RESEARCH ARTICLE

Bayesian maximum entropy-based prediction of the spatiotemporal risk of schistosomiasis in Anhui Province, China

Fuju Wang[1], Xin Liu[1*], Robert Bergquist[5], Xiao Lv[1], Yang Liu[1], Fenghua Gao[3], Chengming Li[4] and Zhijie Zhang[2*]

Abstract

Background: "Schistosomiasis" is a highly recurrent parasitic disease that affects a wide range of areas and a large number of people worldwide. In China, schistosomiasis has seriously affected the life and safety of the people and restricted the economic development. Schistosomiasis is mainly distributed along the Yangtze River and in southern China. Anhui Province is located in the Yangtze River Basin of China, with dense water system, frequent floods and widespread distribution of *Oncomelania hupensis* that is the only intermediate host of schistosomiasis, a large number of cattle, sheep and other livestock, which makes it difficult to control schistosomiasis. It is of great significance to monitor and analyze spatiotemporal risk of schistosomiasis in Anhui Province, China. We compared and analyzed the optimal spatiotemporal interpolation model based on the data of schistosomiasis in Anhui Province, China and the spatiotemporal pattern of schistosomiasis risk was analyzed.

Methods: In this study, the root-mean-square-error (RMSE) and absolute residual (AR) indicators were used to compare the accuracy of Bayesian maximum entropy (BME), spatiotemporal Kriging (STKriging) and geographical and temporal weighted regression (GTWR) models for predicting the spatiotemporal risk of schistosomiasis in Anhui Province, China.

Results: The results showed that (1) daytime land surface temperature, mean minimum temperature, normalized difference vegetation index, soil moisture, soil bulk density and urbanization were significant factors affecting the risk of schistosomiasis; (2) the spatiotemporal distribution trends of schistosomiasis predicted by the three methods were basically consistent with the actual trends, but the prediction accuracy of BME was higher than that of STKriging and GTWR, indicating that BME predicted the prevalence of schistosomiasis more accurately; and (3) schistosomiasis in Anhui Province had a spatial autocorrelation within 20 km and a temporal correlation within 10 years when applying the optimal model BME.

Conclusions: This study suggests that BME exhibited the highest interpolation accuracy among the three spatiotemporal interpolation methods, which could enhance the risk prediction model of infectious diseases thereby providing scientific support for government decision making.

*Correspondence: xinliu1969@126.com; epistat@gmail.com
[1] College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China
[2] School of Public Health, Fudan University, Shanghai 200032, China
Full list of author information is available at the end of the article

www.manaraa.com

Wang *et al. BMC Infectious Diseases*   (2021) 21:1171

Page 2 of 12

Background

Schistosomiasis, an important zoonotic parasitic disease caused by three main (and three less common) species of the trematode worm Schistosoma, is reported from 78 countries on the tropical and subtropical parts of the world where it affects more than 200 million people [1]. *Schistosomiasis japonicum* is endemic in China [2], where its endemic areas are classified into three types based on geographical topography and the ecological characteristics of breeding areas of the only intermediate snail host *Oncomelania*: lakes and swamp areas, plain areas of waterway networks, and hilly and mountainous areas [3, 4]. Compared with the latter two types of area, schistosomiasis control has proven difficult in the lake and swamp areas because of the widespread distribution of breeding areas and difficult-to-control water levels, where over 80% of schistosomiasis cases occur [5, 6]. Frequent flooding of the Yangtze River that runs through Anhui Province forming lakes and swamps adds to the problem. The large number of livestock, such as cattle and sheep that play the role of reservoir hosts in endemic areas, exacerbate the difficulty of controlling transmission of the disease facilitating the persistence of schistosomiasis in the country. This situation contributes to the great significance of the disease and the need to study its risk potential in Anhui Province [2].

Because of the large workload associated with schistosomiasis control, the number of surveillance areas varies from year to year. This causes results in incomplete and irregular schistosomiasis data that not only poses an obstacle for control efforts but also affects people's judgment of the schistosomiasis risk potentially leading to unsafe and hazardous behaviour [7]. Data interpolation is the primary approach to solving the problem of missing spatiotemporal data regarding the risk of schistosomiasis. However, as traditional data interpolation methods tend to study temporal or spatial interpolation separately which makes a global view elusive. Commonly used spatial interpolation methods include inverse distance-weighted interpolation [8] and Kriging interpolation [9] to convert data from discrete points into a continuous data surface. From a spatiotemporal analysis point of view, however, one-sided spatial interpolation analyzes confines the analysis to a particular point or period of time which destroys the uniformity of the spatiotemporal continuum. On the other hand, temporal interpolation is to interpolate the observed time series; commonly used such methods include the autoregressive model [10], the autoregressive moving average model [11] and the generalized additive model [12]. Time series analysis of spatiotemporal data alone greatly reduces the pure spatial correlation. These shortcomings have led to spatiotemporal interpolation methods, which are widely used today for the estimation of missing spatiotemporal datasets, generating high-precision, spatiotemporal surfaces expressing spatiotemporal processes and distributions [3, 13, 14]. The main spatiotemporal interpolation methods are spatiotemporal Kriging (STKriging) [3, 13, 14] and regression-based methods [15–19], such as geographical and temporal weighted regression (GTWR) and Bayesian maximum entropy (BME). Both STKriging interpolation and GTWR interpolation have been applied to the study of schistosomiasis, the former of which has divides the prevalence of schistosomiasis into spatiotemporal trends and residuals [13], whereas the latter of which has used factors that affect schistosomiasis to fit the prevalence [20]. The spatiotemporal variation functions of the residuals are first established and the residuals of the prevalence of schistosomiasis are then predicted based on that, with the final interpolation results obtained by summing the spatiotemporal trends and the predicted residual. Considering the characteristics of the above two interpolation methods, this study attempts to improve the accuracy of predicting the risk of schistosomiasis by taking into account the factors influencing the prevalence value when attempting to predict it around the point to be estimated based on the values measured. This approach is the BME method [21–34], and it refers to high-order statistical estimation of the spatiotemporal prevalence phenomenon and predicts the risk of disease at the point to be estimated based on soft data (e.g., data fitted according to mathematical or statistical methods, uncertain, subjective or qualitative data) and hard data (data actually measured around the point to be estimated). The BME method has been successfully applied to infectious diseases such as syphilis [25], hand-foot-mouth disease [27], influenza [23, 38], dengue fever [30, 32] and Black Death [24]. However, it has been rarely been applied to the prediction of the risk of schistosomiasis.

The prevalence of schistosomiasis is interpolated in Anhui Province with BME in the study. The

Wang *et al. BMC Infectious Diseases*   (2021) 21:1171

Page 3 of 12

spatiotemporal pattern of schistosomiasis risk is analyzed based on interpolation results.

Methods

Study area

Anhui Province is located in southern China, with an area of approximately 140,100 km² and had a population of approximately 636,570,000 in 2019. The province is crossed by the Huai River in the north and the Yangtze River in the south. Climatically, it is a transitional area between tropical and subtropical zones, with warm temperatures north of the Huai River and a subtropical zone south of the Huai River. There is a pronounced monsoon climate with strong rains in June and July, which often lead to flooding. Such geography and climate are well suited for the growth and reproduction of the schistosomiasis-related Oncomelania snails. Thus, Anhui Province is of key concern for control of this disease, and study therefore focused on the endemic lakes and swamps along the Yangtze River Basin as it traverses the province.

