Contents lists available at ScienceDirect



International Journal of Infectious Diseases



INTERNATIONAL SOCIETY FOR INFECTIOUS DISEASES

journal homepage: www.elsevier.com/locate/ijid

## Spatial analysis of dengue fever and exploration of its environmental and socio-economic risk factors using ordinary least squares: A case study in five districts of Guangzhou City, China, 2014



Yujuan Yue<sup>a</sup>, Jimin Sun<sup>a</sup>, Xiaobo Liu<sup>a</sup>, Dongsheng Ren<sup>a</sup>, Qiyong Liu<sup>a</sup>, Xiangming Xiao<sup>b</sup>, Liang Lu<sup>a,\*</sup>

<sup>a</sup> State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, People's Republic of China <sup>b</sup> Department of Microbiology and Plant Biology, University of Oklahoma, OK, USA

#### ARTICLE INFO

Article history: Received 22 May 2018 Received in revised form 24 July 2018 Accepted 27 July 2018 Corresponding Editor: Eskild Petersen, Aarhus, Denmark

Keywords: Dengue fever Environmental and socio-economic factors Spatial pattern analysis Spatial statistics analysis Spearman rank correlation OLS Guangzhou

#### ABSTRACT

*Objective:* Spatial patterns and environmental and socio-economic risk factors of dengue fever have been studied widely on a coarse scale; however, there are few such quantitative studies on a fine scale. There is a need to investigate these factors on a fine scale for dengue fever.

*Methods*: In this study, a dataset of dengue fever cases and environmental and socio-economic factors was constructed at 1-km spatial resolution, in particular 'land types' (LT), obtained from the first high resolution remote sensing satellite launched from China (GF-1 satellite), and 'land surface temperature', obtained from moderate resolution imaging spectroradiometer (MODIS) images. Spatial analysis methods, including point density, average nearest neighbor, spatial autocorrelation, and hot spot analysis, were used to analyze spatial patterns of dengue fever. Spearman rank correlation and ordinary least squares (OLS) were used to explore associated environmental and socio-economic risk factors of dengue fever in five districts of Guangzhou City, China in 2014.

*Results*: A total of 30 553 dengue fever cases were reported in the districts of Baiyun, Haizhu, Yuexiu, Liwan, and Tianhe of Guangzhou, China in 2014. Dengue fever cases showed strong seasonal variation. The cases from August to October accounted for 96.3% of the total cases in 2014. The top three districts for dengue fever morbidity were Baiyun (1.32%), Liwan (0.62%), and Haizhu (0.60%). Strong spatial clusters of dengue fever cases were observed. Areas of high density for dengue fever were located at the district junctions. The dengue fever outbreak was significantly correlated with LT, normalized difference water index (NDWI), land surface temperature of daytime (LSTD), land surface temperature of nighttime (LSTN), population density (PD), and gross domestic product (GDP) (correlation coefficients of 0.483, 0.456, 0.612, 0.699, 0.705, and 0.205, respectively). The OLS equation was built with dengue fever cases as the dependent variable and LT, LSTN, and PD as explanatory variables. The residuals were not spatially autocorrelated. The adjusted R-squared was 0.320. *Conclusions*: The findings of spatio-temporal patterns and risk factors of dengue fever can provide scientific information for public health practitioners to formulate targeted, strategic plans and implement effective public health prevention and control measures.

© 2018 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## Introduction

Dengue fever is a systemic viral infection transmitted by mosquitoes of the *Aedes* genus (Simmons et al., 2012), and is endemic in more than 100 countries of the Southeast Asia, Americas, Western Pacific, Africa, and Eastern Mediterranean

\* Corresponding author.

E-mail address: luliang@icdc.cn (L. Lu).

regions (Guzman and Harris, 2015). A study in 2013 estimated that 390 million people had dengue virus infections with 96 million cases annually worldwide, more than three times higher than the World Health Organization 2012 estimates (Bhatt et al., 2013). Dengue fever has evolved from a sporadic disease to a major public health problem with substantial social and economic impacts because of increasing geographical extension, numbers of cases, and disease severity (Guzman and Harris, 2015).

https://doi.org/10.1016/j.ijid.2018.07.023

<sup>1201-9712/© 2018</sup> The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Dengue fever is a notifiable disease in China. From 1978 to 2008, a total of 655 324 cases and 610 deaths were reported in Mainland China. From 2009 to 2014, a total of 52749 cases and six deaths were notified (Chen and Liu, 2015). Dengue fever outbreaks have spread from Guangdong and Hainan in the southern coastal areas to the relatively northern and western areas including Fujian, Zhejiang, and Yunnan, with shorter outbreak intervals as compared to those before the 1990s (Wu et al., 2010). Guangdong has been the area most seriously affected by dengue fever in China (Liu et al., 2014), and the majority of cases have occurred in Guangzhou, the capital city of Guangdong (Wang et al., 2013a). In recent years, Guangdong has had the highest incidence of dengue fever in China (Wang et al., 2013b; Fan et al., 2014; Li et al., 2013). According to the China National Notifiable Disease Surveillance System, an extensive dengue outbreak that posed a substantial socio-economic burden hit China in 2014 (Chen and Liu, 2015), with 47 127 dengue fever cases diagnosed in the country, 45 231 dengue fever cases in Guangdong, and 37382 dengue fever cases in Guangzhou, among which 30553 cases were aggregated in the districts of Baiyun, Liwan, Yuexiu, Haizhu, and Tianhe. The latter four districts belong to old districts of Guangzhou.

