Meteorological Factors–Based Spatio-Temporal Mapping and Predicting Malaria in Central China

Fang Huang, Shuisen Zhou,* Shaosen Zhang, Hongwei Zhang, and Weidong Li

Malaria Department, National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention; World Health Organization Collaborating Centre for Malaria, Schistosomiasis and Filariasis; Laboratory of Parasite and Vector Biology, Ministry of Health, Shanghai, People’s Republic of China; Department of Parasitology, Henan Center for Disease Control and Prevention, Zhengzhou, People’s Republic of China; Department of Parasitology, Anhui Center for Disease Control and Prevention, Hefei, People’s Republic of China

Abstract. Despite significant reductions in the overall burden of malaria in the 20th century, this disease still represents a significant public health problem in China, especially in central areas. Understanding the spatio-temporal distribution of malaria is essential in the planning and implementing of effective control measures. In this study, normalized meteorological factors were incorporated in spatio-temporal models. Seven models were established in WinBUGS software by using Bayesian hierarchical models and Markov Chain Monte Carlo methods. \( M_1, M_2, \) and \( M_3 \) modeled separate meteorological factors, and \( M_4, \) which modeled rainfall performed better than \( M_1 \) and \( M_2, \) which modeled average temperature and relative humidity, respectively. \( M_5 \) was the best fitting models on the basis of deviance information criterion and predicting errors. The results showed that the way rainfall influencing malaria incidence was different from other factors, which could be interpreted as rainfall having a greater influence than other factors.

INTRODUCTION

Malaria is an important cause of death and illness in children and adults in tropical countries. According to World Malaria Report 2010, half of the world’s population was at risk of malaria and an estimated 225 million cases led to nearly 781,000 deaths in 2009.\(^1\) Despite significant reductions in the overall burden of malaria in the 20th century, the disease still represents a significant public health problem in China, especially in central parts with an unstable malaria profile.\(^2\)

Malaria was severely epidemic in central China, and the number of cases in these areas was up to 21.99 million, accounting for 91.2% of the total reported cases in the country in 1970. With active implementation of malaria control measures for more than 30 years, considerable success had been achieved and the cases decreased dramatically, and many counties have reached the standard of basic malaria elimination (incidence was less than 1/10,000).\(^3\) However, early in the 21st century, malaria re-emerged in these areas, especially in Anhui, Henan, and Hubei Provinces along the Huang-Huai River. A total of 64,178 malaria cases, 52,082 suspected cases, and 38 deaths were reported in 917 counties of 23 provinces in 2006 (this was the highest number of malaria cases in the 21st century). Afterwards, the re-emergence was effectively controlled by an extensive control strategy, but the number of malaria cases and the incidence in central China still has accounted for more than 60% of the total cases in China in recent years.\(^4\) Therefore, understanding the space-time distribution of malaria in central China is essential in planning and implementing effective control measures.

The main vector in these areas is \textit{An. sinensis}, which develops in accumulated water on the ground or in rivers and lakes. Many factors play important roles in the spatial and temporal distribution of malaria. These factors include population immunity, mosquito control measures, social and economic status, and climate variability.\(^5\) Meteorological factors, such as temperature, relative humidity, and rainfall, have potential influence for malaria prediction, and the relationship between meteorological factors and malaria incidence has varied greatly over time and in space.\(^5,12\) Temperature, rainfall, and relative humidity are major natural factors that affect the life cycle and breeding of mosquitoes. Temperature and relative humidity affect mosquito activity and breeding, and rainfall affects populations of mosquitoes by creating additional or decreasing breeding sites. Examples of these factors affected malaria include use of surveillance data to predict epidemics in Ethiopia, rainfall data to predict epidemics in Kenya, and the El Niño Southern Oscillation index to predict epidemics in Kenya and southern Asia.\(^13-16\)

Social and economical status have significantly changed since the 1990s in central China, and malaria control interventions have also changed from vector controls such as indoor residual spraying and insecticide-treated bed nets combined with case management to enhancing case detection and health education, particularly for populations at risk for malaria.\(^17\) The objectives of this study were to identify possible relationships between meteorological, social, and economic factors and malaria incidence and spatial and temporal similarities in counties in central China, identify the roles of rainfall, temperature, and humidity in driving spatio-temporal patterns of malaria incidence, and obtain evidence that intrinsic host-pathogen dynamics might also contribute to malaria incidence. These objectives would contribute to future mathematical modeling, which could help formulate methods for malaria monitoring, forecasting, and early warning.

