

“地理信息分析在公共卫生与健康中的应用”前沿论坛， 南昌，2019.7.29-31

时空变系数 (STVC) 模型： 探测健康地理现象的局域时空非平稳异质成因关系

Spatiotemporally Varying Coefficients (STVC) model
to detect nonstationary heterogeneous causal relationships at local spatiotemporal scales

汇报人：宋 超 副研究员，西南石油大学

导 师：Xun Shi Full Professor, Dartmouth College

王 劲 峰 研究员，中科院地理科学与资源研究所

江西师范大学

地理与环境学院中顺学术报告厅

2019年7月29日 10:30 AM

1. Background——1.1 Regression analysis

Regression analysis purposes

- (1) **Explore relationships among variables.** In restricted circumstances, regression analysis can be used to infer **causal relationships** between the independent (or 'predictors') and dependent variables.
 - Correlational relationship;
 - Intrinsic sequential logic;
 - Controlling other factors.
- (2) **Prediction and forecasting**, where its use has substantial overlap with the field of machine learning.

A regression model relates Y to a function of X and β .

$$Y \approx f(X, \beta)$$

Where,

Y is **dependent (or response, outcome) variable**.

X is **independent (or covariate, explanatory) variables**.

β are **unknown parameters/regression coefficients**, representing Y-X correlated or causal relationships.

