Proceedings of ICBioMed 2026 Symposium: Biomedical Engineering and Public Health: Real-Time Wearables for Disease Detection
DOI: 10.54254/2753-8818/2026.ES38197

# *Study on the Driving Mechanism of Meteorological Factors on the Density of Anopheles sinensis—A Comparative Analysis Based on GWRF-SHAP and the Geodetector*

**Hanxi Wang**

*College of Atmospheric Sciences, Lanzhou University, Lanzhou, China*
*3612384643@qq.com*

***Abstract.*** *Anopheles sinensis* is an important malaria vector in China. Clarifying the driving mechanism of meteorological factors on its population density has practical significance for consolidating the achievements of malaria elimination. Objective: to compare the differences between two models, the geographically weighted random forest (GWRF) combined with SHAP interpretability analysis and the geodetector, in identifying the driving factors of vector mosquitoes, and to analyze the effects of multiple meteorological factors, time-lag effects and their interactions on the density of *Anopheles sinensis*. Methods: based on public datasets of mosquito density and meteorology, 39 meteorological variables covering the current month, 1-month lag and 2-month lag were included, and GWRF-SHAP and the geodetector were used respectively to carry out single-factor contribution and factor interaction effect analysis, and the outputs of the two models were compared. Results: relative humidity, precipitation, wind speed and temperature at 1-month lag are the key meteorological factors affecting the density of *Anopheles sinensis*, and the meteorological effects have a time lag of 1-2 months. The primary driving factor identified by GWRF is relative humidity at 1-month lag, and the strong interaction effects are concentrated in the temperature-temperature and temperature-pressure variable combinations at 2-month lag; the geodetector takes precipitation days at 1-month lag as the primary factor, and the strong interactions are mostly short-term variable combinations of humidity with temperature, pressure and wind speed. The difference between the results of the two models comes from the different algorithm principles: GWRF focuses on the joint contribution of variables at the prediction level, and the geodetector focuses on the explanatory power of factors for spatial stratified differentiation. Conclusion: a single model will bring biased results, and a method of combining multiple models will be considered later, so as to provide methodological reference for zonal early warning and refined control of malaria vectors.

***Keywords:*** GWRF-SHAP, geodetector, *Anopheles sinensis*

## 1. Introduction

*Anopheles sinensis* is one of the important vectors of malaria transmission and has long been the core concern of China's vector surveillance and control system. Although China obtained the

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.ES38197

malaria-free certification from the World Health Organization in 2021, the potential threat of this mosquito species to public health security still cannot be ignored [1]. On the one hand, *Anopheles sinensis* is one of the most widely distributed and most abundant dominant mosquito species in China [2]; on the other hand, persistent imported malaria cases, the emergence of new mosquito-borne viruses, and the development of insecticide resistance may all lead to the resurgence of epidemics [3]. Therefore, long-term research on the population dynamics of *Anopheles sinensis* and the patterns of its spatial distribution is the key scientific basis for maintaining the achievements of malaria elimination. Existing studies have shown that meteorological factors such as temperature, precipitation and humidity have significant effects on the population density and distribution of this mosquito species. However, under the new situation of frequent compound extreme meteorological events, the independent effect of a single meteorological factor can hardly fully characterize the action mechanism of climate change on vector populations, and it is urgent to deepen the relevant theoretical understanding from the paths of multi-factor interaction and nonlinear response. Djègbè et al. [4] found through seasonal observations under laboratory and semi-field conditions that seasonal fluctuations of temperature significantly affect the development rate of anophelines from egg to adult and the longevity of adults, while relative humidity and precipitation indirectly regulate the seasonal fluctuation of anopheline populations by affecting the energy consumption of adults and the availability of breeding water. This finding indicates that combinations of meteorological factors jointly shape the spatiotemporal distribution pattern of vectors by regulating anopheline life-history traits and habitat suitability, and provide key biological parameter support for constructing vector population prediction models based on multiple meteorological elements.