MATERIALS AND METHODS

Study area. Anhui, Henan, and Hubei Provinces are in central China and have a total area of 429,600 km\(^2\). The population of the three provinces was 228.3 million in 2009 (Statistic Bureau of Anhui, Henan and Hubei Provinces). There are 281 counties and 57 meteorological monitoring stations (Figure 1) in the three provinces, which are located at 32°17′–34°18′N, 113°–117°09′E along the Huang-Huai River in central China (Figure 1). This area is hilly, and common
variables that follow a Poisson distribution. The distribution functions of Poisson distribution were

\[ Y_i \sim \text{Poisson}(\lambda_i) \]

where \( \lambda_i \) is the expected counts of cases of the disease, \( r_i \) is the relative risk, and \( n_i \) is the population of subregion \( i \). The expected counts of the cases of the disease and the population were known quantities and play an important role in standardizing the information from areas of different characteristics. Therefore, the unknown quantities of interests are relative risks \( r_i \).

Disease maps are based on maximum-likelihood estimate of \( r_i \), denoted here by \( \hat{r}_i = Y_i / n_i \), are misleading when either both the disease was rare or the areas are small.\(^2\) The spatial dependence was not accommodated in the maximum-likelihood estimates. In addition, the homogeneity within each group of data should also be considered when individual risks are expected to be equal within the area. There are many reports that focused on the modeling of relative risks, as mentioned in the introduction of this report. A classic disease mapping method was proposed, which models the log relative risk of subregion \( i \) as\(^2\)

\[ \log(r_i) = \alpha + \beta X_i + b_i + u_i \]  

where \( \alpha \) is a common intercept for the entire region, \( X_i \) is a vector of covariates, \( b_i \) and \( u_i \) are random effect terms, \( u_i \) is the unstructured noise term that follows a normal distribution, \( u_i \sim N(0, \sigma^2) \), and

\[ b_i \sim N \left( 0, \left( \frac{1}{\sum_{j \in \delta_i} W_{ij}} \right) \right) \]  

where \( \delta_i \) is the set of subregions that adjacent to subregion \( i \), and \( W_{ij} \) is the weight of neighboring subregion \( j \) and \( i \). Weight \( W_{ij} \) is set 1 when subregions \( i \) and \( j \) shares the same boundary and set 0 otherwise. The prior distribution of random effect term \( b_i \) is known as the conditional autoregressive prior, which is used to model the spatial dependence. The conditional autoregressive prior was used in spatial analysis.\(^2\)

There were successive observations of the numbers of cases of sub-region \( i \) that were also important in disease mapping. It was assumed that \( Y_{it} \) represents the number of cases of the disease of sub-region \( i \) at the observation time \( t \). The spatial model that describes above incorporating time effect is

\[ Y_{it} \sim \text{Poisson}(\lambda_{it}) \]

\[ \log(\lambda_{it}) = \alpha + \beta X_{it} + b_i + u_{it} \]  

where \( \alpha \) is a common intercept for the entire region, \( X_{it} \) is a vector of covariates, \( b_i \) and \( u_{it} \) are random effect terms, \( u_{it} \) is the unstructured noise term that follows a normal distribution, \( u_{it} \sim N(0, \sigma^2) \), and

\[ b_i \sim N \left( 0, \left( \frac{1}{\sum_{j \in \delta_i} W_{ij}} \right) \right) \]  

where \( \delta_i \) is the set of subregions that adjacent to subregion \( i \), and \( W_{ij} \) is the weight of neighboring subregion \( j \) and \( i \). Weight \( W_{ij} \) is set 1 when subregions \( i \) and \( j \) shares the same boundary and set 0 otherwise. The prior distribution of random effect term \( b_i \) is known as the conditional autoregressive prior, which is used to model the spatial dependence. The conditional autoregressive prior was used in spatial analysis.\(^2\)