Prevalence data

Data on the prevalence of schistosomiasis infection in Anhui Province between 2000 and 2015 were obtained from field surveys conducted by professional health workers at the Anhui Institute of Parasitic Diseases (AIPD) [3]. A two-step diagnostic approach was used annually to identify cases of schistosomiasis infection: serology was done for all people aged 5 to 65 years in endemic villages using the indirect hemagglutination test (IHT) followed by faecal Kato-Katz parasitological test for those with positive blood test results [35]. The results were reported to the AIPD through the county office [3]. The study covered 29 counties between 2000 and 2015 (Fig. 1).

Environmental data

As transmission of schistosomiasis [13] is closely related to the presence of Oncomelania intermediate host snails in the natural environment as well as social factors, such as urbanization, which combine to influence the spatiotemporal distribution of schistosomiasis risk. In the present study, daytime land surface temperature (LSTd), night-time land surface temperature (LSTn), the normalized difference vegetation index (NDVI), meteorological data (precipitation, mean minimum temperature (MTmin), mean maximum temperature (MTmax) and sunshine hours), soil data (soil moisture, soil pH and soil bulk density), distance from the Yangtze River and the urbanization level (using night-time light data to describe the county-level urbanization level [36]) were chosen as the main factors influencing schistosomiasis presence.
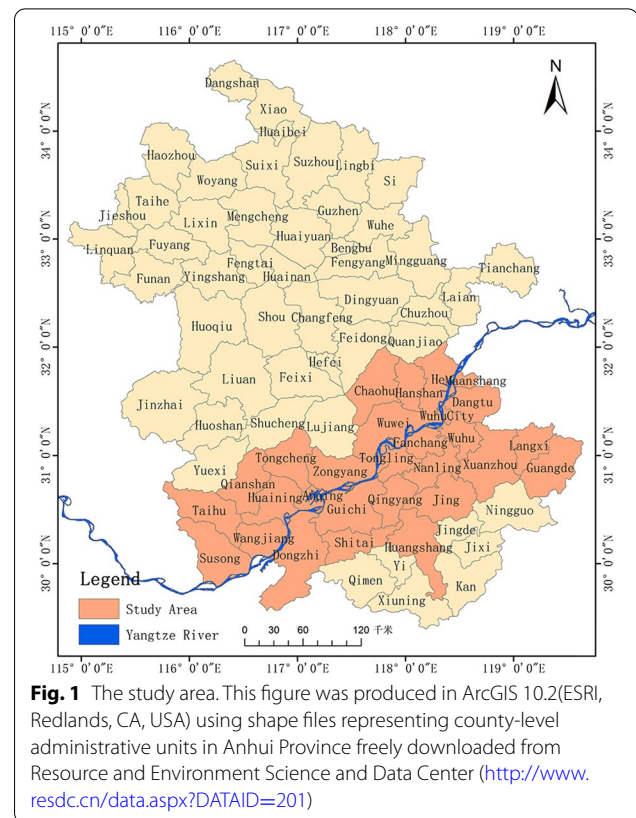

Fig. 1 The study area. This figure was produced in ArcGIS 10.2(ESRI, Redlands, CA, USA) using shape files representing county-level administrative units in Anhui Province freely downloaded from Resource and Environment Science and Data Center (http://www.resdc.cn/data.aspx?DATAID=201)

LSTd, LSTn and NDVI data were obtained from the Level-1 and Atmosphere Archive & Distribution System (LAADS) website (https://ladsweb.modaps.eosdis.nasa.gov/), with a temporal resolution of 1 month and a spatial resolution of 1 km. ENVI Software (version 5.2, Research System Inc. (RSI), Boulder, CO, USA) was used for cropping and stitching the above data. The LSTd, LSTn and NDVI data of each county are pixel accumulation, respectively. Then, the monthly pixel average value of each county is calculated by using the zoning statistical function of ArcGIS software (version 10.4, ESRI Inc., Redlands, CA, USA), to produce county attribute tables. Finally, the annual average value of each county is calculated. Meteorological data, including precipitation, MTmin, MTmax and sunshine hours, came from the website of the China Meteorological Administration (http://data.cma.cn/) with a time resolution of one month. Meteorological data for Anhui Province with a spatial grid of 1 km × 1 km were obtained through Kriging interpolation. The monthly averages for each county were calculated using ArcGIS zoning statistics followed by production attribute tables and calculation of the annual averages for each county.

Soil moisture data were obtained from the European Space Agency (https://www.esa-soilmoisture-cci.org/)

Wang *et al. BMC Infectious Diseases* (2021) 21:1171

Page 4 of 12

with a temporal resolution of 1 day and a spatial resolution of 0.25°. Soil pH and soil bulk density data were obtained from the Cold and Arid Region Scientific Data Center (http://westdc.westgis.ac.cn/data), with a temporal resolution of 1 year and spatial resolution of 30 arcseconds. The obtained data were dealt as described above for the LSTd, LSTn, NDVI and the meteorological data.

Data regarding the distance to the Yangtze River came from the World Wildlife Fund (https://www.worldwildlife.org/) and the Euclidean distance measurement tool in ArcGIS software were used to calculate the distance from the geometry center to the Yangtze River in each county.

The night-time light data included the Defense Meteorological Program Operational Line-Scan System (DMSP-OLS) with its unique capability to detect visible and near-infrared light emission and the Visible Infrared Imaging Radiometer Suite (*VIIRS*), both instruments operated by the US National Oceanic and Atmospheric Administration (NOAA) (http://ngdc.noaa.gov/eog/download.html). We used DMSP-OLS data with a spatial resolution of 1 km, a temporal resolution of 1 year, and a time range of 1992–2013, where transient light, such as lightning and natural gas flares, had been removed. The National Polar-Orbiting Partnership's Visible Infrared Imaging Radiometer Suite (NPP-VIIRS) data had a spatial resolution of 500 m, a temporal resolution of 1 month, and a time range of 2000–2015. The noise in the NPP-VIIRS data was first removed, after which the night-time light index was calculated for each county; finally, the NPP-VIIRS night-time light index was converted to the DMSP-OLS night-time light index using a curve fitting method [36].

Statistical analysis

A univariate analysis of the influencing factor data with the prevalence of schistosomiasis was performed first to exclude variables with $P > 0.1$[13]; then, a multicollinearity test was conducted on the remaining variables with the variance inflation factor (VIF) index < 5; finally, backward stepwise regression modelling was carried out with $P > 0.1$ and $P \leq 0.05$ as the exit and entry criteria, respectively [13]. When modelling, the data were randomly divided into ten equal parts, nine of which were used as training data; the remaining part was used as the test validation data. The BME, STKriging, and GTWR interpolation analyses were performed to select the optimal interpolation model for analysis of the spatiotemporal patterns of schistosomiasis (see the Appendix for a description of the three methods). Statistical analyses (univariate analysis, multicollinearity test and backward stepwise regression) were carried out using the statistical software SPSS version20. BME computations were performed with the software SEKS-GUI v1.0.8 [22]. GTWR

computation were carried using the software ArcGIS version10.4. STKriging were implemented in the R package gstat [37]

The RMSE and absolute residual (AR) of the validation datasets were used for comparison of the prediction accuracy of the different methods, assessing the prediction accuracy of the model over the whole study area, as well as in each county.

$$RMSE = \sqrt{\frac{1}{16n} \sum_{t_i=2000}^{2015} \sum_{i=1}^{n} y(u_i, v_i, t_i) - \hat{y}(u_i, v_i, t_i)^2}$$
(1)

$$AR = \left| y(u_i, v_i, t_i) - \hat{y}(u_i, v_i, t_i) \right|$$
(2)

where (u_i, v_i, t_i) are the spatiotemporal coordinates of the geometric centre of the i-th county; u_i, v_i, and t_i the longitude, latitude and time coordinates of the i-th sample point, respectively; $y(u_i, v_i, t_i)$, and $\hat{y}(u_i, v_i, t_i)$ the observed and predicted values of the prevalence of schistosomiasis in the i-th county in Anhui Province, respectively; and n the number of counties in the province where schistosomiasis still is endemic.