Field monitoring of mosquitoes is limited by manpower and sampling sites, and it is difficult to achieve large-scale, full-coverage sampling. Species distribution models (SDMs) use known distribution points of species and environmental variables to build models, and can estimate the potential distribution of species in areas without field surveys; this method has been widely applied, and the related algorithms are also continuously iterated and updated. At present, the algorithm mechanisms and complexity of SDMs differ significantly, resulting in obvious differences in model accuracy and prediction results. In studies on the distribution of *Anopheles sinensis* and its environmental impact factors, scholars have used generalized linear models (GLM), random forest (RF), artificial neural networks (ANN), maximum entropy models (MaxEnt), the geodetector (GD) and other tools to analyze the response of the distribution of this vector species to climate factors and its relationship with climate change [5,6]. But except for G, these models mostly imply the assumption of 'global homogeneity' and ignore the nonstationarity of environmental effects varying with space. The geographically weighted series of models captures spatial heterogeneity by relying on the local regression idea and has obvious advantages in such research scenarios; the geodetector can quantitatively calculate factor explanatory power at the global and stratified statistical levels and identify the interaction relations between factors. The two types of models each have advantages, and it is necessary to carry out systematic comparative evaluation.

This study takes the geographically weighted random forest (GWRF) model as an example to model the distribution pattern of *Anopheles sinensis* and its climate driving factors, and compares it with the results of the geodetector (Geodetecter, GD) model [6]. The aim is to analyze the similarities and differences between single-factor contribution and multi-factor interaction effects, and to provide a methodological basis for model selection. In the end, this study not only provides a more refined spatial understanding of the response mechanism of the distribution of *Anopheles sinensis* to climate change, but also provides scientific support for the dynamic early warning of malaria vectors and the formulation of regional adaptive control strategies in the future.

DOI: 10.54254/2753-8818/2026.ES38197

## 2. Data and methods

### 2.1. Data sources and processing

The density and variable data of *Anopheles sinensis* used in this study are all from Jia et al. [6]. There are 14 variables such as temperature, humidity, precipitation and wind speed (Table 1). Also considering the potential lag between mosquito density and meteorology, the meteorological scenarios were set as the current month (L0), the previous month (L1) and two months prior (L2), which means the density of *Anopheles sinensis* corresponds to 39 meteorological variables. Longitude and latitude were included in the model as covariates

Table 1. Main variables

| Primary variable | Secondary variable | Abbreviation | Unit |
|---|---|---|---|
| Temperature | Mean air temperature | X1 | ℃ |
| | Mean maximum air temperature | X2 | ℃ |
| | Mean minimum air temperature | X3 | ℃ |
| | Extreme maximum air temperature | X4 | ℃ |
| | Extreme minimum air temperature | X5 | ℃ |
| Humidity | Relative humidity | X6 | % |
| Precipitation | Accumulated precipitation | X7 | mm |
| | Maximum precipitation | X8 | mm |
| | Precipitation days | X9 | d |
| Air pressure | Sea level pressure | X10 | hPa |
| Wind speed | Mean wind speed | X11 | m/s |
| | Mean maximum sustained wind speed | X12 | m/s |
| | Daily maximum mean wind speed | X13 | m/s |
| Geographic coordinates | Longitude | X14 | ° |
| | Latitude | X15 | ° |

### 2.2. Geographically weighted random forest

#### 2.2.1. Model introduction

Geographically weighted random forest combines the nonlinear fitting ability of random forest (RF) with the spatial local modeling advantage of geographically weighted regression (GWR), and is a machine learning algorithm adapted to spatially heterogeneous data. The model introduces a spatial distance weight matrix in the decision tree modeling process of random forest, assigns differentiated weights to samples at different geographic locations, and realizes the local adaptive updating of model parameters, breaking through the limitation of constant parameters of traditional global models [7].

DOI: 10.54254/2753-8818/2026.ES38197

#### 2.2.2. Model evaluation

To objectively and accurately evaluate the prediction accuracy and fitting stability of the GWRF model, and avoid the limitation of single-index evaluation, this study selected three classical regression model evaluation indexes, root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R2), for quantitative evaluation. Among them, RMSE and MAE are used to quantify the absolute error between the predicted values and the measured values of the model, and directly reflect the prediction bias level of the model. R2 is used to characterize the degree of explanation of the variation of the dependent variable by the independent variables, and measure the overall fitting quality of the model. RMSE=1.591, MAE=0.484, $R^2$=0.8711, indicating that the accuracy and fitting degree of the model are high.