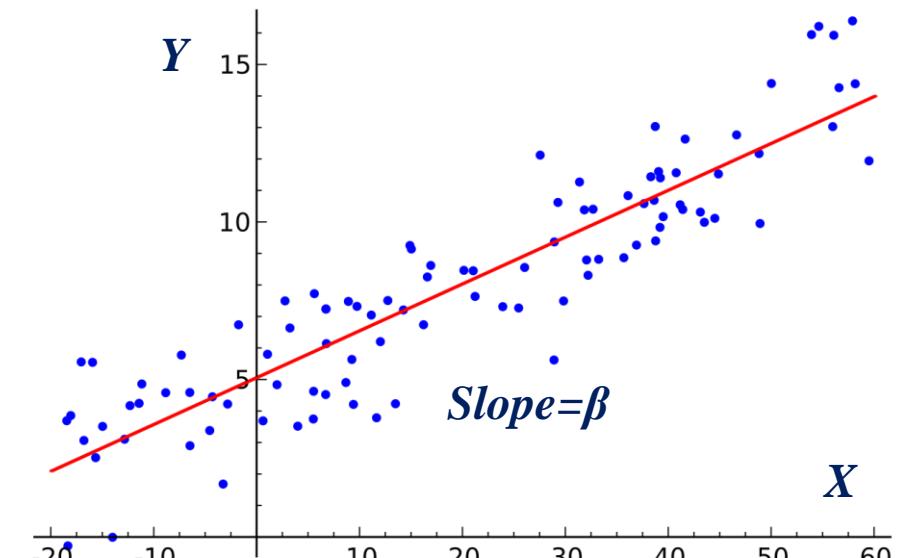
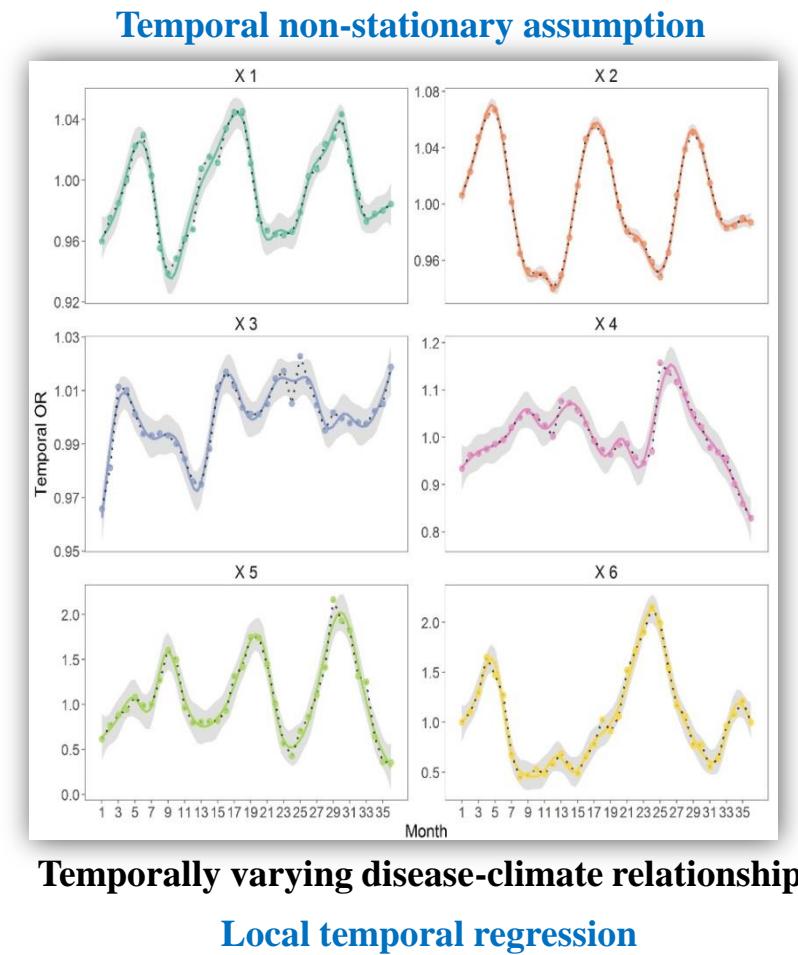