ArcGIS software is used to analyze the temporal and spatial changes of the RMSE and AR.

Results

Analysis of influencing factors and projected results

Table 1 shows the results of the backward stepwise regression significance test on the data of influencing factors of schistosomiasis in Anhui Province between 2000 and 2015. The LSTd, MTmin, NDVI, soil moisture, night-time light, and soil bulk density were included in the model; their VIF values were all less than 5, and their P-values less than 0.05. This indicates that the collinearity between these influencing factors is small and that there is a significant relationship between these factors and the prevalence of schistosomiasis in the province.

The GTWR model was then used to fit the significant influencing factors with regard to the prevalence of schistosomiasis. The goodness-of-fit $R^2 = 0.76$ indicates that the GTWR model can reveal 76% of

Table 1 Significance test results of influencing factors

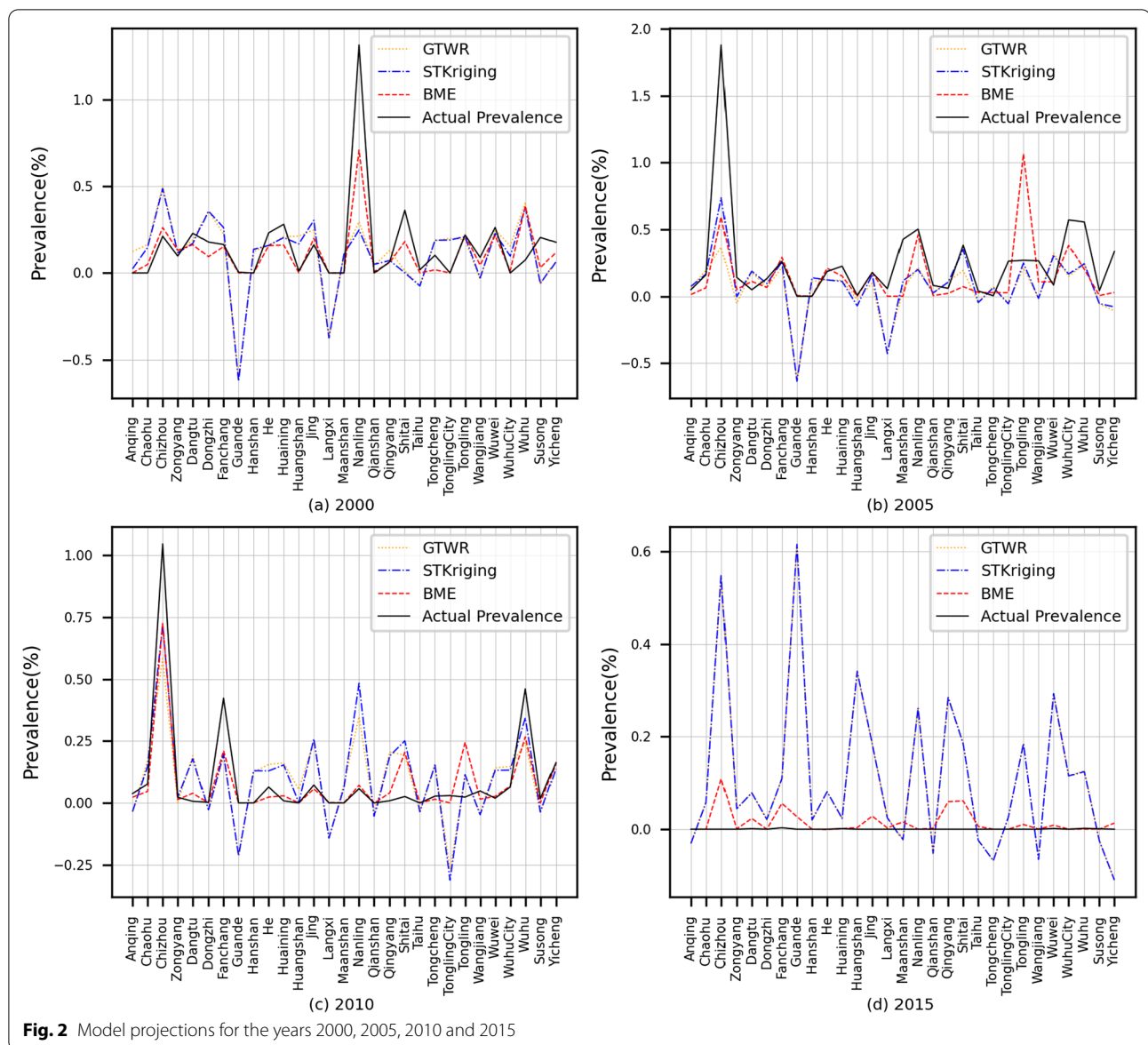| Influencing factor	*P*	VIF
LSTd	0.001	1.164
MTmin	0.041	1.034
NDVI	0.003	1.217
Soil moisture	0.0005	1.190
Night-time light data	0.034	1.033
Soil bulk density	0.000014	1.091

Wang *et al. BMC Infectious Diseases*   (2021) 21:1171

Page 5 of 12

Fig. 2 Model projections for the years 2000, 2005, 2010 and 2015

the spatiotemporal variation in the prevalence of schistosomiasis.

Figure 2 shows a comparison between the predicted and observed values of the three different interpolation methods for the years 2000, 2005, 2010 and 2015. The trends between the predicted and actual values of the three methods were basically consistent. The BME predictions showed a better fit at the maximum and minimum prevalence. In contrast, STKriging and GTWR interpolation results showed a poor fit at the maximum and minimum prevalence and the prediction results were poorer for years with lower prevalence, such as 2015 (e.g., these data were mostly 0).