A better understanding of dengue fever outbreaks, especially spatial patterns, would help in the planning of resource allocation for dengue fever prevention and control (Lai et al., 2015). Most research on the spatio-temporal analysis of dengue fever has been based on a coarse scale (Bhatt et al., 2013; Chen and Liu, 2015; Lai et al., 2015; Hashizume et al., 2012; Corner et al., 2013; Dewan et al., 2017; Lippi et al., 2018; Castro et al., 2018; Shearer et al., 2018), such as the census district (Corner et al., 2013; Dewan et al., 2017), block (Lippi et al., 2018), or municipality (Castro et al., 2018). Most studies on dengue fever in China have been based on an administrative scale, such as the province, city, or district (Liu et al., 2014; Wang et al., 2013a,b; Fan et al., 2014; Li et al., 2013), and only a few studies on the spatio-temporal analysis of dengue fever have been based on the administrative scale of a town with an area in the dozens of square kilometers (Qi et al., 2015). However, dengue fever field monitoring performed by the present authors has shown that adjacent blocks may have significantly different dengue fever outbreaks. In other words, dengue fever outbreaks are similar on a coarse scale, but the spatio-temporal patterns on a fine scale may clearly be different. In order to evaluate the hotspot areas exactly, dengue fever analysis should be explored at as fine a spatial resolution as possible.

Dengue fever outbreaks are known to be strongly influenced by imported cases (Sang et al., 2014, 2015), mosquito density (Sang et al., 2014, 2015; Lai, 2011), meteorological factors (Wang et al., 2013a) (such as air temperature (Sang et al., 2014, 2015; Eastin et al., 2014; Xu et al., 2016; Goto et al., 2013), rainfall (Sang et al., 2014, 2015; Xu et al., 2016; Goto et al., 2013; Castro et al., 2018), relative humidity (Sang et al., 2014), vapor pressure (Sang et al., 2014), air pressure (Sang et al., 2014), and sea surface temperature (Lai, 2011; Laureano-Rosario et al., 2017)), socio-economic factors (Qi et al., 2015; Hagenlocher et al., 2013; Wu et al., 2009), and environmental factors (such as water (Fullerton et al., 2014; Tian et al., 2016), vegetation (Qi et al., 2015), river levels (Hashizume et al., 2012), access to paved roads, and housing conditions (Lippi et al., 2018)). Moreover, in dengue fever field monitoring, the present authors have found that dengue fever is closely related to environmental and socio-economic conditions, such as sanitation status, population density, ventilation conditions, etc.

Meteorological data rather than remote sensing data have been used in most dengue fever research in the past (Fan et al., 2014; Sang et al., 2014, 2015). For an area of 9 600 000 km<sup>2</sup>, there are fewer than 800 meteorological stations in China. Furthermore, air temperature rather than land surface temperature has been used in such research (Sang et al., 2014, 2015). Land surface temperature

derived from remote sensing images at a moderate spatial resolution has a smaller spatial scale and shows the true environmental condition more directly. Some researchers have studied risk factors of dengue fever using remote sensing data on the coarse scale of a city or neighborhood (Laureano-Rosario et al., 2017; Tian et al., 2016; Khormi and Kumar, 2011), such as sea surface temperature (Laureano-Rosario et al., 2017) and surface water areas (Tian et al., 2016). Others have studied dengue fever on a neighborhood scale but with the existing field investigation data (Delmelle et al., 2016). However, the use of remote sensing images on a fine scale has been scarce in previous studies. Such investigations are lacking, particularly in China. China now has its first high spatial resolution satellite – the GF-1 satellite. This has provided the opportunity to use China's own high resolution satellite data for application in disease prevention and control.

In this study, a dataset of dengue fever cases and environmental and socio-economic factors was constructed at 1-km spatial resolution for five districts of Guangzhou City, China in 2014. Spatial analysis methods including point density, average nearest neighbor, spatial autocorrelation, and hot spot analysis were adopted to analyze spatial patterns of dengue fever, and Spearman rank correlation and ordinary least squares were used to confirm environmental and socio-economic risk factors of dengue fever, in particular land types from GF-1 remote sensing images and land surface temperature from moderate resolution imaging spectroradiometer (MODIS) images.