There were successive observations of the numbers of cases of sub-region \( i \) that were also important in disease mapping. It was assumed that \( Y_{it} \) represents the number of cases of the disease of sub-region \( i \) at the observation time \( t \). The spatial model that describes above incorporating time effect is

\[ Y_{it} \sim \text{Poisson}(\lambda_{it}) \]

\[ \log(\lambda_{it}) = \alpha + \beta X_{it} + b_i + u_{it} \]  

The temporal effects can be specified in a number of ways, which will determine the prior distribution that was assumed for the term \( u_{it} \). Most of the studies on disease mapping of past decade do not deal with the spatial and temporal effects jointly.

Knorr-Held and Besag adopted the model (equations 4 and 5) to investigate lung cancer data, and \( b_i \) was fixed for all \( t = 1, \ldots, T \).\(^5\) Knorr-Held and Raser proposed a hierarchical model for space-time surveillance date on meningococcal disease incidence.\(^2\) Lawson proposed a susceptible, infectious, or recovered (SIR) model to analysis the spatio-temporal disease incidence was not accommodated in the maximum-likelihood estimation from areas of different characteristics. Therefore, the unknown quantities of interests are relative risks \( r_i \).

Disease maps are based on maximum-likelihood estimate of \( r_i \), denoted here by \( \hat{r}_i = Y_i / n_i \), are misleading when either both the disease was rare or the areas are small.\(^2\) The spatial dependence was not accommodated in the maximum-likelihood estimates. In addition, the homogeneity within each group of data should also be considered when individual risks are expected to be equal within the area. There are many reports that focused on the modeling of relative risks, as mentioned in the introduction of this report. A classic disease mapping method was proposed, which models the log relative risk of subregion \( i \) as\(^2\)

\[ \log(r_i) = \alpha + \beta X_i + b_i + u_i \]  

where \( \alpha \) is a common intercept for the entire region, \( X_i \) is a vector of covariates, \( b_i \) and \( u_i \) are random effect terms, \( u_i \) is the unstructured noise term that follows a normal distribution, \( u_i \sim N(0, \sigma^2) \), and

\[ b_i \sim N \left( 0, \left( \frac{1}{\sum_{j \in \delta_i} W_{ij}} \right) \right) \]  

where \( \delta_i \) is the set of subregions that adjacent to subregion \( i \), and \( W_{ij} \) is the weight of neighboring subregion \( j \) and \( i \). Weight \( W_{ij} \) is set 1 when subregions \( i \) and \( j \) shares the same boundary and set 0 otherwise. The prior distribution of random effect term \( b_i \) is known as the conditional autoregressive prior, which is used to model the spatial dependence. The conditional autoregressive prior was used in spatial analysis.\(^2\)

There were successive observations of the numbers of cases of sub-region \( i \) that were also important in disease mapping. It was assumed that \( Y_{it} \) represents the number of cases of the disease of sub-region \( i \) at the observation time \( t \). The spatial model that describes above incorporating time effect is

\[ Y_{it} \sim \text{Poisson}(\lambda_{it}) \]

\[ \log(\lambda_{it}) = \alpha + \beta X_{it} + b_i + u_{it} \]  

The temporal effects can be specified in a number of ways, which will determine the prior distribution that was assumed for the term \( u_{it} \). Most of the studies on disease mapping of past decade do not deal with the spatial and temporal effects jointly.

Knorr-Held and Besag adopted the model (equations 4 and 5) to investigate lung cancer data, and \( b_i \) was fixed for all \( t = 1, \ldots, T \).\(^5\) Knorr-Held and Raser proposed a hierarchical model for space-time surveillance date on meningococcal disease incidence.\(^2\) Lawson proposed a susceptible, infectious, or recovered (SIR) model to analysis the spatio-temporal dynamics of season outbreaks and detailed the modeling process and evaluation of spatio-temporal models.\(^2\)

The spatial effects \( b_i \) at \( t = 1, \ldots, T \) in equation 5 could be modeled as fixed or varies with time. This modeling results in

\[ b_i = b_i(t=1, \ldots, T) \]  

\[ b_i \in \{ b_{1i}, \ldots, b_{Ti} \} \]  