### 2.3. SHAP interpretability analysis

To understand the internal decision-making mechanism of the GWRF model and quantify the actual effect of each meteorological factor, this study introduced the SHAP (Shapley Additive exPlanations) analysis method. The SHAP algorithm is based on cooperative game theory, can decompose the model prediction result of a single sample into the independent contribution values of each input feature, and accurately quantify the direction and magnitude of the influence of different meteorological factors on the density of *Anopheles sinensis*. Existing studies have shown that the SHAP framework can not only realize the main effect analysis of a single variable, but its dedicated interaction value index can also complete the effect decomposition of two and more variables, effectively revealing the synergistic promotion or antagonistic inhibition relations between variables, providing scientific support for analyzing the coupling mechanism of multiple factors [8].

Based on the output of the GWRF model, this study carried out SHAP interpretability analysis from two dimensions. One is single-factor main effect analysis, which screens the core dominant factors affecting the spatial distribution of *Anopheles sinensis* by statistically summarizing the SHAP values of each meteorological factor, and clarifies the action direction and influence degree of each factor; the other is multi-factor interaction effect analysis, which relies on the SHAP interaction value index to explore the nonlinear superposition law under the coupling of key meteorological factors.

## 3. Results and analysis

### 3.1. Analysis of the driving effects of single meteorological factors

#### 3.1.1. Single-factor contribution analysis based on the GWRF model

This study used the mean absolute SHAP value to measure the contribution level of each variable to the density prediction of *Anopheles sinensis*, sorted all variables in descending order according to this index, and the top 10 variables in contribution level are L1_X6, L0_X7, L0_X9, L0_X12, L0_X1, L2_X13, L1_X8, L0_X2, L0_X4 and L0_X3 (Table 1). Among them, the mean absolute SHAP value of L1_X6 reaches 7.042, which is the largest among all variables, significantly higher than all other variables, and it is the driving variable with the highest contribution in the GWRF model. The mean absolute SHAP values of the other 9 variables are distributed in the range of 0.318-2.450, and there are obvious gradient differences in the contribution degree among variables.

DOI: 10.54254/2753-8818/2026.ES38197

From the perspective of meteorological element categories, the top-ranked variables cover four categories: humidity, precipitation, wind speed and temperature. In the humidity dimension, only L1_X6 enters the front ranks, and the contribution of this variable is much higher than that of other elements, reflecting the dominant regulating effect of humidity on the population of *Anopheles sinensis*. The precipitation category includes L0_X7, L0_X9 and L1_X8; the contributions of the two precipitation variables of the current month are at a relatively high level, reflecting that precipitation has an immediate effect on the mosquito population, and at the same time the precipitation variable of 1-month lag also shows an influence that cannot be ignored, indicating that the precipitation effect has the characteristic of lagged accumulation. The wind speed category includes L0_X12 and L2_X13, corresponding to the current month and 2-month lag respectively, indicating that wind speed has both current effects and long-period lag effects on mosquito density. The selected temperature variables all belong to the current-month time scale, reflecting that current temperature conditions directly participate in the regulation of the population dynamics of *Anopheles sinensis*.

From the distribution of time scales, among the variables with top contributions there are 6 current-month variables, 3 variables of 1-month lag, and 1 variable of 2-month lag. Current-month variables dominate in number, indicating that the current effect of meteorological factors is very prominent, but the meteorological signals of 1-month lag and 2-month lag can also act on the mosquito population, and the density of *Anopheles sinensis* is the result of the joint action of meteorological conditions in different time windows.

Table 2. Single-factor variables with the top ten largest SHAP values based on the GWRF model

| Variable | Mean SHAP | Mean absolute SHAP | SHAP standard deviation |
|---|---|---|---|
| L1_X6 | 7.032229 | 7.041792 | 15.73791 |
| L0_X7 | -2.43893 | 2.450385 | 5.433855 |
| L0_X9 | -1.5323 | 1.545818 | 3.442311 |
| L0_X12 | -0.99206 | 0.992061 | 2.193307 |
| L0_X1 | -0.4099 | 0.409901 | 0.860224 |
| L2_X13 | -0.35336 | 0.353363 | 0.567615 |
| L1_X8 | 0.334372 | 0.345419 | 0.753332 |
| L0_X2 | -0.33009 | 0.330093 | 0.693177 |
| L0_X4 | -0.3123 | 0.318075 | 0.692611 |
| L0_X3 | -0.31768 | 0.317682 | 0.539852 |

### 3.1.2. Single-factor contribution analysis based on the GD model

The GD model uses the q statistic to characterize the explanatory power of a factor for the spatial differentiation of the density of *Anopheles sinensis*; the larger the q statistic, the stronger the spatial explanatory power of the factor. Sorting all variables by q value in descending order, the top 10 variables in explanatory power are L1_X9, L2_X13, L0_X3, L0_X5, L1_X6, L0_X1, L1_X11, L2_X12, L1_X12 and L1_X7 [6] (Table 2). Among all variables, the q value of L1_X9 is 0.230, the largest of all variables; the q values of the other 9 variables are distributed between 0.149-0.196.