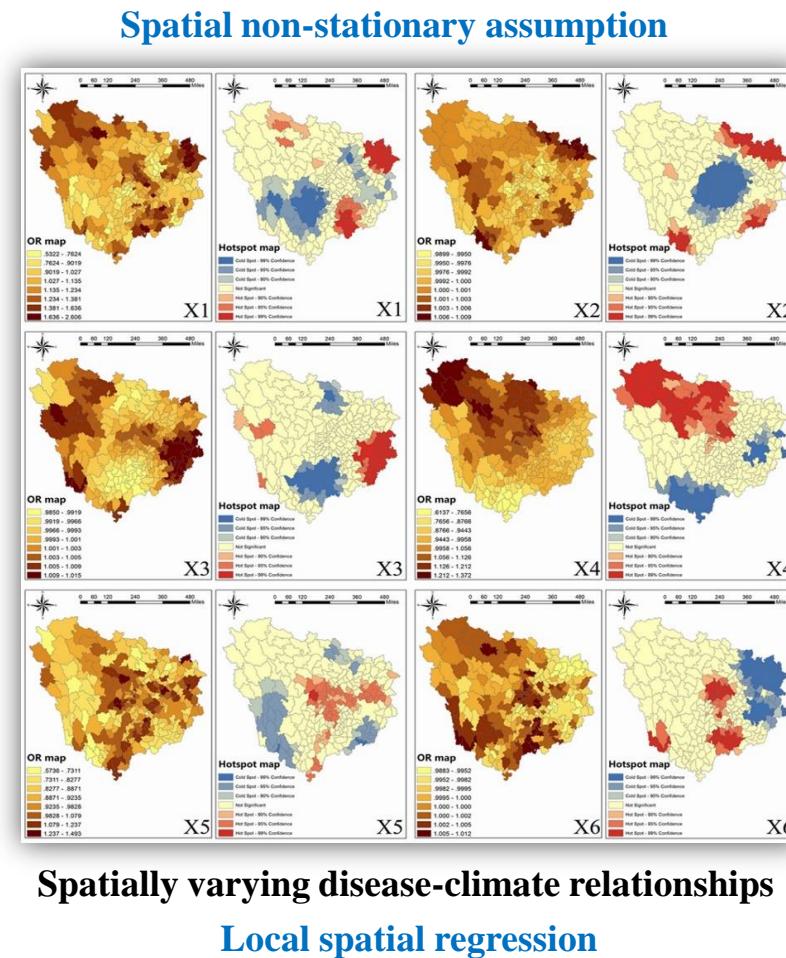
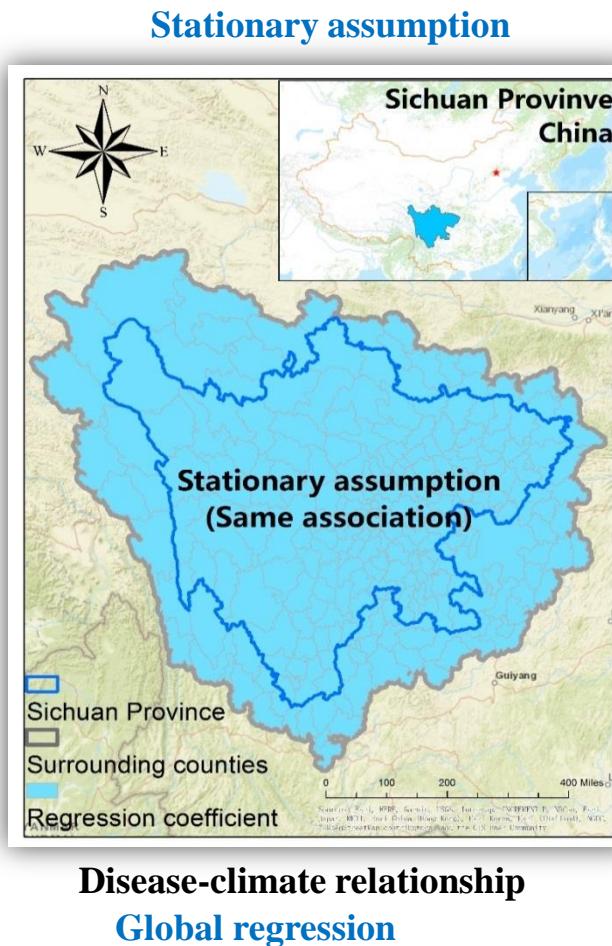