## Materials and methods

#### Study area

Guangzhou is the capital city of Guangdong Province. It is the largest coastal city in southern China, with an area of  $7000 \text{ km}^2$  and about 13.5 million permanent residents. Guangzhou is located in the subtropical coastal area with an oceanic subtropical monsoon climate. In Guangzhou, the summer is long and the winter is short. The average daily temperature in Guangzhou is 16 °C in January and 28.7 °C in July, and the average annual precipitation ranges from 1600 mm to 1900 mm.

A dengue fever outbreak occurred in the five districts of Baiyun, Liwan, Yuexiu, Haizhu, and Tianhe in Guangzhou, China in 2014 (Figure 1).

#### Dengue fever case data

Data on dengue fever cases from January 1, 2014 to December 31, 2014 were collected from the China Information System for Disease Control and Prevention. Each dengue fever case was confirmed through clinical diagnosis or laboratory diagnosis, and details including the residential address, date of illness onset, etc., were available for each case. With geocoding technology, the address of each dengue fever case was transformed to a particular spatial location with latitude and longitude. A total of 30 553 cases from five districts of Guangzhou in 2014 were used in the analysis. The distribution of these cases is shown in Figure 2, in which the grey grid has a spatial resolution of 1 km. The study area comprised 758 grids. The numbers of dengue fever cases in these grids were obtained using the Data Management Tools and Spatial Statistics Tools of ArcMap 10.1 software. Environmental factor values in these grids were also obtained in this way.

## Environmental and socio-economic factors

The environmental factors of land types (LT), normalized difference water index (NDWI), land surface temperature of



Figure 1. Geographic locations of the study areas.

daytime (LSTD), and land surface temperature of nighttime (LSTN), and the socio-economic factors of population density (PD) and gross domestic product (GDP), were considered in this study (Figure 3).

The GF-1 satellite, made in China, was launched on April 26, 2013. Its multi-band image resolution is 16 m. At least two scenes of GF-1 images obtained on August 7, 2015 were chosen. The remote sensing images taken by the GF-1 satellite were downloaded from the China Center for Resources Satellite Data and Application (http://www.cresda.com/CN/sjfw/zxsj/index.shtml). LT and NDWI used in this study were from image interpretation of GF-1 multi-band remote sensing images. LT including water, vegetation, and buildings was interpreted according to the object-oriented classification approach in eCognition software, with a classification accuracy of 0.91. NDWI was interpreted according to band math in ENVI 5.2 software.

LSTD and LSTN at 1-km spatial resolution used in this study were the mean values of MODIS MOD11A2 daily reprocessing products from July 1 to September 30, 2014. These were downloaded from NASA MODIS LAADS DAAC (https://ladsweb.modaps. eosdis.nasa.gov/); the unit of LSTD and LSTN was 0.1 K.

PD and GDP raster data at 1-km spatial resolution were provided by the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (http://www. igsnrr.ac.cn/) and were interpolated from the census data of PD and GDP in 2010.

Finally, all data were obtained at 1-km spatial resolution through image resampling and geometric correction, which coincided with the grids at 1-km spatial resolution.

#### Spatial pattern analysis

Spatial analysis methods of point density (Silverman, 1986), average nearest neighbor (Mitchell, 2005), spatial autocorrelation (global Moran's *I*) (Mitchell, 2005), and hot spot analysis (Scott and Warmerdam, 2005) were adopted to assess dengue fever spatial patterns such as spatial distribution characteristics, spatial clustering, and spatial hot spots in the five districts of Guangzhou. The spatial analysis methods have been applied successfully in disease analysis. Spatial autocorrelation and hot spot analysis have usually been used in dengue fever spatial analysis on a coarse scale (Liu et al., 2014; Wang et al., 2013b; Fan et al., 2014). However these methods were used in dengue fever spatial analysis on a fine scale in this study, such as the point scale and 1-km spatial resolution grid. Point density and average nearest neighbor were applied on the point scale. Spatial autocorrelation and hot spot analysis were applied on the 1-km spatial resolution grid.

'Point density' calculates the density of point features around each output raster cell. Conceptually, a neighborhood is defined around each raster cell center, and the number of points that fall within the neighborhood is totaled and divided by the area of the neighborhood.

'Average nearest neighbor' measures the distance between each feature centroid and its nearest neighbor's centroid location. It then averages all of these nearest neighbor distances. If the average distance is less than the average for a hypothetical random distribution, the distribution of the features being analyzed is considered to be clustered. If the average distance is greater than a hypothetical random distribution, the features are considered to be



Figure 2. Geographic locations of dengue fever cases in the study areas, 2014.

dispersed. If the average nearest neighbor ratio is less than 1, the pattern exhibits clustering. If the index is greater than 1, the trend is towards dispersion.

'Spatial autocorrelation' (global Moran's *I*) measures spatial autocorrelation based on both feature locations and feature values simultaneously. The global Moran's *I* index values fall between -1.0 and +1.0. This index evaluates whether the pattern expressed is clustered (>0), dispersed (=0), or random (<0).

