Africa Using the Vector-Borne Disease Community Model of the International Center for Theoretical Physics (VECTRI) and the Bias-Corrected Phase 6 Coupled Model Intercomparison Project Data (CMIP6). *Microbiology Research*, 14(4), 2148–2180.

Gadiaga, T., Ba, M.F., Sagna, S., Cisse, B., Sène, D., Sarr, S.C., Gueye, B., Thiam, S., Cire, E.B.K., and Ndiaye, J.L.A., 2024. Assessment of the efficiency of the routine epidemiological surveillance system for malaria at the Tambacounda Health District, Senegal. *The Pan African Medical Journal*, 48, 41.

Guo, C., and others, 2015. Malaria incidence from 2005–2013 and its associations with meteorological factors in Guangdong, China. *Malaria Journal*, 14, 1–12.

Hay, S.I., Smith, D.L., and Snow, R.W., 2008. Measuring malaria endemicity from intense to interrupted transmission. *The Lancet Infectious Diseases*, 8(6), 369–378.

Husson, F., Josse, J., and Pagès, J., 2010. Principal component methods–hierarchical clustering. partitional clustering. *Why would we need to choose for visualizing data*.

Husson, F., Josse, J., Le, S., Mazet, J., and Husson, M.F., 2016. Package 'factominer'. *An R package*, 96(96), 698.

Ilunga-Ilunga, F., Leveque, A., and Dramaix, M., 2016. Influence of the age and the level of transmission on the clinical and biological expression of severe malaria in children. *Archives de Pediatrie: Organe Officiel de la Societe Francaise de Pediatrie*, 23(5), 455–460.

Jolliffe, I.T., 2002. *Principal Component Analysis for Special Types of Data*. Springer.

Lancelot, R. and Lesnoff, M., 2005. Selection de modèles avec l'AIC et critères d'information derives. *CIRAD, Montpellier*.

Liu, H., 2008. Generalized additive model. *Department of Mathematics and Statistics University of Minnesota Duluth: Duluth, MN, USA*, 55812.

Manga, I.A., Tairou, F., Seck, A., Kouevidjin, E., Sylla, K., Sow, D., Gueye, A.B., Ba, M., Ndiaye, M., Tine, R.C.K., and others, 2023. Effectiveness of seasonal malaria chemoprevention administered in a mass campaign in the Kédougou region of Senegal in 2016: a case-control study. *Wellcome Open Research*, 7, 216.

Meibalan, E. and Marti, M., 2017. Biology of malaria transmission. *Cold Spring Harbor Perspectives in Medicine*, 7(3), a025452.

Midekisa, A., Beyene, B., Mihretie, A., Bayabil, E., and Wimberly, M.C., 2015. Seasonal associations of climatic drivers and malaria in the highlands of Ethiopia. *Parasites & Vectors*, 8, 1–11.

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC), 2024. Giovanni: A Web Application for Interactive Exploration of NASA Earth Science Data. https://giovanni.gsfc.nasa.gov

National Agency for Statistics and Demography of Senegal. *National Agency for Statistics and Demography of Senegal*. https://www.ansd.sn

National Malaria Control Programme (NMCP), 2023. *Epidemiological Bulletin of Malaria in Senegal Year 2022*.

Nigussie, T.Z., Zewotir, T.T., and Muluneh, E.K., 2023. Seasonal and spatial variations of malaria transmissions in northwest Ethiopia: Evaluating climate and environmental effects using generalized additive model. *Heliyon*, 9(4).

Olliaro, P.L., Delenne, H., Cisse, M., Badiane, M., Olliaro, A., Vaillant, M., and Brasseur, P., 2008. Implementation of intermittent preventive treatment in pregnancy with sulphadoxine/pyrimethamine (IPTp-SP) at a district health centre in rural Senegal. *Malaria Journal*, 7, 1–11.

RAKOTOMALALA, R., 2023. Principal Component Analysis (PCA). *Principles and Practice of PCA, Lumière University Lyon*, 2.

R Core Team, 2020. or_gam: Calculate Odds Ratios of Generalized Additive (Mixed) Models. https://rdrr.io/cran/oddsratio/man/or_gam.html

R Core Team, 2022. R: A language and environment for statistical computing. Version 4.2.0. Vienna, Austria: R Foundation for Statistical Computing.

Regional Committee for Africa, 2021. Status report on the implementation of the regional strategy and regional strategic plan for neglected tropical diseases control 2014-2020 in the African Region: information document. World Health Organization. Regional Office for Africa.

Seck, M.C., Thwing, J., Fall, F.B., Gomis, J.F., Deme, A., Ndiaye, Y.D., Daniels, R., Volkman, S.K., Ndiop, M., Ba, M., and others, 2017. Malaria prevalence, prevention and treatment seeking practices among nomadic pastoralists in northern Senegal. *Malaria Journal*, 16, 1–11.

Siraj, A.S., Bouma, M.J., Santos-Vega, M., Yeshiwondim, A.K., Rothman, D.S., Yadeta, D., Sutton, P.C., and Pascual, M., 2015. Temperature and population density determine reservoir regions of seasonal persistence in highland malaria. *Proceedings of the Royal Society B: Biological Sciences*, 282(1820), 20151383.

Stallman, R.M., 2009. *Using the gnu compiler collection: a gnu manual for gcc version 4.3. 3.*

Tairou, F., Nawaz, S., Tahita, M.C., Herrera, S., Faye, B., and Tine, R.C.K., 2022. Malaria prevention knowledge, attitudes, and practices (KAP) among adolescents living in an area of persistent transmission in Senegal: Results from a cross-sectional study. *Plos One*, 17(12), e0274656.

Taye, G., Kaba, M., Woyessa, A., Deressa, W., Simane, B., Kumie, A., and Berhane, G., 2015. Modeling effect of climate variability on malaria in Ethiopia. *The Ethiopian Journal of Health Development*, 29(3).

The President's Malaria Initiative (PMI), 2020. The PMI VectorLink Project. Abt Associates Inc., Rockville, MD.

Thwing, J., Eckert, E., Dione, D.A., Tine, R., Faye, A., Ye, Y., Ndiop, M., Cisse, M., Ndione, J.A., Diouf, M.B., and others, 2017. Declines in malaria burden and all-cause child mortality following increases in control interventions in Senegal, 2005–2010. *The American Journal of Tropical Medicine and Hygiene*, 97(3 Suppl), 89.

Tusting, L.S., Bousema, T., Smith, D.L., and Drakeley, C., 2014. Measuring changes in Plasmodium falciparum transmission: precision, accuracy and costs of metrics. *Advances in Parasitology*, 84, 151–208.

Ugwu, C.L.J. and Zewotir, T., 2020. Evaluating the effects of climate and environmental factors on under-5 children malaria spatial distribution using generalized additive models (GAMs). *Journal of Epidemiology and Global Health*, 10(4), 304–314.

Ukawuba, I. and Shaman, J., 2022. Inference and dynamic simulation of malaria using a simple climate-driven entomological model of malaria transmission. *PLoS Computational Biology*, 18(6), e1010161.

World Health Organization, 2023. *World malaria report 2023*. World Health Organization.

World Health Organization, 2024. DHIS2 Platform. District Health Information Software 2 (DHIS2). https://dhis2.org

Wood, S.N., 2017. *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.

Yamana, T.K. and Eltahir, E.A.B., 2013. Incorporating the effects of humidity in a mechanistic model of Anopheles gambiae mosquito population dynamics in the Sahel region of Africa. *Parasites & Vectors*, 6, 1–10.