Causal Structural Equation Modeling of Climatic and Environmental Factors Influencing Malaria Incidence in Southeastern Senegal

Leontine Ndogou Bakhoum[1,*], Mor Absa Loum[2], Mouhamad Mounirou Allaya[3], Gouvidé Jean Gbaguidi[4], Khady Ndiaye[1], Almamy Youssouf Ly[1], Mamadou Bousso[3] and Jean Louis Abdourahim Ndiaye[5]

[1]*Faculty of Science and Technology, Iba Der Thiam University, Thiès, Thiès, Senegal*

[2]*Faculty of Engineering Sciences, Iba Der Thiam University, Thiès, Thiès, Senegal*

[3]*Faculty of Economic and Social Sciences, Iba Der Thiam University, Thiès, Thiès, Senegal*

[4]*West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL), Faculty of Human and Social Sciences, University of Lomé, Lomé, Togo, Technology, Engineering and Mathematics of Abomey, National University of Science, Benin, Benin*

[5]*Parasitology and Mycology Department, Faculty of Health Sciences, Iba Der Thiam University, Thiès, Senegal*

Abstract: *Introduction*: Malaria remains a major public health problem in tropical regions due to complex interactions between climatic, environmental, and epidemiological processes.

Objective: To improve the quantitative understanding of these mechanisms, this study introduces a causal analysis framework based on structural equation modeling (SEM) to elucidate the direct and indirect pathways through which climatic and environmental variables influence malaria incidence.

Methods: We analyzed weekly malaria case data collected in four districts of southeastern Senegal (Kédougou, Salémata, Saraya, and Dianké Makha) over the period 2018–2024 using structural equation modeling and climates factors. Climate predictors included temperature, precipitation, relative humidity, wind speed, atmospheric pressure, cloud cover, and lunar phase, obtained from Visual Crossing, complemented by vegetation indices derived from satellite imagery.

Results: The results revealed significant spatial heterogeneity: rainfall and cloud cover had the strongest total effects on malaria incidence in three districts (standardized total effects ranging from 0.51 to 0.66), while high wind speed emerged as the primary determining factor in Kédougou, with a negative association (−0.51). Wind-related factors showed protective effects against malaria transmission, with notable spatial complexity, and atmospheric pressure exhibited consistent indirect effects (0.18 to 0.42), mediated by other climatic predictors, but without a significant direct association.

Conclusion: These results underscore the importance of considering indirect causal pathways in environmental epidemiology and highlight district-specific climatic profiles that can inform localized malaria control strategies. The SEM framework presented here offers a rigorous statistical approach to integrating multi-source climatic and environmental data to advance causal inference in vector-borne disease research, ultimately supporting early warning systems and targeted vector control interventions at the district level.

Keywords: Causal modeling, Structural equation modeling, Exploratory factor analysis, Environmental epidemiology, Malaria, Vector-borne disease, West Africa.

1. INTRODUCTION

Malaria remains a major public health challenge, particularly in tropical regions where climatic and environmental conditions significantly influence its transmission dynamics. 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 [1]. In Senegal, malaria is endemic, and its distribution varies considerably across different regions. 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 [2]. Despite the observed effectiveness of many interventions, malaria remains concentrated in the southeastern districts of Kédougou, Salémata, Saraya, and Dianké Makha since 2016 [3]. This part of Senegal has the highest rainfall combined with high temperatures and permanent humidity, creating exceptionally favorable climatic conditions for the reproduction and survival of mosquito vectors.

Previous studies have analyzed climate-malaria relationships using negative binomial models, quasi-Poisson models, ANOVA, and generalized linear models [4-9]. While these approaches have established associations between meteorological parameters and malaria incidence, they assume temporal independence and favor correlational

*Address correspondence to this author at the Faculty of Science and Technology, Iba Der Thiam University, Thiès, Thiès, Senegal;
E-mail: lndogou.bakhoum@univ-thies.sn

analyses that do not allow identification of direct and mediated causal mechanisms [7, 8, 10, 11].

Understanding how climatic factors affect malaria incidence requires a methodological approach capable of capturing complex causal relationships between variables. In malaria epidemiology, climatic variables can exert direct effects on vector survival while simultaneously influencing intermediate environmental conditions that mediate transmission intensity. Structural equation modeling (SEM) addresses these limitations by simultaneously modeling complex multivariate relationships and decomposing total effects into direct and indirect pathways. Unlike generalized linear models (GLM) and time series models which either ignore indirect pathways or focus primarily on temporal autocorrelation SEM incorporates latent variables representing unobserved climatic constructs, making it particularly suited for capturing the multi-pathway nature of climate-malaria relationships.

This study follows a three-step approach: (1) identification of latent climatic and environmental factors through exploratory factor analysis, (2) validation of the measurement model through confirmatory factor analysis, and (3) quantification of direct and indirect effects on malaria incidence using SEM. This framework contributes to advancing causal inference methods in environmental epidemiology and provides an operational basis for early warning systems and adaptive malaria control strategies.

2. MATERIALS AND METHODS

2.1. Study Area

This study focuses on four districts in the Kédougou and Tambacounda regions of Senegal: Kédougou, Salémata, Saraya, and Dianké Makha (Figure **1**). These regions have a long rainy season from May to October and constitute an area of intense and perennial malaria transmission. The selected districts are among the highest malaria risk areas in the country according to reports from the National Malaria Control Program (NMCP) [12]. Table **1** reveals distinct epidemiological and climatic profiles among the four study districts. District of Saraya exhibits the highest malaria burden ($median$ = 191 cases, $mean$ = 529.9 cases, $maximum$ = 2161 cases), followed by districts of Kédougou and Dianké Makha which display intermediate but comparable levels (respective medians of 89 and 29 cases, means of 216.7 and 199.2 cases), while district of Salémata records the lowest transmission ($median$ = 14 cases, $mean$ = 64.5 cases, $maximum$ = 358 cases). From a climatic perspective, all localities share a similar semi-arid rainfall regime characterized by distinct wet and dry seasons (null or very low precipitation medians: 0 − 0.2 mm) but with variable rainfall episodes (maximums ranging from 72.8 to 132.5 mm), stable tropical temperatures (median UV indices = 9), and homogeneous atmospheric pressures (means around 1012 hPa). The most marked differences concern cloud cover, significantly higher in districts of Saraya,

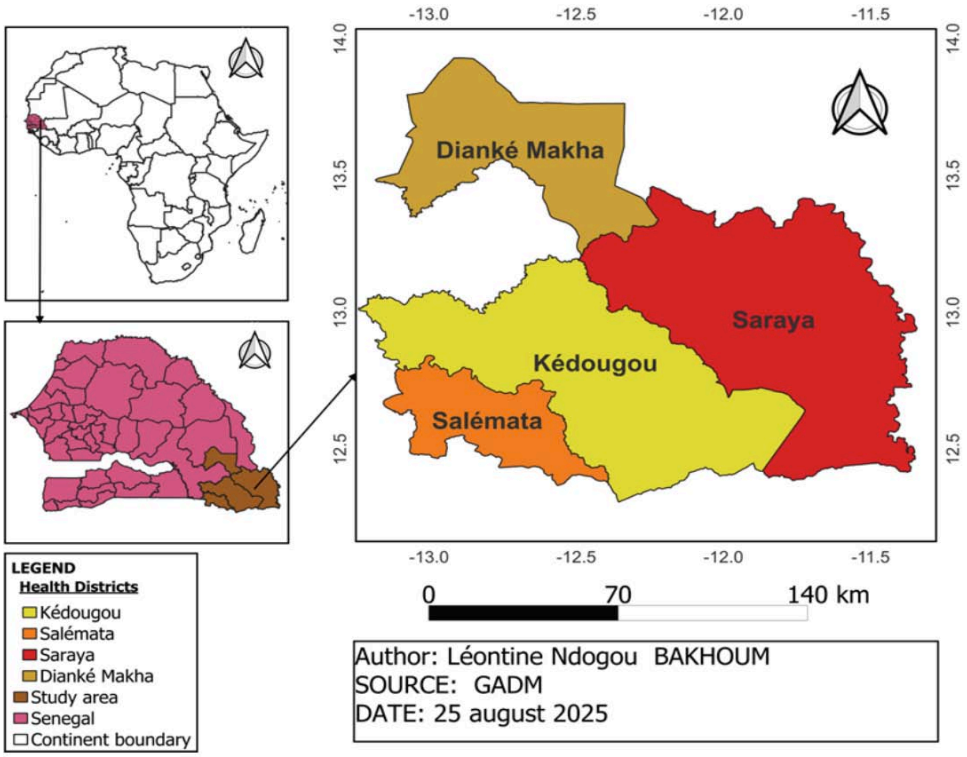


Figure 1: Map of study areas.