The approach used in this study models time effects as first-order auto-regressive AR (equation 1).
correlation between consecutive time periods, which can be assessed via a temporal correlation coefficient ρ. The prior distributions for the time effects \( u_t \) are specified as

\[
\begin{align*}
u_t &\sim \text{Normal}(0, \sigma_u^2) \\
\sigma_u & = \sigma_u \sqrt{1 - \rho^2}, \quad t = 1 \\
\mu (\mu (1), \sigma_u^2) & \sim \text{Normal}(\mu (1, \mu_u), \sigma_u^2), \quad t > 1
\end{align*}
\]

There was no information available for the remaining parameters. Vague inverse gamma distribution was adopted for the variance \( \sigma_u^2 \) and vague normal priors for all other parameters. The prior for the correlation coefficient follows a uniform distribution.

Temperature, relative humidity, and rainfall were the main environmental factors that affect the incidence of malaria. It was difficult to clarify which factor was the major factor. Best spatio-temporal model fitting and meteorological variables evaluating malaria transmission during 1990–2009 in central China were used in this study. Average temperature, relative humidity and rainfall were incorporated in the spatio-temporal model. There were seven models based on different combinations of climatic factors. The models are

\[
M_1 \log(r_i) = \alpha + \beta_1 \text{temp}^{\text{norm}}_i + b_i + u_i, \quad \text{temp}^{\text{norm}}_i = \frac{\text{temp}_i - 15.5417}{72.0000}\ 
\]

\[
M_2 \log(r_i) = \alpha + \beta_2 \text{rh}^{\text{norm}}_i + b_i + u_i, \quad \text{rh}^{\text{norm}}_i = \frac{\text{rh}_i - 72.0000}{9919.0000}\ 
\]

\[
M_3 \log(r_i) = \alpha + \beta_3 \text{rain}^{\text{norm}}_i + b_i + u_i, \quad \text{rain}^{\text{norm}}_i = \frac{\text{rain}_i - 9919.0000}{72.0000}\ 
\]

\[
M_4 \log(r_i) = \alpha + \beta_1 \text{temp}^{\text{norm}}_i + \beta_2 \text{rh}^{\text{norm}}_i + b_i + u_i, \quad \beta_1 = 0.7531, \beta_2 = 1.1695\ 
\]

\[
M_5 \log(r_i) = \alpha + \beta_1 \text{temp}^{\text{norm}}_i + \beta_2 \text{rh}^{\text{norm}}_i + \beta_3 \text{rain}^{\text{norm}}_i + b_i + u_i, \quad \beta_1 = 0.4451, \beta_2 = 1.1654, \beta_3 = 0.2626, \beta_4 = 3.0340\ 
\]

Markov Chain Monte Carlo simulation was used to obtain estimates of the posterior and predictive quantities of interest. The models were implemented by using Gibbs sampling in the free software package WinBUGS version V1.4.3.

**Model comparison.** Considering the dynamics of malaria, an important issue was whether meteorological factors had impact on the incidence of malaria. For each model, we used a burn-in period of 40,000 iterations and retained every 100th set of parameter values to obtain approximately posterior samples of size 4,000. All models were evaluated in terms of deviance information criterion (DIC), which was a hierarchical modeling generalization of the Akaike information criterion and Bayesian information criterion. Models were penalized by the value of B, which favors a good fit, and by (in common with Akaike information criterion and Bayesian information criterion) the effective number of parameters \( p_D \). Because \( B \) would decrease as the number of parameters in a model increases, the \( p_D \) term compensated for this effect by favoring models with a smaller number of parameters. The missing meteorological data were interpolated by Gaussian Kriging interpolation. The meteorological data were inputted into the spatio-temporal model fitting a proper model.

**RESULTS**

The population and number of malaria cases in the 281 counties during 1990–2009 in the study areas were incorporated in the model fitting. Considering the complication and uncertain quality of social and economical factors, only meteorological factors, including average temperature, relative humidity and rainfall, were used in the model fitting. All meteorological data for those counties without meteorological stations were interpolated by Gaussian Kriging interpolation from the data obtained at the 57 meteorological stations. These data were normalized to evaluate influence on malaria.