Comparison of prediction accuracy

The RMSE value of the ten-fold cross-validation of BME was 0.148, which was 0.071 and 0.087 lower than that of the STKriging (RMSE=0.219) and GTWR (RMSE=0.235), respectively. This suggests that the interpolation accuracy of BME is better than that of STKriging and GTWR. Figure 3 shows the RMSE comparison of the three interpolation methods for the annual prevalence of schistosomiasis in the 29 counties in Anhui Province. Overall, the RMSE of the BME interpolation method was lower than that of STKriging and GTWR, and the RMSE of STKriging was lower than that of GTWR in most years; however, these values were similar to each other.
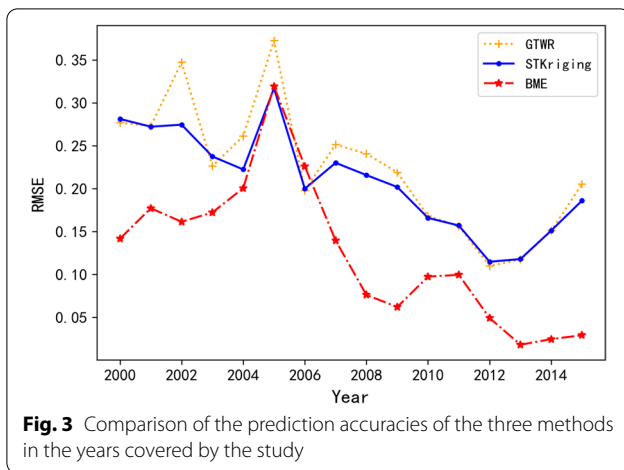
Wang *et al. BMC Infectious Diseases*   (2021) 21:1171

Page 6 of 12

Fig. 3 Comparison of the prediction accuracies of the three methods in the years covered by the study
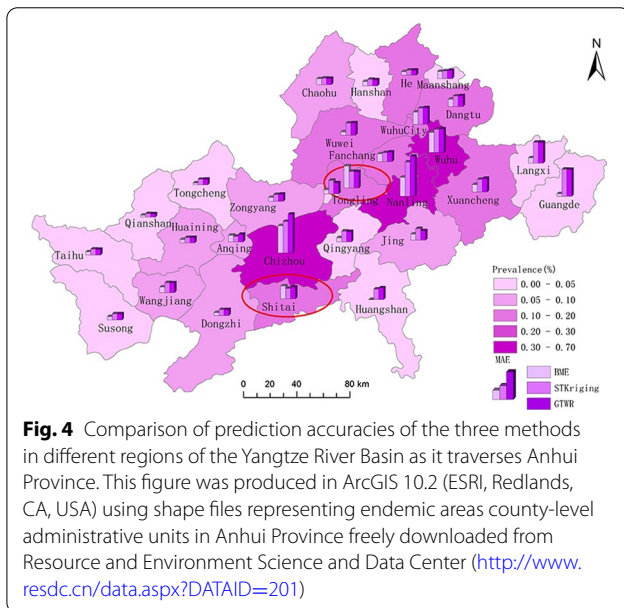

Fig. 4 Comparison of prediction accuracies of the three methods in different regions of the Yangtze River Basin as it traverses Anhui Province. This figure was produced in ArcGIS 10.2 (ESRI, Redlands, CA, USA) using shape files representing endemic areas county-level administrative units in Anhui Province freely downloaded from Resource and Environment Science and Data Center (http://www.resdc.cn/data.aspx?DATAID=201)

Figure 4 shows a comparison of the RMSEs of the three methods for the prevalence of schistosomiasis in the 29 counties in Anhui Province. With the exception of counties Shitai and Tongling, the RMSEs of BME interpolation were lower than those of STKriging and GTWR. Meanwhile, the RMSEs of the STKriging interpolation method were slightly lower than those of the GTWR model in most districts and counties.

To investigate the spatial interpolation accuracy of the three models, the errors of the three models were compared for the years 2000, 2005, 2010, and 2015 (Fig. 5). Overall, the BME model showed larger errors in the surrounding areas, such as Chizhou County and Shitai County. However, in terms of spatial distribution, the interpolation accuracy of BME was overall higher

than that of STKriging and GTWR, whose spatial distributions of error were similar.

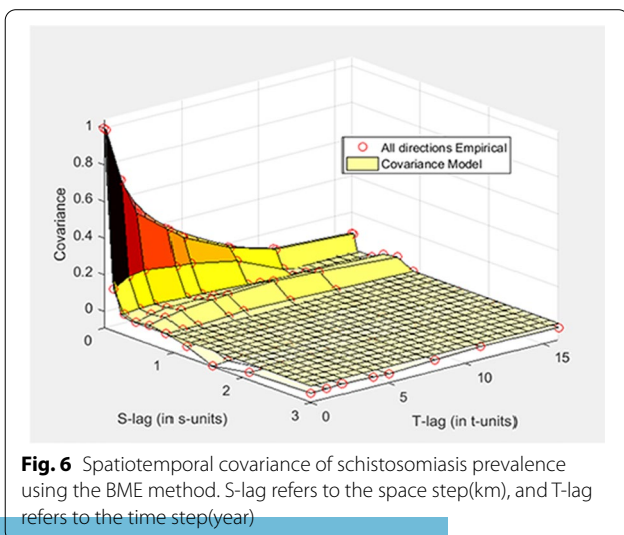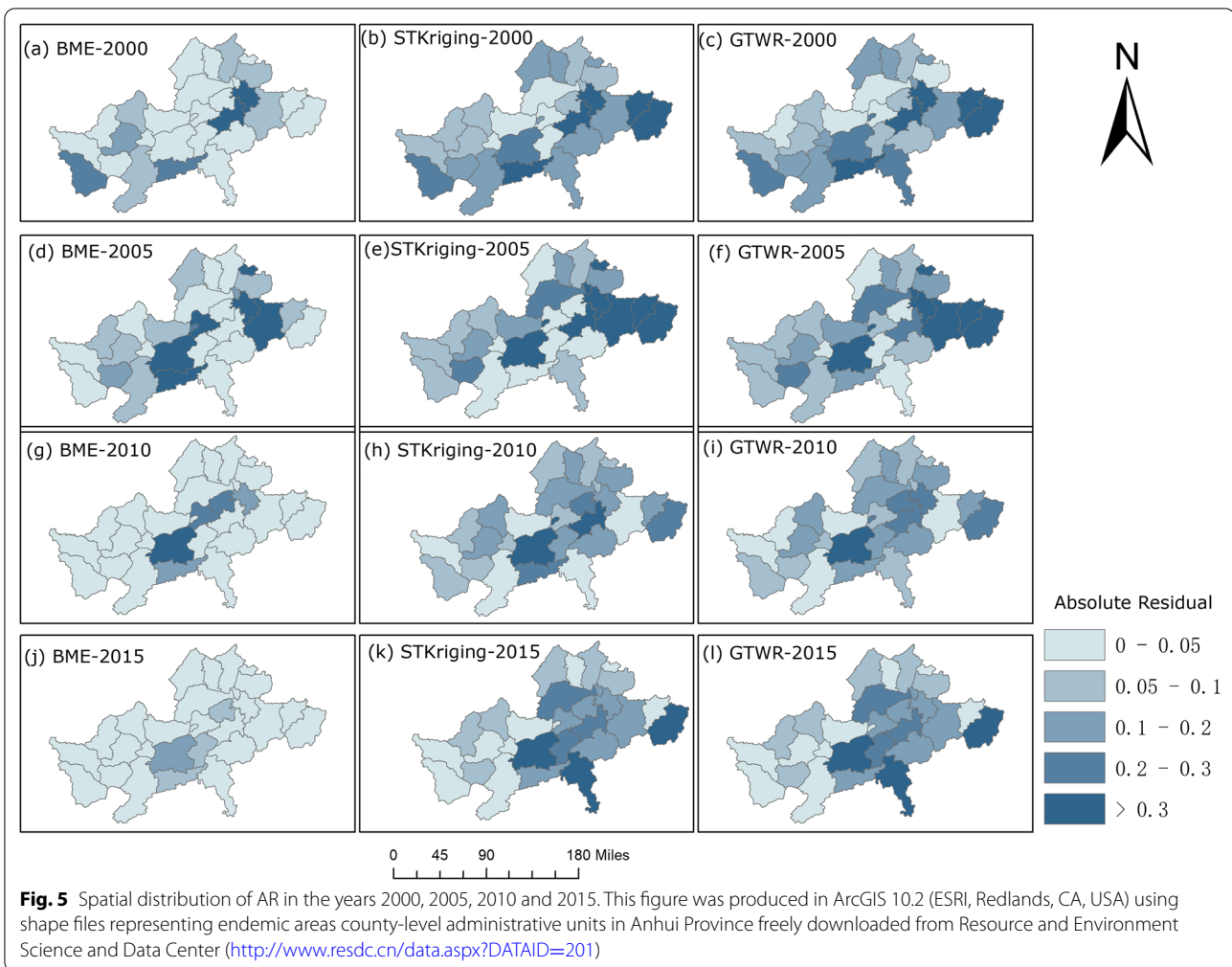**Analysis of the spatiotemporal patterns of schistosomiasis prevalence in Anhui Province**