The 'hot spot analysis' (Getis-Ord Gi<sup>\*</sup>) returned for each feature in the dataset is a *z*-score. For statistically significant positive *z*-scores, the larger the *z*-score is, the more intense the clustering of high values (hot spot). For statistically significant negative *z*-scores, the smaller the *z*-score is, the more intense the clustering of low values (cold spot).

#### Spatial statistical models

Spearman rank correlation (Lehman, 2005) was used to explore the relationships between dengue fever and risk factors. Ordinary least squares (OLS) (Mitchell, 2005) was adopted to establish an equation between dengue fever and several risk factors in this research.

Spearman rank correlation assesses how well the relationship between two variables can be described using a monotonic function.

OLS is a method for estimating the unknown parameters in a linear regression model, with the goal of minimizing the sum of the

squares of the differences between the observed responses (values of the variable being predicted) in the given dataset and those predicted by a linear function of a set of explanatory variables. Visually this is seen as the sum of the squared vertical distances between each data point in the set and the corresponding point on the regression line: the smaller the differences are, the better the model fits the data. OLS is the best known of all regression techniques. It is also the proper starting point for all spatial regression analysis. It provides a global view of the variable or process one is trying to understand or predict; it creates a single regression equation to represent that process. The OLS regression equation is the mathematical formula applied to the explanatory variables to best predict the dependent variable. In the regression equation, the dependent variable is always Y and the explanatory variables are always Xs. Each explanatory variable is associated with a regression coefficient describing the strength and the sign of that variable's relationship to the dependent variable. A regression equation might appear as follows (ESRI, 2017a):  $Y = \beta_0 + \beta_1 X_1 + \beta_2 + \beta_1 X_1 + \beta_2 + \beta_2 + \beta_2 + \beta_1 X_2 + \beta_2 +$  $\beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$ , where Y is the dependent variable,  $X_n$  is the explanatory variable,  $\beta_n$  is the coefficient, and  $\varepsilon$  is the random error residual.

Trust in the model can be assessed according to six rules: (1) the coefficients have the expected signs; (2) there is no redundancy among explanatory variables; (3) the coefficients are statistically significant; (4) the residuals are normally distributed; (5) there is a strong adjusted R-square value; (6) the residuals are not spatially



Figure 3. Environmental factors. A:LT; B:NDWI; C:LSTD; D:LSTN; E:PD; F:GDP.



Figure 4. The monthly dengue fever occurrences in the study areas from July to November, 2014.

Table 1Dengue fever cases in the study area, 2014.

	Baiyun	Yuexiu	Liwan	Tianhe	Haizhu
Number of dengue fever cases	11 843	4806	4469	3436	5999
Morbidity of dengue fever	1.32%	0.41%	0.62%	0.42%	0.60%

autocorrelated (ESRI, 2017b). Collinearity was accounted for among the variables utilized according to the six rules of the OLS model.

#### Results

### Monthly occurrence of dengue fever

The dengue fever outbreaks in the five districts of Guangzhou in 2014 are summarized in Figure 4 and Table 1. The monthly dengue fever cases peaked in October with 15 034 cases; other peaks occurred in September with 12 908 cases and August with 1473 cases. There were few dengue fever cases from January to May. Dengue fever cases were concentrated in the districts of Baiyun, Haizhu, and Yuexiu, accounting for 74.1% of the total cases in the study area. The yearly dengue fever cases reached the highest number in Baiyun District with 11843 cases (38.8%), followed by Haizhu District with 5999 cases and Yuexiu District with 4806 cases (Table 1 and Figure 4). The top three districts for dengue fever morbidity were Baiyun (1.32%), Liwan (0.62%), and Haizhu (0.60%).

#### Spatial analysis of dengue fever

As shown in Figure 5, the high density areas for dengue fever cases (between 370 and 617) at 1-km spatial resolution, which are shown in red and orange, were located at the junctions between Baiyun District, Yuexiu District, Liwan District, and Haizhu District.

The average nearest neighbor result indicated that dengue fever cases in the study area were in clustered patterns, with a nearest neighbor ratio of 0.17, *z*-score of -277.33, and *p*-value of less than 0.01.

The spatial autocorrelation (global Moran's I) result again showed that dengue fever cases in the study area were in clustered patterns, with a Moran's index of 0.59 and p-value of less than 0.01.

The hot spot analysis (Getis-Ord Gi<sup>\*</sup>) results are given in Figure 6. The grids in selected ones were with statistically significantly positive *z*-scores. Most of these grids showed the most intense clustering of high values, which were located in Baiyun District, Yuexiu District, Liwan District, and Haizhu District.

# Identifying environmental and socio-economic risk factors of dengue fever

Grid data with a PD value of zero or any other risk factor value of zero were removed. Finally, data for 683 grids were considered in this section.

The correlations between dengue fever cases and risk factors such as LT, NDWI, LSTD, LSTDN, PD, and GDP are summarized in



Figure 5. Point density analysis of dengue fever in the study areas, 2014.