Table 1: Summary Statistics of Climate Variables by District

Variables	Kédougou			Saraya			Dianké Makha			Salémata		
	Mean	Med	Max	Mean	Med	Max	Mean	Med	Max	Mean	Med	Max
Malaria cases	217	89	1358	529.9	191	2161	199.2	29	1244	64.45	14	358
Moon_phase_mean	0.48	0.48	0.88	0.48	0.48	0.88	0.47	0.46	0.88	0.48	0.47	0.88
Precip_max	12.46	0.2	130	12.49	0.20	132.54	4.38	0.00	72.80	12.01	0.10	130.00
Precip_cover_max	8.86	4.17	54.17	7.83	4.17	54.17	8.77	0.00	67.26	8.63	4.17	66.67
Wind_gust_max	39.91	40.30	59.40	37.97	37.10	59.40	40.24	40.70	60.10	39.75	40.00	59.40
Wind_speed_max	21.0	20.0	45.1	20.22	19.00	62.30	20.21	20.50	29.90	21.17	22.30	45.70
Sea_level Pressure_max	1012	1012	1016	1012	1012	1016	1012	1012	1016	1012	1012	1016
Cloud_cover_max	74.83	80.50	99.00	75.83	80.65	98.30	80.76	89.50	100.00	74.62	79.80	98.80
Uv_index_max	9.12	9.00	10	9.15	9.00	10	9.16	9.00	10.00	9.107	9.00	10.00
Wind_gust_min	25.58	24.80	43.90	24.27	23.80	40.70	26.19	25.90	39.20	25.57	24.50	42.10
Wind_speed_min	12.27	11.20	22.30	11.49	11.20	20.30	12.13	11.90	20.50	12.51	11.20	22.00
Sea_level_ Pressure_min	1009	1010	1013	1009	1010	1010	1009	1009	1010	1009	1010	1013
Cloud_cover_min	26.63	19.70	81.00	31.99	26.75	81.60	19.59	10.00	76.80	26.65	19.40	80.10

Med = Median; Max = Maximum.

Kédougou, and Salémata (medians: 80.6%, 80.5%, and 79.8% respectively) compared to Dianké Makha (median = 10%), suggesting distinct microclimates that may explain the observed variations in malaria transmission. This spatial heterogeneity of malaria incidence despite globally similar climatic conditions underscores the importance of local factors in vector transmission dynamics.

2.2. Data Collection

This study uses weekly data from 2018 to 2024 from four districts, with confirmed malaria cases as the dependent variable and climatic and environmental variables as explanatory variables. Weekly malaria cases data were obtained from the District Health Information System 2 (DHIS2) [13], representing 364 observations per district. Confirmed malaria cases across the entire population diagnosed by rapid diagnostic tests were extracted by week. Weekly meteorological variables including (mean, max, min) temperature (°C), cumulative precipitation (mm), relative humidity, wind speed, cloud cover, cloud precipitation, lunar phase, and sea level pressure were obtained from Visual Crossing in each district of the study areas [14]. In addition, Advanced Very High-Resolution Vegetation health data including Temperature Condition Index (TCI), Vegetation Condition Index (VCI) and Vegetation Health Index (VHI) were retrieved from NOAA [15]. All datasets were complete over the study period, with no missing observations recorded for either malaria cases or meteorological variables. Regarding potential limitations, Visual Crossing data rely on spatial interpolation between ground-based stations, which may introduce estimation bias in areas with sparse coverage. Vegetation indices at 4 km resolution may also introduce spatial aggregation bias at the district level. Years with incomplete malaria case records were excluded; the final dataset covers 2018–2024, representing a complete series with no missing values prior to merging with meteorological variables.

2.3. Data Preprocessing

Before proceeding with the structural equation modeling analysis, several preprocessing steps were performed to ensure data quality and suitability.

We cleaned our database using the Mahalanobis distance with a threshold of 0.95 to remove outliers. The Mahalanobis distance is a powerful tool for identifying outliers in multivariate data. It accounts for covariance between variables, making it more accurate

than other measures such as Euclidean distance, especially when variables are correlated. A detailed description of the calculations is provided in Appendix (A)

We then performed correlation and collinearity analysis between variables for each district using Pearson's correlation matrix. To ensure the robustness of the multiple linear regression model and prevent estimator instability issues, a systematic evaluation of multicollinearity was conducted using the Variance Inflation Factor (VIF). For more details on the VIF, see Appendix (B).

Finaly, two preliminary tests were performed: Bartlett's test of sphericity which allows verification that correlations between variables are sufficient for dimensional reduction and the KMO (Kaiser-Meyer-Olkin) index which evaluates the adequacy of sampling for factor analysis. Details of the calculations are given in Appendix (C).

2.4. Structural Equation Modeling Framework

The analytical framework of this study followed three sequential steps. First, Exploratory Factor Analysis (EFA) was performed to identify latent climatic and environmental factors from observed variables, reducing data dimensionality and revealing underlying structures. Second, Confirmatory Factor Analysis (CFA) was conducted to validate the measurement model derived from EFA, assessing the reliability and validity of latent constructs. Finally, Structural Equation Modeling (SEM) was applied to quantify the direct and indirect effects of latent climatic and environmental factors on malaria incidence for each district.

2.4.1. Exploratory Factor Analysis

Mathematical Formulation of Exploratory Factor Analysis

Exploratory Factor Analysis is a multivariate statistical technique used to identify the underlying structure of a set of observed variables by discovering latent (unobserved) factors that explain the correlations among these variables. In simple mathematical terms, observed variables are expressed as linear combinations of potential factors to which residual terms are added.

Let X_1, X_2, …, X_j be the observed variables assumed to be linearly related to a reduced number of unobservable factors (latent variables) F_1, F_2, …, F_j, with n < p, according to the system:

$$\begin{cases} X_1 = \beta_{10} + \beta_{11}F_1 + \cdots + \beta_{1n}F_n + \varepsilon_1 \\ X_2 = \beta_{20} + \beta_{21}F_1 + \cdots + \beta_{1n}F_n + \varepsilon_2 \\ \vdots \\ X_p = \beta_{p0} + \beta_{p1}F_1 + \cdots + \beta_{pn}F_n + \varepsilon_p \end{cases} \quad (1)$$

where $\varepsilon_1, \ldots, \varepsilon_p$ represent the residual terms, with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \delta_i^2$. The unobservable factors F_j are independent of each other, with $E(F_j) = 0$ and $Var(F_j) = 1$, also ε_i are independent of each factors F_j. These assumptions constitute the robust prerequisites for the application of structural equation modeling (SEM). The saturation coefficients can be obtained from the covariance and variance of any two observed variables using the following formulas.