DOI: 10.54254/2753-8818/2026.ES38197

The statistical test results show that, except for L1_X7 whose p=0.123 fails the 0.05 significance test, the remaining 9 variables all satisfy p<0.01, reaching the extremely significant statistical level.

From the analysis of element categories, variables of precipitation, wind speed, temperature, humidity and pressure all enter the front ranks of explanatory power. Among precipitation elements, L1_X9 is the variable with the highest global q value, indicating that the precipitation condition of 1-month lag has the strongest explanatory power for the spatial differentiation of mosquito density. Wind speed related variables include L2_X13, L1_X11, L2_X12 and L1_X12, covering the three time windows of current month, 1-month lag and 2-month lag, proving that wind speed at different lag stages all participates in shaping the spatial differences of mosquito density. The selected temperature variables are mainly current-month indicators, reflecting the constraining effect of current temperature on the spatial pattern of mosquitoes. The humidity variable L1_X6 ranks 5th, further confirming that humidity is a core factor affecting *Anopheles sinensis*. The pressure variable L0_X10 has certain spatial explanatory power, indicating that sea level pressure also participates in the spatial differentiation process of mosquito density.

From the statistics of time scales, the variables with strong explanatory power cover multiple time scales including the current month, 1-month lag and 2-month lag. Although current-month variables account for a certain proportion, the maximum q value appears in a variable of 1-month lag, reflecting the shaping value of earlier meteorological conditions on the spatial pattern of mosquitoes.

Table 3. Single-factor variables with the top ten largest q values based on the GD model

| Variable | q value | P value |
|---|---|---|
| L1_X9 | 0.230052 | 0.000013 |
| L2_X13 | 0.195637 | 0.000347 |
| L0_X3 | 0.189568 | 0.000987 |
| L0_X5 | 0.187738 | 0.000106 |
| L1_X6 | 0.184854 | 0.00029 |
| L0_X1 | 0.174849 | 0.000229 |
| L1_X11 | 0.174538 | 0.000199 |
| L2_X12 | 0.170452 | 0.000708 |
| L1_X12 | 0.161075 | 0.000456 |
| L1_X7 | 0.149075 | 0.122793 |

### 3.1.3. Comparison of the single-factor analysis results of the two models

Combining the outputs of the two models, temperature, humidity, precipitation and wind speed are all the core meteorological elements driving the density change of *Anopheles sinensis*; pressure only shows a relatively obvious spatial stratified explanatory effect in the geodetector analysis, and its independent contribution is limited under the GWRF prediction framework. The influence of meteorological factors covers the current month, 1-month lag and 2-month lag at the same time, but the response intensity of different elements in different time windows is obviously differentiated.

Humidity shows high importance in both models. L1_X6 is the largest contribution value among all variables in GWRF, and also ranks among the front in the geodetector. From the ecological point

DOI: 10.54254/2753-8818/2026.ES38197

of view, relative humidity directly regulates the water evaporation rate of the body surface of adult mosquitoes; a low-humidity environment will cause dehydration stress of adults, shorten the survival time, and thereby reduce the population size. The prominent effect of humidity is concentrated at 1-month lag, which means that the influence of earlier humidity on adult survival takes a certain cycle to be reflected in the adult population density.

The influence of precipitation exists at both the current-month and 1-month lag scales. In the GWRF analysis the current-month precipitation variables have high contributions, reflecting that precipitation directly changes the breeding water habitats and produces immediate ecological effects; the maximum q value of the geodetector comes from the precipitation variable of 1-month lag, reflecting that the earlier accumulation of precipitation days has stronger explanatory power for the spatial differentiation of mosquito density. Moderate precipitation expands the breeding water environments and is conducive to larval development; heavy rainfall will destroy the existing breeding sites and suppress population growth. Current-month precipitation directly changes the existing habitat conditions, while the precipitation conditions of 1-month lag indirectly regulate the subsequent adult population size through the larval development cycle.