Figure 6. Hot spot analysis for dengue fever cases in the study areas, 2014.

Table 2. The results showed that the dengue fever outbreak was significantly positively correlated with LT, NDWI, LSTD, LSTN, PD, and GDP, with correlation coefficients of 0.483, 0.456, 0.612, 0.699, 0.705, and 0.205, respectively.

Risk factors were first inputted into the model. Then, according to the six rules of OLS model availability, the risk factors were removed from the model step by step. Finally, three risk factors were adopted in the model (Table 3). The residuals of this model were not spatially autocorrelated. The adjusted R-squared was 0.320. The model could be shown as follows:  $\hat{Y} = -5215.584 +$ 20.401 $X_1$  + 17.461 $X_2$  + 31.671 $X_3$ , where  $\hat{Y}$  was dengue fever case,  $X_1$ was LT,  $X_2$  was LSTN, and  $X_3$  was PD, which was expressed as the number per 10 000.

## Discussion

Dengue fever shows a seasonal pattern. During 2005–2014, 99.8% of indigenous dengue fever cases in continental China

#### Table 2

Relationships between dengue fever cases and environmental factors.

	LT	NDWI	LSTD	LSTN	Population	GDP
r <sub>s</sub>	0.483 <sup>**</sup>	0.456 <sup>**</sup>	0.612 <sup>**</sup>	0.699 <sup>**</sup>	0.705 <sup>**</sup>	0.205 <sup>**</sup>
Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000

LT, land types; NDWI, normalized difference water index; LSTD, land surface temperature of daytime; LSTN, land surface temperature of nighttime; GDP, gross domestic product.

Correlation significant at the 0.01 level (2-tailed).

occurred from July to November (Lai et al., 2015). A dengue outbreak hit the city of Guangzhou extensively in 2014. There were 30 553 dengue fever cases in five districts of Guangzhou City. Dengue fever cases were concentrated in August, September, and October, accounting for 96.3% of the total cases. Dengue fever cases peaked in October, accounting for 49.2% of the total cases.

Dengue fever also exhibits a spatial clustering characteristic (Liu et al., 2014; Wang et al., 2013a,b; Fan et al., 2014; Li et al., 2013). Counties around Guangzhou City and Chaoshan Region were at increasing risk of dengue fever in Guangdong Province from 2001 to 2006 (Liu et al., 2014). Dengue fever cases in Guangdong Province during the years 1978–2010 were mostly concentrated in the developed regions of the Pearl River Delta, such as Zhanjiang and Shantou (Wang et al., 2013b). Strong spatial clusters of dengue fever were distributed in the study area in 2014. High density areas of dengue fever were located at the district junctions. The most intense clustering areas for high values were located in the areas adjacent to the four districts of Baiyun, Yuexiu, Liwan, and Haizhu (Figures 5 and 6). In China, a street or town on the prefectural boundary can roughly be regarded as peri-urban, with lower levels of management from both prefectures, because most of the streets/ towns on the prefectural boundary are far away from the prefectural centers. These peri-urban areas can easily suffer from poor hygiene standards and indirectly promote vector clusters (Goto et al., 2013). The results were based on a spatial resolution of 1 km. Compared with previous studies performed on a coarse scale (Liu et al., 2014; Wang et al., 2013a,b; Fan et al., 2014; Li et al., 2013), the results are more detailed and specific, and could provide more

Table 3			
Summary of OLS	results and	i OLS d	iagnostics.

.. .

Summary of OLS results				
Variable	Coefficient <sup>a</sup>	Probability <sup>b</sup>	Robust_Pr <sup>b</sup>	VIF <sup>c</sup>
Intercept	-5215.583917	0.000000*	0.000000*	-
LT	20.400690	0.000087*	0.000004*	1.140798
LSTN	17.460860	0.000000*	0.000000°	1.478022
PD	31.671152	0.000000°	0.000000°	1.476165
OLS diagnostics				
Number of observations:	683	Akaike's information criterion (AICc) <sup>d</sup> :		7955.588258
Multiple R-squared <sup>d</sup> :	0.325421	Adjusted R-squared <sup>d</sup> :		0.320424
Joint F-statistic <sup>e</sup> :	83.726619	Prob(>F), (3679) degrees of freedom:		0.000000
Joint Wald statistic <sup>e</sup> :	191.299979	Prob(>Chi-squared), (3) degrees of freedom:		0.000000
Koenker (BP) statistic <sup>f</sup> :	27.182036	Prob(>Chi-squared), (3) degrees of freedom:		0.557469
Jarque-Bera statistic <sup>g</sup> :	12 276.417501	Prob(>Chi-squared), (2) degrees of freedom:		0.812536

OLS, ordinary least squares; LT, land types; LSTN, land surface temperature of nighttime; PD, population density.

\* An asterisk next to a number indicates a statistically significant *p*-value (p < 0.01).

Coefficient: represents the strength and type of relationship between each explanatory variable and the dependent variable.