Let X_i and X_j be two observed variables for $i \neq j$:

$$X_i = \beta_{i0} + \sum_{k=1}^{N} \beta_{ik}F_k + \varepsilon_i$$

$$X_j = \beta_{j0} + \sum_{k=1}^{N} \beta_{jk}F_k + \varepsilon_j$$

$$E(X_i) = \beta_{i0} + \sum_{k=1}^{N} \beta_{ik}E(F_k) + E(\varepsilon i) = \beta_{i0}$$

Because $E(F_k) = 0$ et $E(\varepsilon_i) = 0$.

$$X_i - E(X_i) = \sum_{k=1}^{N} \beta_{ik}F_k + \varepsilon_i$$

$$Cov(X_i, X_j) = E\left[\left(\sum_{k=1}^{N} \beta_{ik}F_k + \varepsilon_i\right)\left(\sum_{l=1}^{N} \beta_{jl}F_l + \varepsilon_j\right)\right]$$

By expanding,

$$Cov(X_i, X_j) = E\left[\sum_{k=1}^{N}\sum_{l=1}^{N} \beta_{ik}\beta_{jl}F_kF_l + \sum_{k=1}^{N} \beta_{ik}F_k\varepsilon_j + \sum_{l=1}^{N} \beta_{jl}F_l\varepsilon_i + \varepsilon_i\varepsilon_j\right]$$

Under independence assumptions:

$E(F_k\varepsilon_j) = 0$

$E(F_l\varepsilon_i) = 0$

$E(\varepsilon_i\varepsilon_j) = 0$ pour $i \neq j$

$E(F_kF_l) = 0$ pour $k \neq l$

$E(F_k^2) = Var(F_k) = 1$

Thus:

$$Cov(X_i, X_j) = E\left[\sum_{k=1}^{N} \beta_{ik}\beta_{jk}F_k^2\right] = \sum_{k=1}^{N} \beta_{ik}\beta_{jk}E\left(F_k^2\right)$$

$$Cov(X_i, X_j) = \sum_{k=1}^{N} \beta_{ik}\beta_{jk} \quad (2)$$

Variance is a special case of covariance where both variables are the same. i = j

$$Var(X_i) = Cov(X_i, X_i) = E\left[\left(X_i - E(X_i)\right)^2\right]$$

$$X_i - E(X_i) = \sum_{k=1}^{N} \beta_{ik} F_k + \varepsilon_i$$

$$Var(X_i) = E\left[\left(\sum_{k=1}^{N} \beta_{ik} F_k + \varepsilon_i\right)^2\right]$$

By expanding,

$$\text{Var}\ (X_i) = E[\sum_{k=1}^{N}\sum_{l=1}^{N} \beta_{ik}\beta_{il}F_kF_l + 2\sum_{k=1}^{N} \beta_{ik}F_k\varepsilon_i + {\varepsilon_i}^2]$$

Under independence assumptions,

$$Var(X_i) = \sum_{k=1}^{N} {\beta_{ik}}^2 E\left({F_k}^2\right) + E({\varepsilon_i}^2)$$

$$Var(X_i) = \sum_{k=1}^{N} {\beta_{ik}}^2 + {\delta_i}^2 \quad (3)$$

The total variance of each observed variable decomposes into:

- $\sum_{k=1}^{N} {\beta_{ik}}^2$ =1 (The proportion of variance explained by common factors)
- δi² (specific factor + error term)

The communality of the variable X_i is:

$$h_i^2 = \sum_{k=1}^{N} {\beta_{ik}}^2 \quad (4)$$

It represents the proportion of variance of X_i explained by the common factors.

Exploratory Factor Analysis on Malaria Cases, Climate and Environmental Factors

In this study, EFA was employed to identify latent confounding ecological factors present in the set of observed climatic and environmental variables. Before proceeding with exploratory factor analysis, two preliminary tests were performed: Bartlett's test of sphericity which allows verification that correlations between variables are sufficient for dimensional reduction [16] and the KMO (Kaiser-Meyer-Olkin) index to evaluate the adequacy of sampling for factor analysis [17]. The underlying structure of climatic and environmental variables is determined using the principal axis extraction method with oblique rotation (oblimin). The choice of optimal number of factors relies on a multi-criteria approach combining four complementary methods: horn's parallel analysis, Cattell's scree plot, Kaiser's criterion (eigenvalues > 1), VSS (Very Simple Structure) analysis and Likelihood Ratio Test. These methods assume multivariate normality of the data, adequate sample size for reliable eigenvalue estimation, factorial suitability of the variables (KMO >0.5, significant Bartlett's test), and temporal stability of the underlying factor structure. Horn's parallel analysis specifically assumes independence of simulated variables and stability of the factorial structure, while VSS analysis relies on the parsimony principle and simple structure assumption where each variable loads primarily on a single factor. This multiple validation strategy, combined with verification of methodological assumptions, ensures the robustness of the decision regarding optimal data dimensionality. Items (observed variables) were allocated to factors (latent variables) using stringent criteria: maximum loading ≥ 0.50 with cross-loading differences ≥ 0.20 and h_{2i} ≥ 0.60. In other terms, item i is assigned to factor j if

$$\beta_{ij} \geq 0.50\,, \beta_{ij} - \max_{k\neq i}(\beta_{ik}) \geq 0.20 \; and \; h_i^2 \geq 0.60$$

Goodness of Fit Indicators

The evaluation of measurement model fit relies on a set of goodness-of-fit indices: the global significance test (Chi-square), the root mean square error of approximation (RMSEA), the comparative fit index (CFI) and the standardized root mean square residual (SRMR) [18].

$$SRMR = \sqrt{\frac{2\Sigma_{i\leq j}(\sigma_{ij} - \hat{\sigma}_{ij})2}{p(p+1)}} \quad (5)$$

Where σ_{ij} and $\hat{\sigma}_{ij}$ are the observed and predicted covariances, respectively.

$$RMSEA = \sqrt{\frac{\chi^2 - df}{df(n-1)}} \quad (6)$$

$$CFI = 1 - \frac{\max \chi^2_{\text{model}} - df_{\text{model}},0}{\max \chi^2_{\text{baseline}} - df_{\text{baseline}},0} \quad (7)$$

$$TLI = \frac{\frac{\chi^2_{\text{baseline}}}{df_{\text{baseline}}} - \frac{\chi^2_{\text{model}}}{df_{\text{model}}}}{\frac{\chi^2_{\text{baseline}}}{df_{\text{baseline}}} - 1} \quad (8)$$

2.4.2. Confirmatory Factor Analysis

A Confirmatory Factor Analysis is performed on 20% of the data to verify the validity of the measurement model [19, 20]. CFA tests the factor structure previously determined by confirming the specified relationships between observed variables and identified latent variables. Model fit quality is evaluated using RMSEA, SRMR, TLI, and CFI [21]. To ensure model identification, some factor loadings were fixed to values obtained during EFA [22, 23].

2.4.3. Structural Equation Model Specification

Structural equation modeling (SEM) was used to simultaneously analyze relationships between observed and latent variables, accounting for direct and indirect effects while integrating measurement errors and the covariance structure of the data. The use of a structural equation model proves particularly

relevant in this study for several major methodological reasons: SEM allows modeling of latent variables, i.e., important theoretical dimensions that are not directly observable, such as complex climatic components or underlying environmental processes. These dimensions are inferred from several observed variables, enabling a more faithful representation of the studied phenomena. This type of model is particularly suited to the complex relationships characterizing malaria epidemiology, especially when indirect effects or causal chains exist between climatic, environmental, and epidemiological variables. The estimation uses the robust maximum likelihood (MLR) method to handle the non-normality of epidemiological data. Model validation relies on the same fit indices (CFI, TLI, RMSEA, SRMR) supplemented by AIC and BIC information criteria for comparing alternative models. The measurement models are defined from the equation, with variable factor structures according to local characteristics and model identifiably constraints. For certain districts, factor loading parameters were fixed to specific values to resolve identifiably problems. In this study, after EFA and CFA, we consider the system of equations (9) to establish the relationships between climatic and environmental factors and the number of malaria cases in each of the health districts.

$$\begin{cases} F_1(d) = \sum_j \lambda_{1,j}(d) X_j(d) + \delta_1(d) \\ F_2(d) = \sum_j \lambda_{2,j}(d) X_j(d) + \delta_2(d) \\ \vdots \\ F_p(d) = \sum_j \lambda_{p,j}(d) X_j(d) + \delta_p(d) \\ Y^{(d)} = \sum_i \gamma_i(d) F_i(d) + \varepsilon^{(d)} \end{cases} \quad (9)$$

for (d) in {1, 2, 3, 4} (Kédougou, Dianké Makha, Salémata, Saraya) where:

X_j ∈ {precip_max, precip_cover_max, cloud_cover_min, cloud_cover_max,

Sea_level_pressure_min, sea_level_pressure_max, wind_gust_min,

Wind_speed_min, wind_gust_max, wind_speed_max,

moonphase_min, moon_phase_mean, moon_phase_max}

$Y^{(d)}$ = cas_palu_confirme (confirmed malaria cases)

All analyses are performed with R software, version 4.5.1 [24]. Main packages include lavaan, psych and ggplot2.