In the modeling, each meteorological factor was normalized by its average value in 281 counties for 20 years. The average value of each factor was

\[
\begin{align*}
\text{temp}_{\text{avg}} &= 15.5417 \text{ (°C)} \\
\text{rh}_{\text{avg}} &= 72.0000\% \\
\text{rain}_{\text{avg}} &= 9919.0000 \text{ (mm)}
\end{align*}
\]

The interval of each normalized meteorological data was

\[
\begin{align*}
\text{temp}^{\text{norm}}_i &= (0.7531, 1.1695) \\
\text{rh}^{\text{norm}}_i &= (0.4451, 1.1654) \\
\text{rain}^{\text{norm}}_i &= (0.2626, 3.0340)
\end{align*}
\]
The DIC of each model is shown in Table 1. $M_1$, $M_2$, and $M_3$ modeled average temperature, relative humidity, and rainfall, respectively. $M_3$, which modeled rainfall, was better fitting than the other two models. $M_2$, which models relative humidity, had the highest DIC value. This model could then interpret that malaria transmission was more sensitive to change in rainfall. Two of the three meteorological factors were grouped in the fitting of $M_4$–$M_6$. $M_5$ was the best fit of $M_4$–$M_6$. This finding may be obvious because the DIC value when modeling rainfall and average temperature was less than that of relative humidity. According to the DIC value, $M_7$ was considered as the best model.

The weights of the three meteorological factors of $M_1$–$M_7$ are shown in Table 2. The sign of $\beta_1$, $\beta_2$, and $\beta_3$ were not changed over the seven models with a confidence level of 97.5%. $\beta_1$, $\beta_2$, and $\beta_3$ were relative stable in the models. The relationship between single meteorological factors and malaria incidence is shown in Figure 2. The x-axis was the interval of normalized meteorological factors, and the y-axis was the quantized affect on malaria incidence when meteorological factors varied. The parameter of $\beta_1$, $\beta_2$, and $\beta_3$ were taken from $M_7$.

The profile between the actual incidence and fitted incidence from during 1990–2009 based on meteorological data is shown in Figure 3. Although under different fittings according to variable DIC values, there was no significant difference between actual incidences and fitted incidences. However, when models were used to predict malaria incidence, the effect of predicting varied dramatically under different models with variable DIC values. The model example with meteorological data for 1990 for predicting the actual incidence and predicted incidence are shown in Figure 4. $M_7$ was best predicting model. The variance of predicting error for a random variable was

$$
\varepsilon = E\left[(Y - \hat{Y})^2\right]
$$

The approximately variance of predicting error for $M_7$–$M_7$ is shown in Table 3. The 2009 prediction maps by different models with 2008 data of malaria incidence is shown in Figure 5.

**DISCUSSION**

Malaria is a vector-borne disease with a limited extent of space and time. Uneven population distribution, socioeconomic structure, and mosquito breeding sites create a spatial clustering in malaria. However, the fluctuation in mosquito populations resulting from meteorological variables such as rainfall, temperature, and humidity has formed a temporal cluster in malaria. This work indicated the important roles of rainfall, temperature, and humidity in driving spatio-temporal
patterns of malaria incidence in central China and provided evidence that intrinsic host-pathogen dynamics might also contribute to the scene for future mathematical modeling.

In this report, seven models under different DIC values were established and compared by Bayesian hierarchical models and Markov Chain Monte Carlo methods based on meteorological factors. Among the fitting models of single meteorological factor, the model $M_3$, in which incorporating rainfall was better than in the other two models, implies that rainfall was more sensitive to malaria incidence. For incidence prediction, $M_3$ was best model. The year and data were two important factors in the accuracy of predicting. The variance of predicting error (Table 3) showed that model $M_3$ performed best with the data for 1990. All models in this study were used to predict the malaria incidence of 2009 using data of 2008. The spatial clustering and malaria transmission trend of 2009 were mapped for all models and illustrated.