<sup>b</sup> Probability and robust probability (Robust\_Pr): the asterisk (\*) indicates a coefficient is statistically significant (p < 0.01); if the Koenker (BP) statistic [f] is statistically significant, use the robust probability column (Robust\_Pr) to determine coefficient significance.

<sup>c</sup> Variance inflation factor (VIF): large variance inflation factor (VIF) values (>7.5) indicate redundancy among explanatory variables.

<sup>d</sup> R-squared and Akaike's information criterion (AICc): measures of model fit/performance.

<sup>e</sup> Joint F and Wald statistics: the asterisk (\*) indicates overall model significance (*p* < 0.01); if the Koenker (BP) statistic [f] is statistically significant, use the Wald statistic to determine overall model significance.

<sup>f</sup> Koenker (BP) statistic: when this test is statistically significant (p < 0.01), the relationships modeled are not consistent (either due to non-stationarity or heteroscedasticity). You should rely on the robust probabilities (Robust\_Pr) to determine coefficient significance and on the Wald statistic to determine overall model significance.

<sup>g</sup> Jarque-Bera statistic: when this test is statistically significant (*p* < 0.01) model predictions are biased (the residuals are not normally distributed).

useful information for decision-making and recommendations for dengue fever prevention and control.

Dengue fever is an acute infectious disease caused by infection with any one of four serotypes of dengue virus (DENV 1-4), which are transmitted by Aedes mosquitoes (Lai et al., 2015). There is sufficient evidence to argue that dengue fever is most likely influenced by a complex combination of factors rather than a single focus pathogenic factor, including environmental, demographic, entomological, and epidemiological factors (Hashizume et al., 2012; Lippi et al., 2018; Castro et al., 2018; Vanwambeke et al., 2006; Thammapalo et al., 2005). The dengue fever outbreak was positively correlated with LT, NDWI, LSTD, LSTN, PD, and GDP, and LSTN and PD were found to contribute more in this research. A developed economy and convenient transportation can promote population movement (Lippi et al., 2018; Qi et al., 2015; Hagenlocher et al., 2013; Wu et al., 2009), and this can promote dengue fever transmission. There was no doubt that dengue fever was positively correlated with PD and GDP in this research. Water and a suitable temperature are essential factors for the larvae of the dengue virus vector, the Aedes mosquito. Human activities are closely related to the house address, work units, and other places of activity. Dengue fever was highly positively correlated with NDWI, LSTD, LSTN, and LT in this research, which is in agreement with the existing research (Fullerton et al., 2014). LSTN is usually lower than LSTD in the same spatial place. The correlation coefficient between dengue fever and LSTN was slightly larger than that between dengue fever and LSTD, which further confirms that the dengue fever outbreak was significantly associated with daily minimum temperature (Sang et al., 2014). Finally, the OLS model of dengue fever was built successfully with dengue fever cases as the dependent variable and the three risk factors of LT, LSTN, and PD as the explanatory variables.

There are several reasons why the adjusted R-squared of the OLS model of dengue fever was not high in this study. First, dengue fever is most likely influenced by a complex combination of factors, including some that were not considered in this research. For example, dengue fever is directly correlated with Aedes mosquito density. Second, several data sources from different periods were used in this research: the final number of dengue fever cases in 2014, GF-1 images from August 7, 2015, MODIS products from July

1 to September 30, 2014, and PD and GDP data interpolated from 2010, which might have influenced the model results. However, these data were adaptable. There were image quality problems, including image blurring, image stripe, etc. for GF-1 multi-band images of the study area from 2013 to 2015. Thus, two scenes of GF-1 images in August 7, 2015 were chosen. Guangzhou is a well developed city and there was little change in LT between 2014 and 2015. Thus LT in 2015 could be used instead of that in 2014. Comprehensively considering the similarity in dengue fever seasonality, influence of precipitation on dengue fever outbreaks, and rainfall characteristics in the study area, the NDWI interpretation product from GF-1 images in August 7, 2015 was adopted. At the same time, comprehensively considering the similarity in dengue fever seasonality, influence of temperature on dengue fever outbreaks, representativeness of land surface temperature, and weather conditions at the same time among the years, as well as the limitations in quality of MODIS images, such as the loss of some local parts in the images, etc., LSTD and LSTN from July 1 to September 30, 2014 were adopted. PD and GDP are relatively stable factors. It was thus considered acceptable to use the interpolated data from 2010 to represent data for 2014. Third, there were machine errors and artificial errors in data processing. Some spatial environmental data were missing, especially in the study marginal areas. A follow-up study will be performed in which other data mining methods will be explored, as well as the relationships between dengue fever outbreaks and environmental factors.