3. RESULTS

3.1. Exploratory Factor Analysis

Preliminary tests confirm the relevance of exploratory factor analysis for these data. The Kaiser-Meyer-Olkin (KMO) indices range from 0.69 to 0.80, exceeding the recommended threshold of 0.60, indicating that the data is adequately suited for factor analysis. Bartlett's sphericity test reveals highly significant p-values ($p\text{-}value$ < 0.001) for all districts, rejecting the null hypothesis of an identity correlation matrix and confirming the existence of sufficient correlations between variables to justify factor extraction. In addition, the multi-normality test also shows p-values below 0.001, indicating a violation of the multivariate normality assumption. However, this violation is common with large epidemiological datasets and does not invalidate the SEM analysis, given the large sample size (n ≈ 364 observations per district) and the robustness of the estimation method used.

Exploratory factor analysis was conducted separately for the four districts. The optimal number of factors was determined using three complementary methods: Kaiser's criterion, Horn's parallel analysis, and likelihood ratio tests. Results consistently indicated 5 factors for Kédougou, Salémata, and Saraya, confirmed by likelihood ratio tests (p-value < 0.001), and 4 factors for Dianké Makha (p-value > 0.05). These results were further supported by scree plots (Figure **2**). Detailed results by district are presented in Table **2**.

We considered the number of factors obtained and performed an EFA with oblique rotation and the maximum likelihood estimator. Although the chi-square test shows a significant p-value ($p\text{-}value$ < 0.05), suggesting model rejection, the other fit indices reveal a different reality. The RMSEA, RMSA, CFI, and TLI all indicate good fit of the factorial model, while the $\chi 2/df$ ratio also confirms satisfactory fit quality (Table **3**). This apparent contradiction is explained by the chi-square test's excessive sensitivity to sample size: with large samples, this test detects even the smallest deviations between the theoretical model and observed data, leading to systematic rejections of otherwise well-fitted models. The retained factorial structures explain 63.5%, 64.5%, 65% and 70.3% of the total variance for Saraya, Kédougou, Salémata and Dianké Makha respectively, revealing distinct patterns between districts (Appendix D: Table **8**).

3.2. Confirmatory Factor Analysis

The confirmatory factor analysis validates the proposed factorial structure in the four districts. All districts show excellent fit indices. The RMSEA and SRMR indices are below 0.05, while the CFI and TLI indices exceed 0.95. Additionally, the χ^2/df ratio remains below 1.5 across all districts (Table **5**).

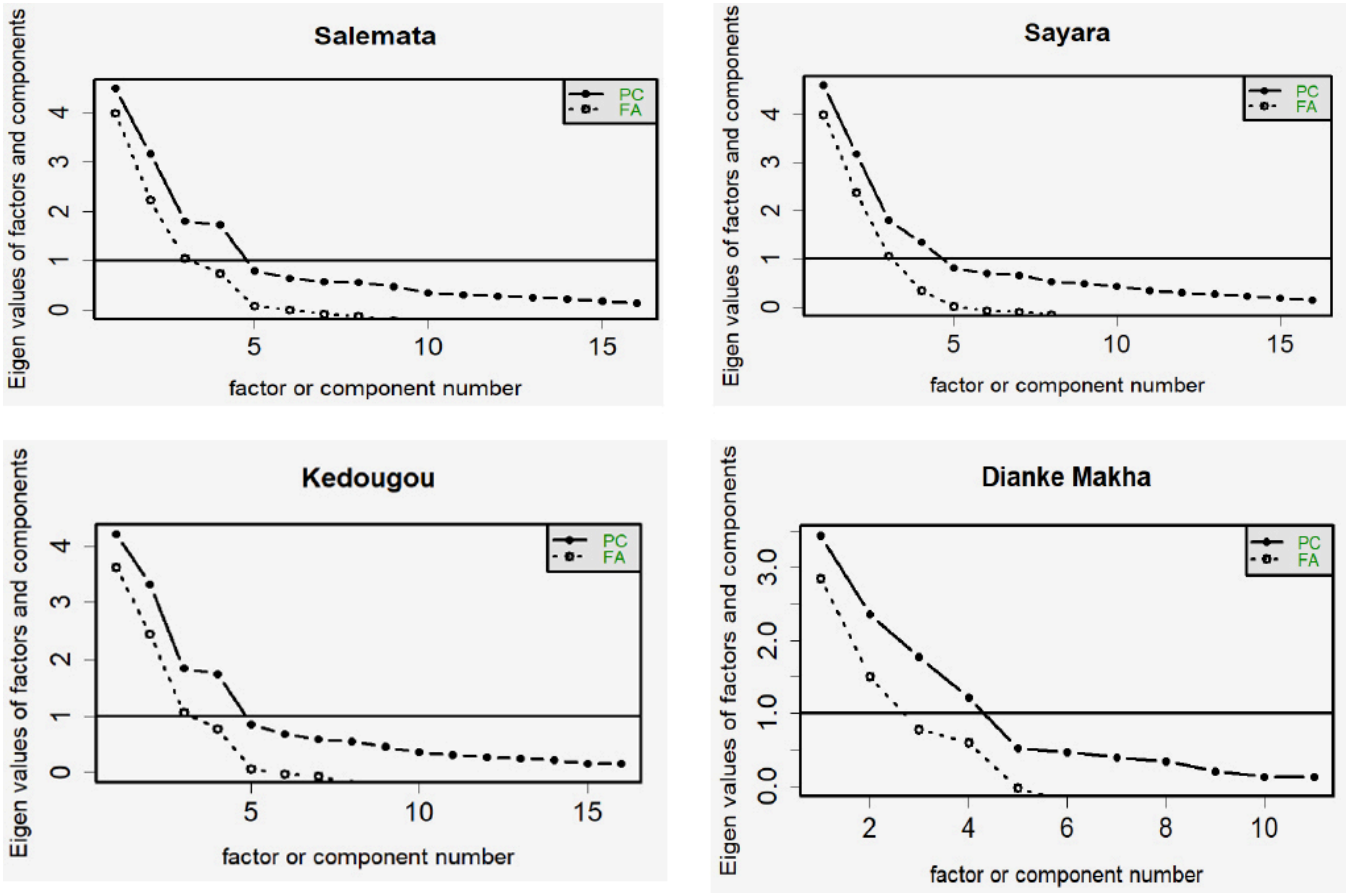


Figure 2: Scree plots of factor analyses by District.

Table 2: Factor Retention Criteria for Exploratory Factor Analysis by District

District	Eigen values						Horn's Parallel Analysis	Models	P- value
	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	N°factors		
Kédougou	4.20	3.31	1.84	1.74	0.85	0.68	5	M5-M4	7.6×10^{-5}
Dianké	3.43	2.36	1.77	1.21	0.52	0.47	4	M5-M4	0.303
Salémata	4.49	3.16	1.79	1.72	0.79	0.64	5	M5-M4	0.0021
Saraya	4.59	3.17	1.79	1.34	0.81	0.70	5	M5-M4	1.64×10^{-5}

Criteria: M5-M4: chi-square difference test comparing 5-factor vs 4-factor models. H_0: the more parsimonious model fits the data as well as the more complex model. Decision rule: p-value ≥ 0.05: reject H_0. p-value > 0.05: fail to reject H_0.