The wind speed effect runs through the three time windows of current month, 1-month lag and 2-month lag, and appears many times among the important variables of both models. Wind speed on the one hand interferes with the flight and blood-feeding activities of adults, and on the other hand accelerates the loss of body surface water and aggravates the dehydration stress of mosquitoes. Current-month wind speed brings immediate physiological pressure; the wind speed of 1-2 months lag acts more on the larval living environment, and then affects the adult population density after generational accumulation.

The significant single-factor effect of temperature is mainly concentrated at the current-month scale. Most of the important temperature variables screened by the two models are current-month variables, indicating that current temperature directly regulates the metabolism and development rate of *Anopheles sinensis* and has a direct effect on the population; lagged temperature variables did not enter the single-factor important sequence, indicating that the lag effect of the temperature single factor is weaker than that of humidity and precipitation.

Pressure only shows certain spatial explanatory power in the geodetector results; the current-month sea level pressure enters the front sequence of explanatory power, while its contribution in the GWRF single-factor analysis is low. This phenomenon shows that pressure has a certain effect on the spatial stratification pattern of mosquito density, but its independent predictive contribution to the population density value is limited.

## 3.2. Analysis of the interactive driving effects of meteorological factors

### 3.2.1. Factor interactions of the GWRF model

Based on the SHAP interaction analysis of the GWRF model, which characterizes the influence intensity of variable pairs acting together on the density prediction of *Anopheles sinensis*, and sorted in descending order by this index, the top 10 variable combinations in interaction intensity are L2_X1 with L2_X2, L2_X1 with L2_X3, L2_X2 with L2_X3, L2_X3 with L2_X5, L2_X1 with L2_X5, L2_X2 with L2_X5, L2_X3 with L2_X10, L2_X1 with L2_X10, L2_X2 with L2_X4, and L2_X2 with L2_X10 (Figure 1). The maximum SHAP value of all interaction combinations is 0.997, the minimum is 0.947, and the overall values are very close to the theoretical upper limit of 1.0, indicating that this group of variable pairs has an extremely strong synergistic prediction effect.

DOI: 10.54254/2753-8818/2026.ES38197

From the perspective of element combination types, the top interaction pairs are divided into two categories: temperature-temperature interactions and temperature-pressure interactions; the pairs involving humidity, precipitation and wind speed did not enter the front ranks of interaction intensity. In terms of time scale, all 10 interaction pairs are composed of variables of 2-month lag, and there is no combination of current-month or 1-month lag variables. This result shows that, under the GWRF prediction framework, the synergy between different temperature indicators and between temperature and sea level pressure needs 2 months of time accumulation before it can significantly affect the density of *Anopheles sinensis*. The interactions among factors within the current month and the 1-month lag, and the pairs involving humidity, precipitation and wind speed, have much lower synergistic contributions at the density prediction level than the temperature-related combinations of 2-month lag. Different temperature indicators constrain each other, and together with pressure they change the regional thermal environment conditions, and after a complete generational development cycle of mosquitoes they form a strong effect on the adult population density.

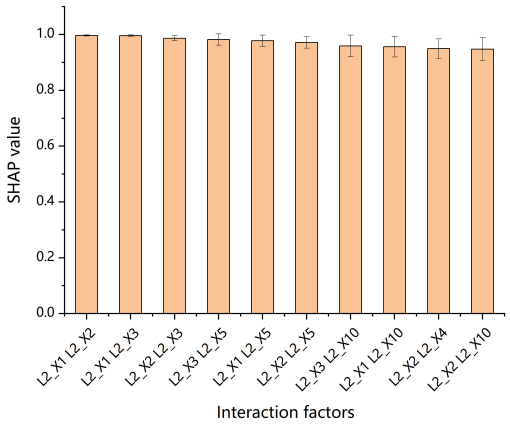



Figure 1. Interaction factor variables with the top ten largest SHAP values based on the GWRF model

### 3.2.2. Factor interactions of the GD model

The top 10 combinations in interactive explanatory power are L0_X6 with L1_X4, L0_X10 with L1_X6, L0_X6 with L0_X12, L0_X12 with L1_X6, L0_X5 with L1_X9, L0_X5 with L1_X12, L0_X6 with L2_X9, L1_X3 with L1_X11, L0_X3 with L1_X11, and L0_X4 with L1_X11 (Figure 2). The maximum interaction q value is 0.537 and the minimum is 0.469. The interaction effects of all pairs show nonlinear enhancement, and the interaction q value of each pair is higher than the q statistics of the two single factors themselves.