Domestic GF-1 remote sensing images at 16-m spatial resolution were used in this research. Chinese Center for Disease Control and Prevention dengue fever field interventions consider a circular area of 200 m radius from dengue fever case points as the core area for intervention, and the area between 200 m and 400 m radius from the dengue fever case point as the area in which precautions should be taken. Thus more remote sensing images at higher spatial resolution, especially domestic images, should be used in disease prevention and control in the future. The research was based on a mixture of geographical and statistical information and a combination of spatial geographic data and spatial statistical methods. This research quantitatively obtained spatial patterns of dengue fever outbreaks and the correlations between dengue fever

outbreaks and risk factors using spatial analysis methods and spatial statistical methods at a relatively small scale. This work improves our understanding of differences in spatial patterns of dengue fever and the effects of environmental and socio-economic risk factors on dengue fever, and thus can effectively support targeted prevention and control.

#### Conclusions

This research provides valuable information on the spatial patterns and associated environmental and socio-economic risk factors of dengue fever using remote sensing (RS) and geographical information system (GIS) technology in Guangzhou City. Dengue fever cases were clustered in August, September, and October, and were clustered in the areas adjacent to the districts of Baiyun, Yuexiu, Liwan, and Haizhu. The dengue fever outbreak was positively correlated with LT, NDWI, LSTD, LSTN, PD, and GDP. The model equation between dengue fever and LT, LSTN, and PD was constructed, which will provide scientific information for public health practitioners to formulate targeted, strategic plans and implement effective public health prevention and control measures.

## Acknowledgements

This study was supported by the National Major Research and Development ProgramAuthor C (grant numbers 10-Y30B11-9001-14/16 and 2016YFC1200802).and State Key Laboratory of Infectious Disease Prevention and Control.

## **Author contributions**

Liang Lu and Qiyong Liu initiated the study. Yujuan Yue collected the data, cleaned the data, performed the statistical analysis and drafted the manuscript. Jimin Sun, Xiaobo Liu, and Dongsheng Ren revised the manuscript.

#### Ethics statement

No human or animal samples were included in the research presented in this article; therefore ethical approval was not necessary for this research.

## **Conflict of interest**

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## References

- Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature 2013;496(7446):504–7.
- Castro DB, Sampaio VS, Albuquerque BC, Pinto RS, Sadahiro M, Passos RA, et al. Dengue epidemic typology and risk factors for extensive epidemic in Amazonas state, Brazil, 2010–2011. BMC Public Health 2018;18:356.
- Chen B, Liu QY. Dengue fever in China. Lancet 2015;385(9978):1621-2.
- Corner RJ, Dewan AM, Hashizume M. Modelling typhoid risk in Dhaka Metropolitan Area of Bangladesh: the role of socio-economic and environmental factors. Int J Health Geogr 2013;12:13.
- Delmelle E, Hagenlocher M, Kienberger S, Casas T. A spatial model of socioeconomic and environmental determinants of dengue fever in Cali, Colombia. Acta Trop 2016;164:169–76.
- Dewan AM, Abdullah AYM, Shogib MRI, Karim R, Rahman MM. Exploring spatial and temporal patterns of visceral leishmaniasis in endemic areas of Bangladesh. Trop Med Health 2017;45:29.
- ESRI. ArcGIS 10.3 help. ESRI Press; 2017.
- ESRI. Regression analysis tutorial for ArcGIS 10. ESRI Press; 2017.
- Eastin MD, Delmelle E, Casas I, Wexler J, Self C. Intra-and interseasonal autoregressive prediction of dengue outbreaks using local weather and regional

climate for a tropical environment in Colombia. Am J Trop Med Hyg 2014;91 (3):598–610.