Table 3: Summary of Fit Indices for the Exploratory Factor Analysis Models Across all Four Districts

Fit Index	Salémata	Saraya	Kédougou	Dianké Makha	Threshold
TLI	0.94	0.92	0.94	0.97	> 0.90
RMSEA	0.06 [0.04;0.07]	0.06 [0.04;0.08]	0.06 [0.05;0.08]	0.054 [0.01;0.08]	< 0.08
RMSR	0.03	0.02	0.003	0.02	< 0.05
χ^2/df	2.00	2.05	2.00	1.76	< 3.00

Note: TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation [90% CI]; RMSR = Root Mean Square of Residuals; χ^2/df = Chi-square divided by degrees of freedom. Criteria: TLI > 0.90 (acceptable), > 0.95 (good); RMSEA < 0.08 (acceptable), < 0.05 (good); χ^2/df < 3 (acceptable).

Table 4: Main Factor Loading by District

Only main factor loadings (≥ |0.30|) are displayed per district, grouped by variable type.

Variables	Salémata	Saraya	Kédougou	Dianké Makha
Wind factor				
Wind_gust_max	F1 (0.98)	F1 (0.99)	F5 (0.78)	—
Wind_speed_max	—	—	F5 (0.73)	—
Wind_gust_min	F5 (0.77)	F4 (0.91)	F1 (0.97)	F3 (0.92)
Wind_speed_min	F5 (0.85)	F4 (0.68)	F1 (0.65)	F3 (0.93)
Lunar phase				
Moon_phase_max	F2 (0.99)	F2 (0.99)	F2 (0.99)	—
Moon_phase_mean	F2 (0.73)	—	F2 (0.76)	F1 (0.99)
Moon_phase_min	—	—	—	F1 (0.79)
Precipitation & Cloud cover				
Precip_max	F3 (0.71)	F3 (0.71)	F4 (0.68)	F4 (0.69)
Precip_cover_max	F3 (0.78)	F3 (0.78)	F4 (0.80)	—
Cloud_cover_min	F3 (0.86)	F3 (0.86)	_ F4 (0.57)	F4 (0.82)
Cloud_cover_max	F3 (0.54)	—	—	—
Sea level pressure				
Sea_level_pressure_max	F4 (0.82)	F5(0.86)	F3 (0.85)	F2 (0.83)
Sea_level_pressure_min	F4 (0.90)	F5 (0.84)	F3 (0.87)	F2 (0.99)

Note: Values indicate factor number and standardized loading (e.g., F1 (0.98) = Factor 1, loading = 0.98). — = no notable loading.

Structural equations per district:

$$Kédougou: \begin{cases} F4^{(1)} = \lambda_{4,1}^{(1)}\text{precip\_max} + \lambda_{4,2}^{(1)}\text{precipcover\_max} + \\ \quad \lambda_{4,3}^{(1)}\text{cloudcover\_min} + \delta_4^{(1)} \\ F3^{(1)} = \lambda_{3,1}^{(1)}\text{sealevelpressure\_min} + \lambda_{3,2}^{(1)}\text{sealevelpressure\_max} + \delta_3^{(1)} \\ F1^{(1)} = \lambda_{1,1}^{(1)}\text{windgust\_min} + \lambda_{1,2}^{(1)}\text{windspeed\_min} + \delta_1^{(1)} \\ F5^{(1)} = \lambda_{5,1}^{(1)}\text{windgust\_max} + \lambda_{5,2}^{(1)}\text{windspeed\_max} + \delta_5^{(1)} \\ F2^{(1)} = \lambda_{2,1}^{(1)}\text{moonphase\_max} + \lambda_{2,2}^{(1)}\text{moonphase\_mean} + \delta_2^{(1)} \end{cases}$$

$$Dianké \begin{cases} F4^{(2)} = \lambda_{4,1}^{(2)}\text{precip\_max} + \lambda_{4,2}^{(2)}\text{cloudcover\_min} + \delta_4^{(2)} \\ F2^{(2)} = \lambda_{2,1}^{(2)}\text{sealevelpressure\_min} + \lambda_{2,2}^{(2)}\text{sealevelpressure\_max} + \delta_2^{(2)} \\ F3^{(2)} = \lambda_{3,1}^{(2)}\text{windgust\_min} + \lambda_{3,2}^{(2)}\text{windspeed\_min} + \delta_3^{(2)} \\ F1^{(2)} = \lambda_{1,1}^{(2)}\text{moonphase\_mean} + \lambda_{1,2}^{(2)}\text{moonphase\_min} + \delta_1^{(2)} \end{cases}$$

$$Salémata: \begin{cases} F5^{(3)} = \lambda_{5,1}^{(3)}\text{windgust\_min} + \lambda_{5,2}^{(3)}\text{windspeed\_min} + \delta_5^{(3)} \\ F4^{(3)} = \lambda_{4,1}^{(3)}\text{sealevelpressure\_max} + \lambda_{4,2}^{(3)}\text{sealevelpressure\_min} + \delta_4^{(3)} \\ F1^{(3)} = \lambda_{1,1}^{(3)}\text{windgust\_max} + \lambda_{1,2}^{(3)}\text{windspeed\_max} + \delta_1^{(3)} \\ F3^{(3)} = \lambda_{3,1}^{(3)}\text{precip\_max} + \lambda_{3,2}^{(3)}\text{precipcover\_max} + \\ \quad \lambda_{3,3}^{(3)}\text{cloudcover\_min} + \lambda_{3,4}^{(3)}\text{cloudcover\_max} + \delta_3^{(3)} \\ F2^{(3)} = 0.99 \cdot \text{moonphase\_max} + \lambda_{2,1}^{(3)}\text{moonphase\_mean} + \delta_2^{(3)} \end{cases}$$

$$Saraya: \begin{cases} F4^{(4)} &= \lambda_{4,1}^{(4)}\text{windgust\_min} + \lambda_{4,2}^{(4)}\text{windspeed\_min} + \delta_4^{(4)} \\ F5^{(4)} &= \lambda_{5,1}^{(4)}\text{sealevelpressure\_max} + \lambda_{5,2}^{(4)}\text{sealevelpressure\_min} + \delta_5^{(4)} \\ F1^{(4)} &= \lambda_{1,1}^{(4)}\text{windgust\_max} + \lambda_{1,2}^{(4)}\text{windspeed\_max} + \delta_1^{(4)} \\ F3^{(4)} &= \lambda_{3,1}^{(4)}\text{precip\_max} + \lambda_{3,1}^{(4)}\text{precipcover\_max} + \\ & \quad \lambda_{3,2}^{(4)}\text{cloudcover\_min} + \delta_3^{(4)} \\ F2^{(4)} &= \lambda_{2,1}^{(4)}\text{moonphase\_max} + \delta_2^{(4)} \end{cases}$$

Table 5: Fit Measures for Confirmatory Factor Analysis

Index	Salémata	Kédougou	Saraya	Dianké Makha	Accepted level
P-value (Chi-square)	0.046	0.348	0.151	0.185	—
Chi-square (χ^2) /df	1.386	1.074	1.899	1.293	< 3
CFI	0.990	0.975	0.983	0.980	≥ 0.90
TCI	0.985	0.963	0.965	0.967	≥ 0.90
SRMR	0.033	0.041	0.030	0.034	< 0.08
RMSEA	0.036 [0.006, 0.059]	0.064 [0.000, 0.093]	0.058 [0.000,0.122]	0.067 [0.000, 0.139]	< 0.05