The influence of meteorological factors on malaria incidence was complex. The values of $\beta_1$ and $\beta_2$ were less than 0, which means that malaria incidence decreased when the value of average temperature and relative humidity increased. However, the value of $\beta_3$ was more than 0, which indicated that the influence of rainfall on malaria incidence was different from that of other factors. It was understandable that malaria incidence increased when average temperature increased, but this was not the case in this study. Malaria incidence was proportional to rainfall when other factors were fixed, and the combination influence of these three meteorological factors on malaria incidence was not evident. Malaria incidence was affected not only meteorological factors but also social factors as interventions.

To improve the accuracy of models in fitting and predicting, data for malaria incidence was obtained from case reports.
in the Disease Reporting Information System through the internet in township hospitals and county centers for disease control throughout China. Although malaria case reports were of high quality and reliable, before reporting to the system, case investigations and confirmations must be conducted by the county center for disease control after reports are received from township hospitals. According to national guidelines, blood smear–positive cases and clinic cases in the community should be included in the routine reporting data. However, a large proportion of confirmed cases and only few clinic diagnostic cases were reported to the system because case confirmation was enhanced in the Malaria Elimination Strategy in China. Under implementation of rounds of Global Fund programs in malaria-endemic areas, because awareness of self-protection and going to clinics was high, most fever cases were timely diagnosed by the clinics. Some malaria cases cannot be diagnosed by blood examination at clinics because of lack of microscopic equipment, but all malaria cases at clinics were requested to send blood smears to township hospitals for confirmation of the diagnosis. Thus, incidence data, including case diagnosis and data recording, at the clinics were reliable.

Despite the difficulties and limitations of the modeling process, models draw attention to the potential health impact of global climate changes. Models also play an important role in the systematic analysis of multiple cause-and-effect relationships on the basis of available knowledge. Also, identification of key gaps in data and knowledge are needed to improve analysis of these effects.34

As in other studies in Asia and Africa, a decreasing trend in malaria incidence in central China was identified.31–33 This finding was likely associated with general economic development, changing agricultural practices, increasing urbanization and better access to healthcare. Trends in malaria incidence are likely influenced by interventions. In addition to support from the ongoing national malaria control program, Anhui, Hubei, and Henan Provinces have been recipients of three rounds of funding from the Global Fund to Fight AIDS, Tuberculosis, and Malaria.35

Socioeconomic and housing factors also played an important role in malaria transmission, including the presence of open eaves or the lack of ceilings, population density and the presence of animal close to the house, education, and available income in households.36 However, in this study, these socioeconomic factors were not taken into account. In addition, it was reported that there were likely imperfections in the data because they were obtained in a passive surveillance system.

These findings should be considered in future malaria prevention and control projects, especially during malaria outbreaks.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>( \varepsilon ) values of ( M_1 )–( M_7 ) for malaria, Anhui, Henan, and Hubei Provinces, China</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>( M_1 )</td>
</tr>
<tr>
<td>Variance</td>
<td>36.47</td>
</tr>
</tbody>
</table>
and in areas of re-emergence of malaria. In 2009, an action plan for malaria elimination was proposed by the Ministry of Health in China. Therefore, it is important for authorities to know how to distribute limited resources effectively.

Received March 20, 2011. Accepted for publication June 6, 2011.

Financial support: This study was supported by the special social commonweal research programs of the Ministry of Science and Technology of the People’s Republic of China (2005 DIB1J092) and the National S. and T. Mayor Project (no. 2008ZX10004-011).

Disclaimer: None of the authors have any conflicts of interest.

Authors’ addresses: Fang Huang, Shuisen Zhou, and Shaosen Zhang, National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Ministry of Health, Shanghai, People’s Republic of China, E-mails: huangfang78@yahoo.com.cn, ccdczss@sh163.net, and shaosen413@163.com. Hongwei Zhang, Department of Parasitology, Henan Center for Disease Control and Prevention, Zhengzhou, People’s Republic of China, E-mail: zhwei69@163.com. Weidong Li, Department of Parasitology, Anhui Center for Disease Control and Prevention, Hefei, People’s Republic of PR China, E-mail: ahcdclwd@163.com.

REFERENCES