The BME method showed the highest interpolation accuracy and was therefore used to examine the spatiotemporal patterns of schistosomiasis in Anhui Province. In this study, the temporal and spatial components of the combined exponential and spherical models were used to fit the spatiotemporal covariance of schistosomiasis prevalence as shown in Fig. 6. The spatiotemporal covariance model is expressed by Eq. (5) in the Appendix, where $c_1 = 0.55$, $c_2 = 0.45$, $a_{h_s1} = 0.1$ (ca. 10 km), $a_{h_t1} = 17$ years, $a_{h_s2} = 0.3$ (ca. 30 km), and $a_{h_t2} = 4$ years. A larger covariance value indicates a stronger spatiotemporal correlation of schistosomiasis. The figure shows that larger values of the spatiotemporal covariance occur within a spatial range of 0.20° (ca. 20 km) and a time range of 10 years.

Discussion

We first investigated and identified the significant influencing factors closely related to schistosomiasis in Anhui Province. We then used the prevalence predicted by the GTWR model as soft data and the measured prevalence in the field as hard data for BME interpolation.

As an introduction of the use of the BME method for the prediction of the prevalence of schistosomiasis, we compared BME with the STKriging and GTWR methods for spatiotemporal interpolation in Anhui Province, China. The GTWR model established the regression relationship between the influencing factors and the prevalence of schistosomiasis; the STKriging model took the predicted values of the GTWR model as spatiotemporal trends and interpolated the residuals; and the BME model used for the GTWR predicted values as soft data together with the actual measured prevalence data as hard data for interpolation.

Overall, the interpolation accuracy of BME was higher than that of STKriging and GTWR, with the accuracy of STKriging slightly higher than that of GTWR. The STKriging model divided the prevalence of schistosomiasis into two components, the spatiotemporal trend and the residuals. The former was expressed by the GTWR and the residuals, for which the trend was not available, by the STKriging interpolation, assuming that the residuals satisfied second-order stationarity. The STKriging model incorporated more information on schistosomiasis and was thus more accurate for prediction than the GTWR model. The BME, using soft and hard data during interpolation, where the former were the GTWR fitted data and the latter the prevalence data that had been

Wang *et al. BMC Infectious Diseases* (2021) 21:1171

Page 7 of 12

Fig. 5 Spatial distribution of AR in the years 2000, 2005, 2010 and 2015. This figure was produced in ArcGIS 10.2 (ESRI, Redlands, CA, USA) using shape files representing endemic areas county-level administrative units in Anhui Province freely downloaded from Resource and Environment Science and Data Center (http://www.resdc.cn/data.aspx?DATAID=201)

Fig. 6 Spatiotemporal covariance of schistosomiasis prevalence using the BME method. S-lag refers to the space step(km), and T-lag refers to the time step(year)

measured in the field, proved superior to both the other approaches. Importantly, the data did not need to meet the second-order stationarity and could automatically fit the nonlinear estimator [38, 39].

Spatially, all three models showed lower prediction accuracy in areas with high prevalence. Areas susceptible to the Yangtze River water level represent environments difficult to control. Examples of such places include Chizhou County, Shitai County, Nanling County and Wuhu County characterized by a multitude of rivers, lakes and beaches and the Oncomelania snails therefore widely distributed. The poverty alleviation policy in China have regrettably led to the development of a large snail areas results in some places with an increase schistosome-infected cattle and sheep acting as reservoirs of the disease. Consequently, the prevalence of schistosomiasis increases, with the complex influencing factors leading to a lower prediction accuracy. Only the BME model showed high prediction accuracy also in areas with low prevalence, such as Tongcheng County, Qianshan

Wang *et al. BMC Infectious Diseases*   (2021) 21:1171

Page 8 of 12

County, Taihu County, Susong County, Dongzhi County, Huangshan County, Langxi County, and Guangde County. Because of the low prevalence of schistosomiasis in these areas, there were continuous changes in significant influencing factors, leading to a high accuracy of BME soft data calculated based on the GTWR model.

From the temporal point of view, the interpolation results of all three models were basically consistent with the trends of actual prevalence measured, but the interpolation results of STKriging and GTWR models were poor after 2012. The prevalence of schistosomiasis in Anhui Province decreased appreciably after 2012, whereas the natural environment did not show significant changes. The factors influencing the prevalence of schistosomiasis are complex and varied, and it was difficult to simulate the spatiotemporal trend of schistosomiasis using natural and social factors alone. This, surely influenced the GTWR interpolation results negatively after 2012. The STKriging interpolation consisted of the spatiotemporal trend and residuals, and the interpolation results of this model agreed with those of GTWR, indicating that the interpolation accuracy of STKriging was primarily determined by the fitting accuracy of the spatiotemporal trend.

The parameters of the spatiotemporal covariance of the schistosomiasis infection rate characterized the spatial and temporal variations of the disease. Larger values of spatiotemporal covariance emerged within the spatial range of 0.20° (ca. 20 km). This revealed that there was spatial autocorrelation within 20 km and negligible spatial correlation beyond 20 km in Anhui Province, suggesting that schistosomiasis transmission was within districts and counties. In addition, the influence of the time scale was as long as 10 years, and the temporal correlation was negligible beyond 10 years, suggesting that the prevalence of schistosomiasis in endemic areas at the current low epidemic level could be autocorrelations and the correlation could be up to 10 years. This result essentially agrees with our previous findings [13].