From the perspective of element combination types, most of the pairs with high interactive explanatory power contain L0_X6; humidity is the core element of the interactions, forming coupled pairs with temperature, pressure, wind speed and precipitation respectively; the temperature-temperature pairs appearing in large numbers in GWRF do not appear in the front sequence. In terms of time scale, the interaction pairs are mainly cross-time combinations of the current month and 1-month lag, only a few pairs involve variables of 2-month lag, and there is no interaction combination composed entirely of variables of 2-month lag. This result shows that, from the perspective of spatial differentiation explanation, the short-term coupling of humidity with other

DOI: 10.54254/2753-8818/2026.ES38197

meteorological elements is an important way of shaping the spatial pattern of the density of *Anopheles sinensis*; the current humidity superimposed with the earlier temperature, precipitation, wind speed and pressure conditions jointly cause the spatial differences of regional mosquito density.

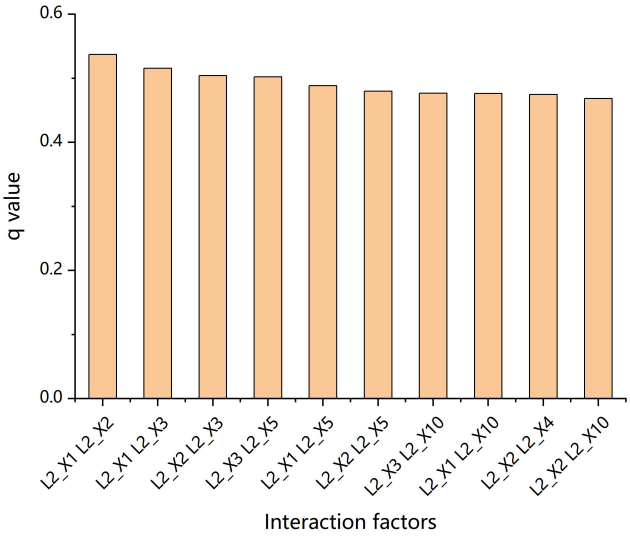



Figure 2. Interaction factor variables with the top ten largest q values based on the GD model

### 3.2.3. Comparative analysis of interaction factors

The strong interaction pairs identified by the two models differ obviously in dominant elements, combination types, time scales and numerical magnitude, and this difference comes from the different definitions of interaction effects by the two algorithms.

First, the dominant elements and combination types of interactions are different. The high-intensity interactions identified by GWRF are concentrated in the temperature-temperature and temperature-pressure pairs of 2-month lag, and the interaction intensity of the pairs involving humidity, precipitation and wind speed is low; the pairs with high interactive explanatory power of the geodetector all take humidity as the core, realizing the coupling of humidity with temperature, wind speed, pressure and precipitation respectively. The interaction effects output by GWRF represent the synergistic contribution of variables in the density prediction process, characterizing the joint influence of the joint change of multiple variables on the value of mosquito density; the interaction effects of the geodetector represent the superimposed explanatory power of variables in spatial stratification, measuring the increase of explanatory power for the spatial differentiation of mosquito density after the superposition of two factors [4].

Second, the difference in time lag scale is significant. The high-intensity interactions screened by GWRF all come from 2-month lag, indicating that the synergistic effect of temperature-pressure needs a long time of accumulation through the generational development of mosquitoes before it can be reflected in the adult density. The high explanatory power interactions of the geodetector are concentrated in the current month and 1-month lag, reflecting that the short-term coupling of humidity with other elements can directly explain the spatial differentiation characteristics of mosquito density. The two models differ in sensitivity to time signals: GWRF is better at capturing long-period accumulated interactions, and the geodetector more easily identifies interaction signals across short-term time scales.

DOI: 10.54254/2753-8818/2026.ES38197

Third, the numerical magnitude of interaction intensity differs greatly. The maximum interaction intensity of GWRF reaches 0.997, the top 10 interactions are all above 0.947, and the interaction changes of the temperature-related variables of 2-month lag contribute extremely high to the model prediction results. The maximum interaction q value of the geodetector is only 0.537; even if multiple factors form a nonlinear enhancement effect, meteorological elements alone still cannot fully explain the spatial differentiation of the density of *Anopheles sinensis*, and non-meteorological factors such as land use, control measures and host distribution also play a non-negligible role.