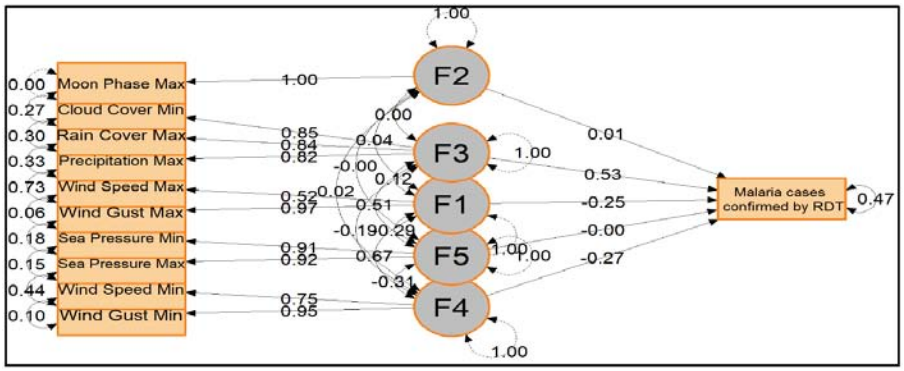
90% confidence interval for RMSEA in brackets.

3.3. SEM Model

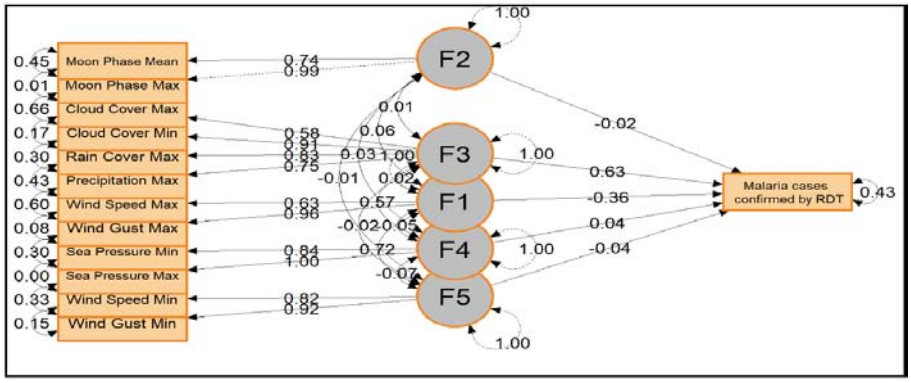
To determine the different effects of each ecological factor, SEM was applied [10]. Figure **3a-3d** presents the path diagrams visualize the architecture of causal relationships identified in each district. Structural equation models demonstrated satisfactory to excellent fit across all districts (CFI: 0.977–0.988, TLI: 0.965–0.981, RMSEA: 0.041–0.064, SRMR: 0.025–0.045), confirming the robustness of the adopted methodological approach. Results revealed significant spatial heterogeneity in climate-malaria relationships across districts (Table **6**). Precipitation and cloud cover emerged as the primary drivers of malaria transmission in three districts, with the strongest total effects observed in Saraya (0.66), Salémata (0.63), and Kédougou (0.51). In Dianké Makha, cloud cover max showed the strongest direct positive effect (0.54). Wind-related factors consistently provided protective effects across districts, though through different pathways. In Kédougou, high wind speed showed the strongest protective effect (−0.51), primarily through direct mechanisms, while in Saraya, minimum wind conditions provided the most pronounced protection (−0.64) through both direct and indirect pathways. Atmospheric pressure showed consistent indirect positive effects across districts (0.18–0.42), suggesting that pressure variations indirectly favor malaria transmission. Lunar phase showed no significant effect in any district. Meteorological factors explained more than half the variance in malaria cases in all districts (Kédougou: 56.4%, Salémata: 56.4%, Saraya: 53.6%, Dianké Makha: 50.0%), with unmeasured factors playing a relatively larger role in Dianké Makha.

DISCUSSION

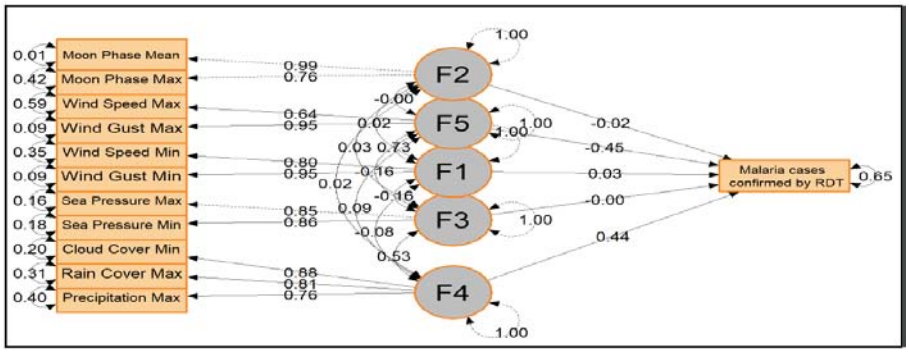
This study developed structural equation models to identify causal relationships between climatic and environmental variables and confirmed malaria cases in four districts of southeastern Senegal. The satisfactory to excellent fit indices (RMSEA: 0.041–0.064, CFI: 0.977–0.988) confirm the robustness of the adopted methodological approach, with climatic factors explaining more than 53% of the variance in malaria cases across all districts. From a statistical methodology perspective, this study makes several contributions to medical research. First, the decomposition of total effects into direct and indirect components revealed that variables such as atmospheric pressure which would have appeared non-significant in standard regression models exerted significant indirect effects on malaria incidence (0.18–0.42). Second, the integration of latent variables through EFA and CFA allowed the identification of unobserved climatic constructs that directly or indirectly influence malaria transmission, overcoming the limitations of univariate approaches. Third, the use of weekly rather than monthly data improved the capture of critical temporal variations in climate-malaria relationships. Together, these contributions demonstrate the applicability of a sequential



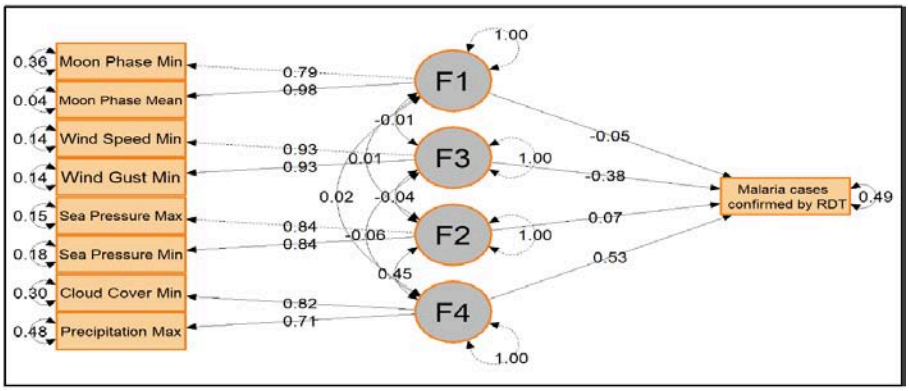
a



b



c



d

Figure 3: a: Causal relationships between latent variables with standardized path coefficient in Saraya.

b: Causal relationships between latent variables with standardized path coefficient in Salémata.

c: Causal relationships between latent variables with standardized path coefficient in Kédougou.

d: Causal relationships between latent variables with standardized path coefficient in Dianké Makha.