Although the spatiotemporal distribution of schistosomiasis predicted by BME, STKriging, and GTWR in Anhui Province agrees with the observed spatiotemporal distribution of schistosomiasis, a discrepancy between the predicted and observed values of schistosomiasis remained (Fig. 2). The spread of schistosomiasis usually depends on factors that cause non-linear and rapid changes in schistosomiasis, whereas such changes are often not characterized by geostatistical models (e.g., GTWR), thus affecting the prediction accuracy of BME and STKriging. Figure 2 shows the highest prevalence and lowest interpolation accuracy in 2005. Possible reasons for this are the following

two developments: (1) as the government initiated the World Bank Loan Project (1992–2001) for schistosomiasis control in Anhui province, the number of Oncomelania snails, showed a declining trend between 2001 and 2004. As the 10-year project was replaced in 2005 by an integrated control strategy more focused on infectious source control, some increase of Oncomelania snails may have started [2, 40]; and (2) the Yangtze River Basin had a warm winter and early spring in 2004 and 2005 followed by high humidity in the spring and summer leading to increased snail breeding and reproduction of Oncomelania snails [41]. These two changes were difficult to characterize using the GTWR model, reducing the prediction accuracy and further affecting the prediction accuracy of BME and STKriging. We have previously studied the impact of the schistosomiasis control project in Anhui province [41, 42], and our future priority will be to study the impact of nonlinear and rapidly changing factors on the risk of schistosomiasis.

Conclusion

This study suggests that BME exhibited the highest interpolation accuracy among the mainstream spatiotemporal interpolation methods, which could enhance the risk prediction model of infectious diseases thereby providing scientific support for government decision making. The concluding findings were the following:

(1) Urbanization together with five factors characterizing the environment in the rural areas influenced the prevalence of schistosomiasis at the county level in Anhui Province. The goodness of fit between these influencing factors and the schistosomiasis prevalence using GTWR was $R^2 = 0.76$, which means that the influencing factors could explain 76% of the spatiotemporal distribution of schistosomiasis.

(2) The prediction accuracy of BME was better what STKriging and GTWR could provide. Moreover, the predicted values all agreed with the actual spatiotemporal distribution trend of schistosomiasis.

(3) Schistosomiasis outbreaks occurred in more than 20 counties per year in Anhui Province between 2000 and 2012, with the spatial range gradually decreasing after 2012. Schistosomiasis spread up to 20 km from the original areas and affected the province for up to 10 years, which may be related to the low level of disease intensity and testing results.

Wang *et al. BMC Infectious Diseases*   (2021) 21:1171

Page 9 of 12

Appendix

Bayesian maximum entropy, BME

BME, initially proposed as a theoretical framework by Christakos in 1990, proves to be a method of nonlinear interpolation with higher precision. BME, used for spatiotemporal heterogeneity research, is designed to integrate data of various precision and quality which are called the Knowledge Base, KB. KB includes General Knowledge Base, G-KB and Specific Knowledge Base, S-KB. G-KB is used to describe the integral character of S/TRF, which can be signified as the statistical law such as expectation, variance as well as covariance. The S-KB includes hard data and soft data. Hard data are those with negligible errors, such as measured point data, while soft data are relatively imprecise data. The processes of BME are as follows:

(1) Calculating prior probability density function (PDF). According to the maximization principle of expected value of information (entropy), priori PDF with the distribution of unobserved variables can be calculated in the research area by utilizing G-KB.

$\chi_{map} = (\chi_{hard}, \chi_{soft}, \chi_k)$ is spatio-temporal random variable, and its priori PDF is represented by $f_G(\chi_{map})$. Therefore, based on the principle of information entropy, the expectation of maximum entropy is

$$E\big[Info(\chi_{map})\big] = - \int \ln\big[f_G(\chi_{map})\big] f_G(\chi_{map}) d\chi_{map} \tag{3}$$

where $\chi_{hard}, \chi_{soft}, \chi_k$ respectively indicate hard data (annual measured morbidity of Schistosomiasis in 29 epidemic counties in Anhui Province from 2000 to 2015), soft data (morbidity of Schistosomiasis measured annually in 29 epidemic counties in Anhui Province from 2000 to 2015 according to calculation of GTWR model) and morbidity in the point to be assessed. The spatial distribution is shown in Fig. 1.

To derive the maximum amount of information from G-KB, the expectation of entropy must reach the maximum, which means Format (3) needs to reach the maximum under the constraint of

$$E[g_\alpha] = \int g_\alpha(\chi_{map}) f_G(\chi_{map}) d\chi_{map}, \quad \alpha = (1, 2, \ldots, N_C) \tag{4}$$

where N_C is total number of constraints. g_α generally represents normalized constraints, expectation (first-order moment) constraints, variance (second-order moment) constraints or covariance (second-order mixed moment) constraints. In this paper, g_α indicates covariance of morbidity in (s, t) and $(s + h_s, t + h_t)$ in Anhui Province [31–33]:

$$g_\alpha = c_\varepsilon(h_s, h_t) = c_1 \exp\left(-\frac{3h_s}{a_{h_s 1}}\right)\left(1 - \frac{3h_t}{2a_{h_t 1}} + \frac{h_t^3}{2a_{h_t 1}^3}\right)$$
$$+ c_2 \exp\left(-\frac{3h_t}{a_{h_t 1}}\right)\left(1 - \frac{3h_s}{2a_{h_s 2}} + \frac{h_s^3}{2a_{h_s 2}^3}\right) \tag{5}$$

where c_1 and c_2 are sill coefficients; $a_{h_s i}$ and $a_{h_t i} (i = 1, 2)$ represent spatial and temporal range of spatio-temporal covariance, respectively.

According to Format (3) and Constraint (4), the maximum of priori PDF is

$$f_G(\chi_{map}) = \frac{1}{A} \exp\left(\sum_{\alpha=1}^{N_C} \mu_\alpha g_\alpha(\chi_{map})\right) \tag{6}$$

where $A = \int \exp\left(\sum_{\alpha=1}^{N_C} \mu_\alpha g_\alpha(\chi_{map})\right) d\chi_{map}$ and μ_α is lagrange multiplier, μ_α is given by

$$E[g_\alpha] = \frac{1}{A} \int g_\alpha(\chi_{map}) \exp\left[\sum_{\alpha=1}^{N_C} \mu_\alpha g_\alpha(\chi_{map})\right] d\chi_{map} \tag{7}$$

(2) Calculating posterior PDF. Based on Bayesian conditional probability, the posterior PDF of Schistosoma at unobserved point χ_k in Anhui Province can be derived from the posterior PDF obtained by using G-KB in the first update phase [43, 44], as is given by

$$f_K(\chi_k) = f_G(\chi_k | \chi_{data}) = \frac{f_G(\chi_k, \chi_{data})}{f_G(\chi_{data})} \tag{8}$$

where, $\chi_{data} = [\chi_{hard}, \chi_{soft}]$.

Geographical and temporal weighted regression, GTWR

The geographically weighted regression model is a classic model for spatial heterogeneity research while GTWR is an in-depth study of geographically weighted regression model. Based on the idea of local regression, the GTWR method, considering spatial temporal heterogeneity of data, embeds the temporal data, a new dimension, into regression parameters to simultaneously measure changes of data in space and time. The processes are as follows: Firstly, according to the adjustable bandwidth criterion, observation points are determined that affect the regression point. Secondly, the weight matrix is calculated by the spatial-temporal distance between observation points and the regression point and the weight function. Finally, the regression coefficient value is calculated by the weighted least square method. The GTWR model is expressed as [15, 17, 18, 45]:

Wang *et al. BMC Infectious Diseases*   (2021) 21:1171

Page 10 of 12

$$y(u_i, v_i, t_i) = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^{d} \beta_k(u_i, v_i, t_i)x_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{9}$$

where (u_i, v_i, t_i) is the space-time coordinate of the i-th sample point; u_i, v_i and t_i are longitude coordinates, latitude coordinates and time coordinates of the i-th sample point, respectively. $y(u_i, v_i, t_i)$ is the dependent variable of the i-th sample point, which is morbidity of Schistosoma in Anhui Province in this paper. $(x_{i1}, x_{i2}, \ldots, x_{id})$, $i = 1, 2, \ldots, n$ is an independent variable, representing significant influencing factors. $\beta_0(u_i, v_i, t_i)$ is the constant term of the i-th sample point. $\beta_k(u_i, v_i, t_i)$ is the regression coefficient of the k-th independent variable at the i-th sample point. ε_i is an independent and identically distributed error term, which is usually assumed to obey $N(0, \sigma^2)$ for distribution.