Illustration of linear regression on a data set

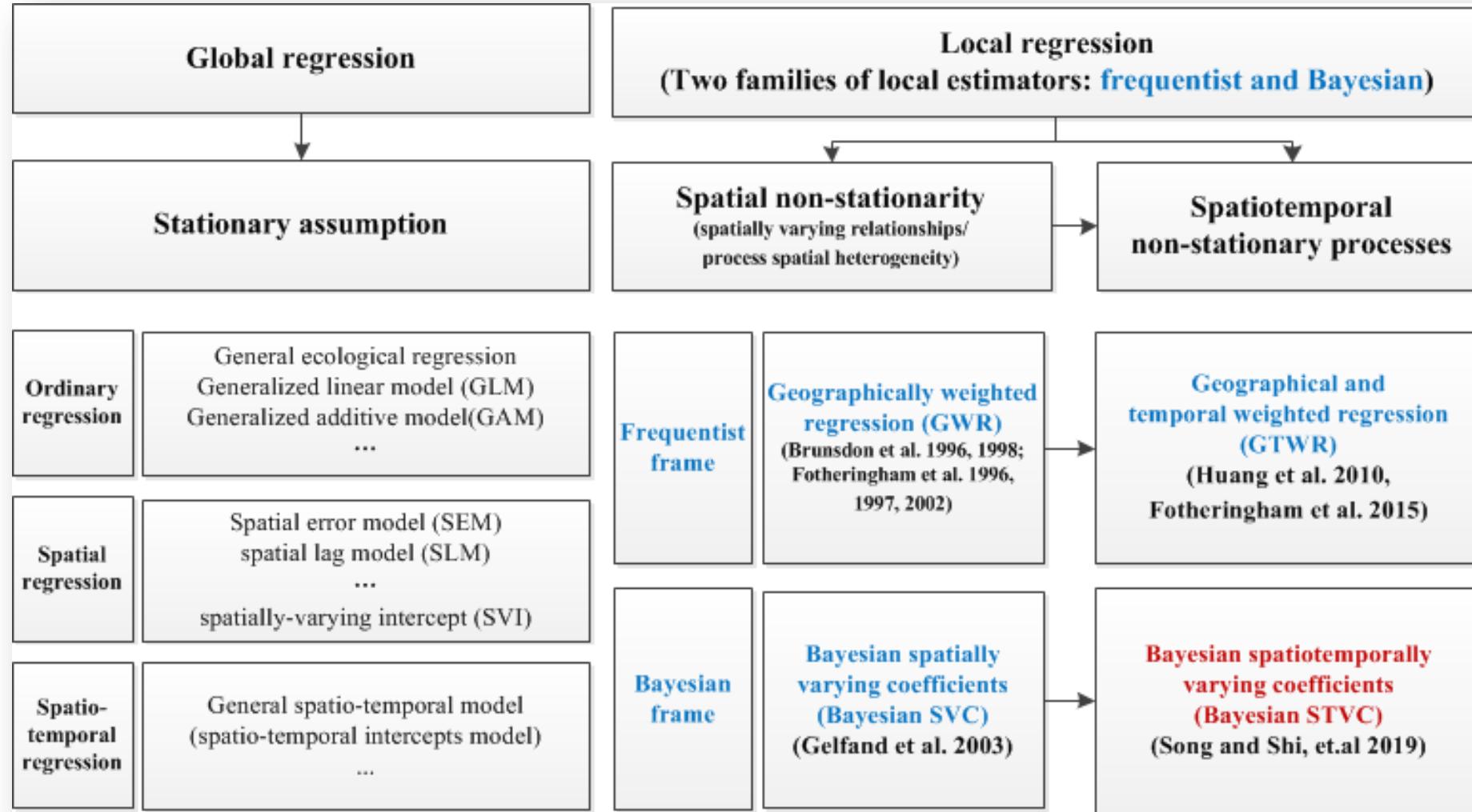
1. Background——1.2 Global and local regression

- Global regression based on **stationary assumption**: Y - X variables relationships (i.e., regression coefficients) are homogeneous across the entire area and time period.
- Local regression based on **non-stationary assumption**: Y - X variables relationships (i.e., regression coefficients) vary among different spatial units and/or time frames.



1. Background——1.3 Why proposing Bayesian STVC model?

- Note that the core difference between global and local regression is whether consider nonstationary effects for **regression coefficients** which represent **dependent-independent relations**, not for intercepts which represent residuals or random errors.



1. Background——1.4 First official named model as Bayesian STVC

≡ Google 学术搜索

Bayesian STVC

文章 找到约 31 条结果 (用时0.03秒)

时间不限 2019以来 2018以来 2015以来 自定义范围...

按相关性排序 按日期排序

不限语言 中文网页 简体中文网页

包括专利 包含引用

创建快讯

... nonstationary effects of climate factors on hand, foot, and mouth disease using Bayesian Spatiotemporally Varying Coefficients (STVC) model in Sichuan, China
C Song, X Shi, Y Bo, J Wang, Y Wang... - *Science of The Total ...*, 2019 - Elsevier
Background Pediatric hand, foot, and mouth disease (HFMD) has generally been found to be associated with climate. However, knowledge about how this association varies spatiotemporally is very limited, especially when considering the influence of local ...
★ 99 被引用次数 : 5 相关文章 所有 6 个版本

[PDF] ... Non-stationary Effects of Climate Factors on Hand, Foot, and Mouth Disease Using Spatiotemporally Varying Coefficient (STVC) Model in Sichuan, China
C Song, X Shi, Y Bo, J Wang, Y Wang, D Huang - [researchgate.net](https://www.researchgate.net)
... Hand, Foot, and Mouth Disease Using Spatiotemporally Varying Coefficient (STVC) Model in Sichuan, China Authors ... covariate is. Table S4. Bayesian model selection with DIC evaluation
Removed covariate each time DIC DIC change Selection SE9 9027.82 135.76 Y ...
☆ 99 相关文章 88