Table 6: Descriptions of Factors Items and Effects of Latent Climatic Factors on Malaria Cases by District

Districts	Factor	Description of items	Dir	Indir	Tot
Salémata	F1	Wind gust max and wind speed max	−0.36	—	−0.36
	F2	Moon phase means and moon phase max	—	—	—
	F3	Max precipitation, max precipitation coverage, min cloud coverage, max cloud coverage	0.63	—	0.63
	F4	Sea level pressure max and Sea level pressure min	—	0.36	0.36
	F5	Wind gust min and wind speed min	—	−0.26	−0.26
Saraya	F1	Wind gust max and wind speed max	−0.25	−0.28	−0.53
	F2	Moon phase max	—	—	—
	F3	Max precipitation, max precipitation coverage, min cloud coverage	0.53	0.13	0.66
	F4	Wind gust min and wind speed min	−0.27	−0.37	−0.64
	F5	Sea level pressure max and Sea level pressure min	—	0.42	0.42
Kédougou	F1	Wind gust min and wind speed min	—	−0.086	−0.086
	F2	Moon phase mean and moon phase max	—	—	—
	F3	Sea level pressure max and Sea level pressure min	—	0.184	0.184
	F4	Max precipitation, max precipitation coverage, min cloud coverage	0.44	0.07	0.51
	F5	Wind gust max and wind speed max	−0.45	−0.06	−0.51
Dianké Makha	F1	Moon phase mean and moon phase min	—	—	—
	F2	Sea level pressure max and Sea level pressure min	—	0.24	0.24
	F3	Wind gust min and wind speed min	−0.38	—	−0.38
	F4	Maximum precipitation and minimum cloud coverage	0.54	—	0.54

Dir. = Direct Effects; Indir. = Indirect Effects; Tot. = Total Effect; — = No-significant effect.

EFA-CFA-SEM framework as a replicable analytical pipeline for causal inference in environmental epidemiology, with broader implications for other vector-borne diseases such as dengue and leishmaniasis. The results revealed complex and geographically differentiated causal mechanisms. Precipitation and cloud cover were the primary drivers of transmission across three districts (coefficients: 0.51–0.66), while wind speed emerged as the dominant protective factor, particularly in Saraya (−0.64) and Kédougou (−0.51). The absence of significant lunar phase effects was consistent across all districts. These distinct spatial patterns underscore the importance of adapting surveillance strategies to local climatic specificity. Most existing studies have focused on temperature, rainfall, and humidity, leaving limited knowledge on wind, atmospheric pressure, cloud cover, and lunar phase. Our study helps fill this gap by examining their role in the causal chain of malaria transmission. The model provides a basis for developing early warning systems linked to weather forecasts, enabling the national malaria control program to adapt preventive interventions to district-specific climatic contexts.

LIMITATIONS

Several limitations must be considered when interpreting these results. First, malaria case data from the routine surveillance system (DHIS2) may underestimate actual transmission intensity due to detection bias related to healthcare accessibility and variable diagnostic practices, particularly in rural areas. Second, the analysis does not account for malaria control interventions such as bed net distribution and indoor residual spraying implemented differentially across districts and periods, which may independently influence observed patterns. Third, the absence of entomological data limits our ability to directly link observed temporal patterns to underlying vector dynamics. From a methodological standpoint, the

reliance on Visual Crossing for meteorological data introduces uncertainties related to spatial interpolation and local-scale accuracy. Additionally, EFA and SEM rely on assumptions of linearity and parameter invariance over time that may be partially violated in climate data, where extreme events constitute informative phenomena rather than aberrations [25], precipitation distributions exhibit strong positive skewness [26], and nonlinear threshold effects are common [27]. Furthermore, the temporal non-stationarity of climate processes, particularly in the context of climate change, may challenge the assumption of parameter invariance in SEM models. Future research should integrate entomological data, intervention coverage indicators, and dynamic Bayesian networks to improve causal inference and temporal modeling.

CONCLUSION

This study demonstrates that structural equation modeling provides a rigorous and innovative framework for identifying the complex causal pathways through which climatic and environmental factors influence malaria transmission in southeastern Senegal. By integrating latent variable modeling with a sequential EFA-CFA-SEM approach, the analysis revealed direct and indirect effects that would have remained undetected by classical statistical methods. Key findings show significant spatial heterogeneity across the four districts: precipitation and cloud cover emerged as the primary drivers of malaria transmission, while wind dynamics consistently provided protective effects. These district-specific climatic profiles support the development of localized surveillance strategies and early warning systems linked to weather forecasts. Despite some limitations, the SEM framework developed here offers a replicable methodological contribution for causal inference in vector-borne disease research and provides an operational basis for adaptive malaria control strategies in high-burden districts.

DECLARATION

This study uses exclusively aggregated epidemiological surveillance data without identifiable individual information, in accordance with national and international ethical guidelines.

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.

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.

AUTHORSHIP CONTRIBUTION

Leontine Ndogou Bakhoum: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing original draft, Review & editing, Validation.

Mor Absa Loum: Conceptualization, Methodology, Review & editing, Validation.

Mouhamad Mounirou Allaya: Conceptualization, Review & editing, Validation.

Gouvidé Jean Gbaguidi: Conceptualization, Formal analysis, Methodology, Review & editing, Validation.

Khady Ndiaye: Conceptualization, Writing– review and editing, Validation.

Almamy Youssouf Ly: Conceptualization, Writing– review and editing, Validation.

Mamadou Bousso: Conceptualization, Review & editing, Validation.

Jean Louis Abdourahim Ndiaye: Conceptualization, Funding acquisition, Supervision, Validation.

APPENDIX

Appendix A: Mahalanobis Distance

Let X be a random variable taking values in $\mathbb{R}^d$ with non-singular covariance matrix Σ. The Mahalanobis distance is a natural generalization of the Euclidean distance that accounts for the covariance structure of multivariate data [28]. In some cases, the usual Euclidean distance is not a suitable choice because it treats all dimensions equally, ignoring:

– Differences in variance between components

– Correlations between variables

The Mahalanobis distance is defined as the square root of the quadratic form:

$$D_M(x,\mu) = d(x,\mu) = \sqrt{(x-\mu)T^T\Sigma^{-1}(x-\mu)}$$

Σ^{-1} is the inverse of the covariance matrix Σ.

μ is the mean of X variable

Appendix B: Variance Inflation Factor (VIF)

To ensure the robustness of the multiple linear regression model and prevent estimator instability issues, a systematic evaluation of multicollinearity was conducted using the Variance Inflation Factor (VIF). The multiple linear regression model is expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$$

where Y is the dependent variable, X_1, X_2, …, X_k are the independent variables, β_0, β_1, …, β_k are the regression coefficients, and ε is the error term. The VIF, defined by the formula

$$VIF_j = \frac{1}{1-R_j^2}$$

quantifies the increase in the variance of the regression coefficient for variable X_j due to linear correlations with other explanatory variables in the model. In this formulation, R^2_j represents the coefficient of determination obtained by regressing X_j on all other independent variables through the auxiliary regression:

$$X_j = \alpha_0 + \alpha_1 X_1 + \ldots + \alpha_{j-1} X_{j-1} + \alpha_{j+1} X_{j+1} + \ldots + \alpha_k X_k + \nu$$

The R^2_j is calculated as:

$$R_j^2 = 1 - \frac{SCR_j}{SCT_j} = 1 - \frac{\sum_{i=1}^{n}\left(X_{ji} - \hat{X}_{ji}\right)^2}{\sum_{i=1}^{n}\left(X_{ji} - \bar{X}_j\right)^2}$$

where X_{ij} is the observed value of X_j for observation i, $\hat{X}_{ji}$ is the predicted value from the auxiliary regression, $\bar{X}_j$ is the mean of variable X_j, SCR_j is the sum of squared residuals, and SCT_j is the total sum of squares for variable X_j. [29], variables exhibiting VIF values greater than 10 were systematically excluded from the final model, as this threshold represents the conventional upper bound beyond which multicollinearity seriously affects the reliability of regression estimates, thereby ensuring the stability of structural equation model parameters while retaining a sufficient number of climatic predictors.

Appendix C: KMO, Bartlett's Test and Multivariate Normality

The assumption of univariate and multivariate normality was evaluated using the MVN package in R, which implements multiple statistical tests including univariate (Anderson-Darling) and multivariate (Henze-Zirkler) normality tests [30,31,32]. A significance level of α = 0.05 was used to determine whether the null hypothesis of multivariate normality should be rejected.