Spatiotemporal Kriging, STKriging

On the ground of Kriging model (ordinary kriging model is widely applied), STKriging [7] model replaces the two-dimensional spatial variogram with the three-dimensional spatio-temporal variogram to objectively describe the proximity of environmental variables in spatial-temporal domain. STKriging theory [13, 46, 47] primarily consists of spatial-temporal variable, spatial-temporal data stationarity, spatial-temporal variogram and spatial-temporal interpolation.

Providing that $Z(s, t)$ indicates morbidity of Schistosoma in county s at time t (unit/ year), STKriging model assumes that the spatio-temporal process[3, 13] of schistosomiasis morbidity is composed of spatio-temporal trend $m(s, t)$ and stochastic residual $\varepsilon(s, t)$:

$$Z(s, t) = m(s, t) + \varepsilon(s, t) \tag{10}$$

where $m(s, t)$ is determined by the result $y(u, v, t)$ of GTWR model ((u, v) is the coordinates of the county geographical center).

The fitting residue covariance of morbidity of Schistosoma analyzed by GTWR at (s, t) and $(s + h_s, t + h_t)$ in Anhui Province merely depends on (h_s, h_t), where h_s is the Euclidean space distance and h_t is temporal distance. The spatio-temporal covariance of $\varepsilon(s, t)$ is

$$C(h_s, h_t) = Cov(\varepsilon(s + h_s, t + h_t), \varepsilon(s, t)) \tag{11}$$

The variogram is

$$\gamma(h_s, h_t) = \frac{1}{2}E\left\{(\varepsilon(s + h_s, t + h_t) - \varepsilon(s, t))^2\right\} \tag{12}$$

The variogram adopted in the paper is nonseparable spatio-temporal Cressie-Huang model [46, 48]:

$$\gamma(h_s, h_t) = \begin{cases} 0 & h_t = h_s = 0 \\ \sigma^2\left(1 - \dfrac{(a|h_t| + 1)}{\left((a|h_t| + 1)^2 + b^2\|h_s\|^2\right)^{1.5}}\right) \\ \quad + C_0 + a_1\|h_s\|^{a_2} & \text{other} \end{cases} \tag{13}$$

In the spatio-temporal variability of fitting residual $\varepsilon(s, t)$, a is temporal scale parameter, b is spatial scale parameter, C_0 is nugget effect, $a_1\|h_s\|^{a_2}$ is pure spatial variability, a_1, a_2 is spatial smoothing parameter, and σ^2 is the variance of $\varepsilon(s, t)$.

Declarations

Wang *et al. BMC Infectious Diseases*   (2021) 21:1171

Page 11 of 12

Author details
[1]College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China. [2]School of Public Health, Fudan University, Shanghai 200032, China. [3]Anhui Institute of Schisomiasis Control and Research, Hefei 230061, China. [4]Chinese Academy of Surveying and Mapping, Beijing 100036, China. [5]Geospatial Health, Ingerod, Brastad, Sweden.