- Fan J, Lin H, Wang C, Bai I, Yang S, Chu C, et al. Identifying the high-risk areas and associated meteorological factors of dengue transmission in Guangdong Province, Chinafrom 2005 to 2011. Epidemiol Infect 2014;142(3):634–43.
- Fullerton LM, Dickin SK, Schuster-Wallace CJ. Mapping global vulnerability to dengue using the water associated disease index. United Nations University; 2014.
- Goto K, Kumarendran B, Mettananda S, Gunasekara D, Fujii Y, Kaneko S. Analysis of effects of meteorological factors on dengue incidence in Sri Lanka using time series data. PLoS One 2013;8(5)e63717, doi:http://dx.doi.org/10.1371/journal. pone.0063717 PMID: 23671694.
- Guzman MG, Harris E. Dengue. Lancet 2015;385:453-65.
- Hagenlocher M, Delmelle E, Casas I, Kienberger S. Assessing socioeconomic vulnerability to dengue fever in Cali, Colombia: statistical vs expert-based modeling. Int J Health Geogr 2013;12(1):36.
- Hashizume M, Dewan AM, Sunahara T, Rahman MZ, Yamamoto T. Hydroclimatological variability and dengue transmission in Dhaka, Bangladesh: a time-series study. BMC Infect Dis 2012;12:98.
- Khormi HM, Kumar L. Modelling dengue fever risk based on socioeconomic parameters, nationality and age groups: GIS and remote sensing based case study. Sci Total Environ 2011;409:4713–9.
- Lai SJ, Huang ZJ, Zhou H, Anders KL, Perkins TA, Yin W, et al. The changing epedemiology of dengue in china 1990-2014: a descriptive analysis of 25 years of nationwide surveillance data. BMC Med 2015;13:100.
- Lai LW. Influence of environmental conditions on asynchronous outbreaks of dengue disease and increasing vector population in Kaohsiung, Taiwan. Int J Environ Health Res 2011;21(2):133–46.
- Laureano-Rosario AE, Garcia-Region JE, Gomez-Carro S, Farfan-Ale JA, Muller-Karger FE. Modelling dengue fever risk in the state of Yucatan, Mexico using regional-scale satellite-derived sea surface temperature. Acta Trop 2017;172:50–7.
- Lehman A. Jmp for basic univariate and multivariate statistics: a step-by-step guide. Cary, NC: SAS Press; 2005.
- Li S, Tao H, Xu Y. Abiotic determinants to the spatial dynamics of dengue fever in Guangzhou. Asia Pac J Public Health 2013;25(3):239–47.
- Lippi CA, Stewart-Ibarra AM, Muñoz AG, Borbor-Cordova MJ, Mejía R, Rivero K, et al. The social and spatial ecology of dengue presence and burden during an outbreak in Guayaquil, Ecuador, 2012. Int J Environ Res Public Health 2018;15:827.
- Liu CX, Liu QY, Lin HL, Xin BQ, Nie J. Spatial analysis of dengue fever in Guangdong Province, China, 2001-2006. Asia Pac J Public Health 2014;6(1):58–66.
- Mitchell A. The ESRI guide to GIS analysis, Volume 2. ESRI Press; 2005.
- Qi XP, Wang Y, Li Y, Meng YJ, Chen QQ, Ma JQ, et al. The effects of socioeconomic and environmental factors on the incidence of dengue fever in the Pearl River Delta, China, 2013. PLoS Negl Trop Dis 2015;9(10)e0004159, <u>doi:http://dx.doi.org/</u> 10.1371/journal.pntd.0004159.
- Sang S, Yin W, Bi P, Zhang HL, Wang CG, Liu XB, et al. Predicting local dengue transmission in Guangzhou, China, through the influence of imported Cases, mosquito density and climate variability. PLoS One 2014;9(7)e102755, <u>doi:</u> http://dx.doi.org/10.1371/journal.pone.0102755.
- Sang S, Gu S, Bi P, Yang WZ, Xu L, Yang J, et al. Predicting unprecedented dengue outbreak using imported cases and climatic factors in Guangzhou, 2014. PLoS Negl Trop Dis 2015;9(5)e0003808, <u>doi:http://dx.doi.org/10.1371/journal.</u> pntd.0003808.
- Scott I, Warmerdam N. Extend crime analysis with ArcGIS spatial statistics tools in ArcUser online. 2005.
- Shearer FM, Longbottom J, Browne AJ, Pigott DM, Brady OJ, Kraemer MUG, et al. Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. Lancet Glob Health 2018;6:270–8.
- Silverman BW. Density estimation for statistics and data analysis. New York: Chapman and Hall; 1986.

Simmons CP, Farrar JJ, Chau NV, Wills B. Dengue. N Engl J Med 2012;366:1423–32.

- Thammapalo S, Chongsuwiwatwong V, McNeil D, Greater A. The climatic factors influencing the occurrence of dengue hemorrhagic fever in Thailand. Southeast Asian J Trop Med Public Health 2005;36:191–6.
- Tian H, Huang S, Zhou S, Bi P, Yang ZC, Li XJ, et al. Surface water areas significantlyimpacted 2014 dengue outbreaks in Guangzhou, China. Environ Res 2016;150:299–305.
- Vanwambeke SO, van Benthem BH, Khantikul N, Burghoorn-Maas C, Panart K, Oskam L, et al. Multi-level analyses of spatial and temporal determinants for dengue infection. Int J Health Geogr 2006;5:5.
- Wang CG, Jiang BF, Fan JC, Wang FR, Liu QY. A study of the dengue outbreak and meteorological factors in Guangzhou, China, by using a Zero-Inflated Poisson Regression model. Asia Pac J Public Health 2013a;26(1):48–57.
- Wang CG, Yang WZ, Fan JC, Wang FR, Jiang BF, Liu QY. Spatial and temporal patterns of dengue in Guangdong Province of China. Asia Pac J Public Health 2013b;27 (2):844–53.
- Wu PC, Lay JG, Guo HR, Lin CY, Lung SC, Su HJ. Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. Sci Total Environ 2009;407(7):2224–33.
- Wu JY, Lun ZR, James AA. Review: dengue fever in Mainland China. Am J Trop Med Hyg 2010;83(3):664–71.
- Xu L, Stige LC, Chan KS, Zhou J, Yang J, Sang SW, et al. Climate variation drives dengue dynamics. PNAS 2016;114(1):113–8.