Table 7: Assumptions of Factorability; Results of KMO, Bartlett's Test and Multivariate Normality

	Kédougou	Salémata	Saraya	Dianké Makha
KMO measure	0.74	0.75	0.8	0.69
Bartlett's test : chi-square	1090.21	966.98	748.25	569.27
Df	153	153	120	55
p. Value	3.95e−141	2.70e−118	1.8e−91	1.39e−86

Henze-Zirkler test: Statistic	1.068	1.102	1.076	1.452
P-value	< 0.001	< 0.001	< 0.001	< 0.001

KMO: KMO < 0.50 (unacceptable), 0.60–0.69 (mediocre), 0.70–0.79 (good), 0.80–0.89 (very good), ≥ 0.90 (excellent).
Henze-Zirkler: p-value > 0.05 (acceptable normality), ≤ 0.05 (violation).

Appendix D

Table 8: Descriptive Statistics of Factors

Zone	Facteur	SS loadings	Proportion Var (%)	Cumulative Var (%)	Proportion Explained (%)
Salémata	F1	1.46	9.1	9.1	14.1
	F2	1.56	9.7	18.8	14.9
	F3	3.26	20.4	39.2	31.3
	F4	2.27	14.2	53.4	21.8
	F5	1.87	11.7	65.1	17.9
Saraya	F1	1.63	10.0	10.0	16.0
	F2	1.12	7.0	17.0	11.1
	F3	3.18	20.0	37.0	31.3
	F4	1.90	12.0	49.0	18.8
	F5	2.29	14.0	63.0	22.6
Kédougou	F1	1.82	11.0	11.0	17.6
	F2	1.59	9.0	20.0	15.4
	F3	2.21	13.8	33.8	21.5
	F4	2.95	18.4	52.2	28.6
	F5	1.72	10.8	64.5	16.7
Dianké Makha	F1	1.60	14.7	14.7	20.9
	F2	2.17	19.7	34.4	28.1
	F3	2.28	20.7	55.1	29.5
	F4	1.65	15.1	70.3	21.4

REFERENCE

[1] World Health Organization. World malaria report 2023. Geneva: World Health Organization; 2023.

[2] Tairou F, *et al*. Malaria prevention knowledge, attitudes, and practices (KAP) among adolescents in Senegal. PLoS One 2022; 17(12): e0274656. https://doi.org/10.1371/journal.pone.0274656

[3] National Malaria Control Program (NMCP). Epidemiological bulletin of malaria in Senegal year 2022. Dakar: NMCP; 2023.

[4] Segun OE, *et al*. Statistical modelling of the effects of weather factors on malaria occurrence in Abuja, Nigeria. Int J Environ Res Public Health 2020; 17(10): 3474. https://doi.org/10.3390/ijerph17103474

[5] Liu Q, *et al*. Association of temperature and precipitation with malaria incidence in 57 countries. J Glob Health 2024; 14: 04021. https://doi.org/10.7189/jogh.14.04021

[6] Kurup R, Deonarine G, Ansari AA. Malaria trend and effect of rainfall and temperature within regions 7 and 8, Guyana. Int J Mosq Res 2017; 4(6): 48-55.

[7] Assefa GM, Muluneh MD, Alemu ZA. The relationship of climate change and malaria incidence in the Gambella region, Ethiopia. Climate 2025; 13(5): 104. https://doi.org/10.3390/cli13050104

[8] Mayilsamy M, *et al*. Impact of climatic factors on the occurrence of malaria in India from 1995 to 2023. Malar J 2025; 24(1): 113. https://doi.org/10.1186/s12936-025-05326-5

[9] Rubuga FK, *et al*. Potential impact of climatic factors on malaria in Rwanda between 2012 and 2021. Malar J 2024; 23(1): 274. https://doi.org/10.1186/s12936-024-05097-5

[10] Beloconi A, *et al*. Malaria, climate variability, and interventions: modelling transmission dynamics. Sci Rep 2023; 13(1): 7367. https://doi.org/10.1038/s41598-023-33868-8

[11] Murdock CC, Sternberg ED, Thomas MB. Malaria transmission potential could be reduced with current and future climate change. Sci Rep 2016; 6(1): 27771. https://doi.org/10.1038/srep27771

[12] National Malaria Control Program (NMCP). National strategic plan for malaria control in Senegal, 2016-2020. Dakar: Ministry of Health and Social Action, Republic of Senegal; 2016.

[13] Ministry of Health and Social Action, Senegal. District health information system 2 (DHIS2) - Senegal. Dakar: Ministry of Health; 2024.

[14] Lin D, Du S, Zhao Z, Zhang T, Wang L, Zhang Q, *et al*. Climate warming fuels the global antibiotic resistome by altering soil bacterial traits. Nat Ecol Evol 2025; 9(8): 1512-1526. https://doi.org/10.1038/s41559-025-02740-5

[15] National Oceanic and Atmospheric Administration. NOAA. 2024. Available from: https://www.noaa.gov

[16] Bartlett MS. The effect of standardization on a χ^2 approximation in factor analysis. Biometrika 1951; 38(3/4): 337-344. https://doi.org/10.1093/biomet/38.3-4.337

[17] Kaiser HF, Rice J. Little jiffy, mark IV. Educ Psychol Meas 1974; 34(1): 111-117. https://doi.org/10.1177/001316447403400115

[18] Revelle W. psych: Procedures for psychological, psychometric, and personality research. R package version 2.3.9; 2023. Available from: https://CRAN.R-project.org/package=psych

[19] Brown TA. Confirmatory factor analysis for applied research. 2nd ed. New York: Guilford Publications; 2015.

[20] Rosseel Y. lavaan: An R package for structural equation modeling. J Stat Softw 2012; 48: 1-36. https://doi.org/10.18637/jss.v048.i02

[21] Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis. Struct Equ Modeling 1999; 6(1): 1-55. https://doi.org/10.1080/10705519909540118

[22] Bullerjahn M. The influence of childhood familial dysfunction on emotional well-being and substance use behaviors in young adults. Can Stud Popul 2018; 45(3-4): 188-195. https://doi.org/10.25336/csp29418

[23] Kline RB. Principles and practice of structural equation modeling. 4th ed. New York: Guilford Publications; 2015.

[24] R Core Team. R: A language and environment for statistical computing. Version 4.5.1. Vienna: R Foundation for Statistical Computing; 2025.

[25] Tsonis AA, Roebber PJ. The architecture of the climate network. Physica A 2004; 333: 497-504. https://doi.org/10.1016/j.physa.2003.10.045

[26] Queiroga F. Manuels des analyses psychométriques - analyse factorielle exploratoire. LAPCOS; 2024. Available from: https://hal.science/hal-04729096v1

[27] Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. 7th ed. Upper Saddle River: Pearson Prentice Hall; 2010.

[28] R Core Team. R stats package documentation: Mahalanobis distance function [Internet]. Vienna: R Foundation for Statistical Computing; 2024. Available from: https://www.r-project.org

[29] Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. 5th ed. Hoboken: Wiley; 2012.

[30] Gbaguidi GJ, *et al*. Towards an intelligent malaria outbreak warning model in the northern part of Benin, west Africa. BMC Public Health 2024; 24(1): 450. https://doi.org/10.1186/s12889-024-17847-w

[31] Korkmaz S, Goksuluk D, Zararsiz G. MVN: An R package for assessing multivariate normality. R J 2014; 6(2): 151-162. https://doi.org/10.32614/RJ-2014-031

[32] Anderson TW, Darling DA. A test of goodness of fit. J Am Stat Assoc 1954; 49(268): 765-769. https://doi.org/10.1080/01621459.1954.10501232

Received on 16-03-2026 Accepted on 14-04-2026 Published on 07-05-2026

https://doi.org/10.6000/1929-6029.2026.15.17