References
1. World Health Organisation's Factsheet on schistosomiasis https://www.who.int/news-room/fact-sheets/detail/schistosomiasis accessed 2 Dec 2020
2. Yang JR, Xu MX, Tan XD. Healthy China strategy and schistosomiasis control. Chin J Schistosomiasis Control. 2020;32(04):419–22.
3. Hu Y, Gao J, Chi M, Luo C, Lynn H, Sun L, Tao B, Wang D, Zhang Z, Jiang Q. Spatio-temporal patterns of *Schistosomiasis japonica* in lake and marshland areas in China: the effect of snail habitats. Am J Trop Med Hyg. 2014;91(3):547–54.
4. Zhou DR, Li YS, Yang XM. Schistosomiasis control in China. World Health Forum. 1994;90(4):387–9.
5. Zhou XN, Wang LY, Chen MG, Wu XH, Utzinger J. The public health significance and control of schistosomiasis in China - then and now. Acta Trop. 2005;96(2–3):97–105.
6. Gray DJ, Williams GM, Li Y, McManus DP. Transmission dynamics of *Schistosoma japonicum* in the Lakes and Marshlands of China. PLoS ONE. 2009;3(12):e4058.
7. Ye JT, Ji SM, Yang Y. Spatio-temporal geotatistics method research and progress. Geomatics Spatial Inf Technol. 2014;37(01):38–43.
8. Gao FH, Zhang SQ, Wang TP, Yu BB, He JC, Zhang GH, Wang H. Spatial analysis of distribution of schistosomiasis in Anhui Province. Chin J Schistosomiasis Control. 2011;23(02):125–7.
9. Xin X, Gu HZ, Zhao B, Hao LP, Zhu WP. Epidemiological characteristics of clustering outbreak of hand foot and mouth disease in kindergarte in Pudong New Area of Shanghai. Chin J School Health. 2018;39(07):1057–9.
10. Xu BY, Chen BW, Ni ZZ, Li DY. Analysis on variation of urinary iodine of children by spatial autoregressive model. J Hygiene Res. 2004;05:578–80.
11. Zhang XF, You AG, Pan JJ, Cui FZ, Wu SX, Sun CQ. Application of autoregressive integrated moving average model to predicting the incidence of hand, foot and mouth disease in Sanmenxia city. Practical Prev Med. 2020;27(02):168–70.
12. Xu K, Dang SQ, Dong JY, Wang YM, Li SX. Association between incidence of hand foot and mouth disease and meteorological factors in Jiayuguan. Chin J Public Health Manag. 2020;36(02):214–6.
13. Hu Y, Li R, Bergquist R, Lynn H, Gao F, Wang Q, Zhang S, Sun L, Zhang Z, Jiang Q. Spatio-temporal transmission and environmental determinants of *Schistosomiasis japonica* in Anhui Province, China. PLoS Negl Trop Dis. 2015;9(2):e0003470.
14. Gething PW, Atkinson PM, Noor AM, Gikandi PW, Hay SI, Nixon MS. A local space–time kriging approach applied to a national outpatient malaria data set. Comput Geosci. 2007;33(10):1337–50.
15. Ge L, Zhao Y, Sheng Z, Wang N, Zhou K, Mu X, Guo L, Wang T, Yang Z, Huo X. Construction of a seasonal difference-geographically and temporally weighted regression (SD-GTWR) model and comparative analysis with GWR-based models for hemorrhagic fever with renal syndrome (HFRS) in Hubei Province (China). Int J Environ Res Public Health. 2016;13(11):1.
16. Ge L. The application of hemorrhagic fever with renal syndrome (HFRS) Analysis based on seasonal difference-geographically and temporally weighted regression (SD-GTWR). Urban Geotechn Investig Surv. 2017;5:34–8.
17. Ge L: The application of spatial-temporal analysis and modeling methods on Hemorrhagic Fever with Renal Syndrome. Ph.D. dissertation. Wuhan University; 2017.
18. Hao H: Spatio-temporal data analysis model and its application in the prediction of hand, foot and mouth disease. M.S. thesis. Inner Mongolia University of Technology; 2018.
19. Xiao Y, He ZY, Miao J, Pan F, Yang H. Modelling the spatial distribution of epidemic by search engine data. Bull Surv Mapping. 2018;2:94–8.
20. Sun Y. The influence of urbanization on schistosomiasis based on night light data. Shangdong: Shangdong University of Science and Tectnology; 2020.
21. He J, Kolovos A. Bayesian maximum entropy approach and its applications: a review. Stoch Env Res Risk Assess. 2018;32(4):859–77.
22. Cao C, Chen W, Zheng S, Zhao J, Wang J, Cao W. Analysis of spatiotemporal characteristics of pandemic SARS spread in mainland China. Biomed Res Int. 2016. https://doi.org/10.1155/2016/7247983.
23. Cao C, Xu M, Chang C, Xue Y, Zhong S, Fang L, Cao W, Zhang H, Gao M, He Q, et al. Risk analysis for the highly pathogenic avian influenza in Mainland China using meta-modeling. Chin Sci Bull. 2010;55(36):4168–78.
24. Christakos G, Olea RA, Yu HL. Recent results on the spatiotemporal modelling and comparative analysis of Black Death and bubonic plague epidemics. Public Health. 2007;121(9):700–20.
25. Gesink Law DC, Bernstein KT, Serre ML, Schumacher CM, Leone PA, Zenilman JM, Miller WC, Rompalo AM. Modeling a syphilis outbreak through space and time using the Bayesian maximum entropy approach. Ann Epidemiol. 2006;16(11):797–804.
26. Hampton KH, Serre ML, Gesink DC, Pilcher CD, Miller WC. Adjusting for sampling variability in sparse data: geostatistical approaches to disease mapping. Int J Health Geogr. 2011;10(1):54–54.
27. He J, Christakos G, Wu J, Jankowski P, Langousis A, Wang Y, Yin W, Zhang W. Probabilistic logic analysis of the highly heterogeneous spatiotemporal HFRS incidence distribution in Heilongjiang province (China) during 2005–2013. PLoS Negl Trop Dis. 2019;13(1):7091.
28. Lee S-J, Yeatts KB, Serre ML. A Bayesian Maximum Entropy approach to address the change of support problem in the spatial analysis of childhood asthma prevalence across North Carolina. Spatial Spatio-Temporal Epidemiol. 2009;1(1):49–60.
29. Wang J-F, Guo Y-S, Christakos G, Yang W-Z, Liao Y-L, Li Z-J, Li X-Z, Lai S-J, Chen H-Y. Hand, foot and mouth disease: spatiotemporal transmission and climate. Int J Health Geogr. 2011;10:25–25.
30. Yu HL, Angulo JM, Cheng MH, Wu J, Christakos G. An online spatiotemporal prediction model for dengue fever epidemic in Kaohsiung (Taiwan). Biometr J Biometrische Zeitschrift. 2014;56(3):428–40.
31. Yu H-L, Chiang C-T, Lin S-D, Chang T-K. Spatiotemporal analysis and mapping of oral cancer risk in Changhua County (Taiwan): an application of generalized Bayesian maximum entropy method. Ann Epidemiol. 2010;20(2):99–107.
32. Yu H-L, Lee C-H, Chien L-C. A spatiotemporal dengue fever early warning model accounting for nonlinear associations with hydrological factors: a Bayesian maximum entropy approach. Stoch Env Res Risk Assess. 2016;30(8):2127–41.
33. Yu HL, Yang SJ, Yen HJ, Christakos G. A spatio-temporal climate-based model of early dengue fever warning in southern Taiwan. Stoch Env Res Risk Assess. 2011;25(4):485–94.
34. Zhang CT: Research on key issues of bayesian maximum entropy spatiotemporal prediction and its application. Ph.D. dissertation. Huazhong Agricultural University; 2016.
35. Yu JM, Vlas SJD, Jiang QW, Gryseels B. Comparison of the Kato-Katz technique, hatching test and indirect hemagglutination assay (IHA) for the diagnosis of *Schistosoma japonicum* infection in China. Parasitol Int. 2007;56(1):45–9.
36. Sun Y, Liu X, Su YC, Xu S, Ji B, Zhang ZJ. County urbanization level estimated from nighttime light data in Anhui province. J Geo-Inf Sci. 2020;22(09):1837–47.
37. Pebesma E, Heuvelink G. Spatio-temporal interpolation using gstat. RFID J. 2016;8(1):204–18.
38. Wibrin MA, Bogaert P, Fasbender D. Combining categorical and continuous spatial information within the Bayesian maximum entropy paradigm. Stoch Env Res Risk Assess. 2006;20(6):423–33.

Wang *et al. BMC Infectious Diseases*   (2021) 21:1171

Page 12 of 12

39. Choi K-M, Yu H-L, Wilson ML. Spatiotemporal statistical analysis of influenza mortality risk in the State of California during the period 1997–2001. Stoch Env Res Risk Assess. 2008;22(1):15–25.

40. Wang HY, Zhang ZJ, Peng WX, Zhou YB, Zhao GM, Chen GX, Cui DY, Jiang QW. Analysis of endemic situation of schistosomiasis in Guichi District of Chizhou City, Anhui Province from 2000 to 2006. Chin J Schistosomiasis Control. 2008;02:89–92.

41. Luo JP, Yang WP, Gao FH, Wang TP, Zhang SQ, Zhang C. Snail situation in surveillance sites for Sehistosomiasis in Anhui Porvince from 2005 to 2008. J Trop Dis Parasitol. 2009;7(04):206–9.

42. Hu Y, Li S, Xia C, Chen Y, Zhang Z. Assessment of the national schistosomiasis control program in a typical region along the Yangtze River, China. Int J Parasitol. 2016;47(1):21.

43. Yang Y, Zhang RX. Review on Bayesian maximum entropy geostatistics method. Soils. 2014;46(03):402–6.

44. Zhang B, Li WD, Yang Y, Wang SQ, Cai CF. The bayesian maximum entropy geostatistical approach and its application in soil and environmental sciences. Acta Pedol Sin. 2011;48(04):831–9.

45. Sun SM, Li ZM, Zhang HG, Hu XJ. Temporal-spatial characteristic analysis of AIDS/HIV epidemic during 2011–2016 in China. Chin J Dis Control Prev. 2018;22(12):1207–10.

46. Mei Y: Key technology and its application for spatio-temporal kriging. M.S. thesis. Huazhong Agricultural University; 2016.

47. Iglesias I, Montes F, Martínez M, Perez A, Gogin A, Kolbasov D, de la Torre A. Spatio-temporal kriging analysis to identify the role of wild boar in the spread of African swine fever in the Russian Federation. Spatial Stat. 2018;28:226–35.

48. Noel C, Hsin-Cheng H. Classes of nonseparable, spatio-temporal stationary covariance functions. J Am Stat Assoc. 1999;94:1330–9.

Publisher's Note