Combining the single-factor and interaction results, the action modes of different meteorological elements are different. At the single-factor level, humidity, precipitation and wind speed have both current-month and lag effects, temperature is prominently a current-month effect, and the independent contribution of pressure is limited; at the interaction level, under long-period conditions the temperature-pressure coupling plays a prominent role in density prediction, while at short-term scales the coupling of humidity with the other meteorological elements dominates the explanation of spatial differentiation. Relying only on the interaction results of a single model will lead to a one-sided understanding of the coupling mechanism of meteorological factors, and it is necessary to combine the results of multiple models for joint analysis.

## 4. Conclusions and prospects

### 4.1. Main conclusions

This study used the two analysis frameworks of GWRF-SHAP and the geodetector, combined with meteorological variables of multiple time scales, to analyze the meteorological driving mechanism of the density of *Anopheles sinensis*, and obtained the following main conclusions: (1) relative humidity, precipitation, wind speed and temperature are all the key meteorological elements regulating the population density of *Anopheles sinensis*, and the meteorological conditions have a time lag effect of 12 months on mosquito density. Relative humidity of 1-month lag has the highest contribution in the GWRF model, reflecting the important constraining effect of humidity conditions on adult survival. (2) The core driving factors identified by GWRF and the geodetector partially overlap, but the factor ranking and the interaction effect results differ significantly. GWRF is good at capturing strong synergistic interactions between temperature variables at long time scales; the geodetector is more inclined to identify short-term interaction combinations involving humidity. This difference comes from the algorithm mechanisms of the two models: GWRF focuses on the joint contribution of variables at the prediction level, and the geodetector focuses on the explanatory ability of factors for spatial stratified heterogeneity. When carrying out research on vector driving factors, a single model easily brings biased results, and it is recommended to use the two types of models in combination to achieve complementary advantages.

### 4.2. Research limitations and prospects

This study still has certain limitations. The dataset used in the study is dominated by meteorological variables, and human and landscape environmental variables such as land use, insecticide use intensity and population density are not included, which will limit the explanatory ability of the model for the real population pattern of mosquitoes to a certain extent. At the same time, this study only identified factor lag effects and interaction effects at the statistical level, and did not combine laboratory controlled experiments to verify the relevant ecological mechanisms.

DOI: 10.54254/2753-8818/2026.ES38197

Follow-up research can be further expanded from three aspects: first, integrate multi-source data such as land use and vector control to improve the input variables of the GWRF model and further improve the explanatory and predictive performance of the model; second, combine field controlled observations to verify the action mechanism of the multi-factor coupling of humidity, temperature and precipitation on the life history of *Anopheles sinensis*; third, rely on the GWRF model to build a dynamic early warning tool for the regional-scale density of *Anopheles sinensis*, and combine local weather forecasts to output vector risk zoning, providing technical support for the refined vector control in China after malaria elimination.

## References

[1] Zhao, T., & Xue, R.-D. (2024). Vector biology and integrated management of malaria vectors in China. *Annual Review of Entomology, 69*, 333–354.

[2] Wang, Z., Jia, X., Wang, Y., et al. (2025). Spatial variation of population, density, and composition of mosquitoes in mainland China. *Scientific Data, 12*(1), 20.

[3] Tian, F., He, J., Shang, S., et al. (2023). Survey of mosquito species and mosquito-borne viruses in residential areas along the Sino–Vietnam border in Yunnan Province in China. *Frontiers in Microbiology, 14*.

[4] Djègbè, N. D. C., Da, D. F., Somé, B. M., et al. (2024). Anopheles aquatic development kinetic and adults' longevity through different seasons in laboratory and semi-field conditions in Burkina Faso. *Parasites and Vectors, 17*(1), 181.

[5] Jiang, D., Wang, K., Chen, S., et al. (2026). Climate-driven distribution and ecological niche modeling of three Anopheles species in China using the Biomod2 ensemble framework. *Tropical Medicine and Infectious Disease, 11*(7), 189.

[6] Jia, X., Wang, Y., Wang, Z., et al. (2025). New insights into the main meteorological factors affecting mosquito densities: Time lags, interaction effects, and species specificity. *Environmental and Ecological Statistics, 32*(3), 1091–1113.

[7] Georganos, S., Grippa, T., Niang Gadiaga, A., et al. (2019). Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International, 36*, 121–136.

[8] Muschalik, M., Baniecki, H., Fumagalli, F., et al. (2024). shapiq: Shapley interactions for machine learning. *ArXiv, abs/2410.01649*.