[引用] An analysis of inflation persistence and the Great Moderation using a model of deterministic structural change
G Kapetanios, T Yates - 2008 - Mimeo, Bank of England
☆ 99 被引用次数 : 2 相关文章

Quantitative trait loci for fertility traits in Finnish Ayrshire cattle
NF Schulman, G Sahana... - *Genetics* ..., 2008 - gsejournal.biomedcentral.com
... Since the pleiotropic and the linked-QTL models are not nested, the Bayesian Information Criterion (BIC ... model relative to the posterior probability of the linked QTL model (Bayes factor ... I). The single-trait model with variance component analysis (STVC) confirms QTL on BTA1 and ...
☆ 99 被引用次数 : 36 相关文章 所有 19 个版本 88

CyberGIS-enabled spatial decision support for supply chain optimization with uncertainty quantification
H Hu - 2018 - ideals.illinois.edu
... advanced cyberinfrastructure. Furthermore, an innovative method combining Bayesian hierarchical modeling with stochastic programming is proposed to explicitly account for ... estimation; 3. Develop a Bayesian statistical model to quantify spatiotemporal uncertainty in the supply ...
☆ 99 相关文章 所有 2 个版本 88

Spatio-temporal disease mapping using INLA

B Schrödle, L Held - *Environmetrics*, 2011 - Wiley Online Library

... Abstract. Spatio-temporal disease mapping models are a popular tool to describe the pattern of disease counts. They are usually formulated in a hierarchical Bayesian framework with latent Gaussian model ... 2. APPROXIMATE BAYESIAN INFERENCE USING INLA ...

☆ 99 被引用次数 : 130 相关文章 所有 3 个版本

Local-scale landslide susceptibility mapping using the B-GeoSVC model

Y Yang, J Yang, C Xu, C Xu, C Song - *Landslides*, 2019 - Springer

... 2016), naive Bayes (Tsangaratos and Ilia 2016a), and decision tree (Tsangaratos and Ilia 2016b) ... To address the problems discussed above, we propose a new integrated statistical method under the hierarchical Bayesian framework, named the B-GeoSVC model, of which the ...

☆ 99 相关文章 所有 2 个版本

[HTML] Spatiotemporal Distribution of Hand, Foot, and Mouth Disease in Guangdong Province, China and Potential Predictors, 2009–2012

Y Wang, Y Lai, Z Du, W Zhang, C Feng, R Li... - *International journal of ...*, 2019 - mdpi.com

... The Bayesian spatiotemporal model was used to evaluate the spatiotemporal variations of HFMD and identify the potential association with meteorological and socioeconomic factors ... 2.4. Statistical Methods. 2.4.1. Bayesian Spatiotemporal Model ...

☆ 99 相关文章 所有 7 个版本 88

[HTML] Space-time heterogeneity of hand, foot and mouth disease in children and its potential driving factors in Henan, China

X Zhang, C Xu, G Xiao - *BMC infectious diseases*, 2018 - bmcinfectdis.biomedcentral.com

... h ($h = 1, 2, \dots, L$), respectively. Bayesian space-time hierarchy model. Bayesian space-time hierarchy model (BSTHM) was used to analyze the temporal and spatial variations of disease risk. This model can explore the spatial ...

☆ 99 被引用次数 : 1 相关文章 所有 10 个版本 88

Single and multiscale models of process spatial heterogeneity

LJ Wolf, TM Oshan, AS Fotheringham - *Geographical Analysis*, 2018 - Wiley Online Library

... Two broad approaches have emerged: Geographically Weighted Regression (GWR) which follows a frequentist perspective and Bayesian Spatially Varying Coefficients models ... Bayesian spatially varying coefficients model ...

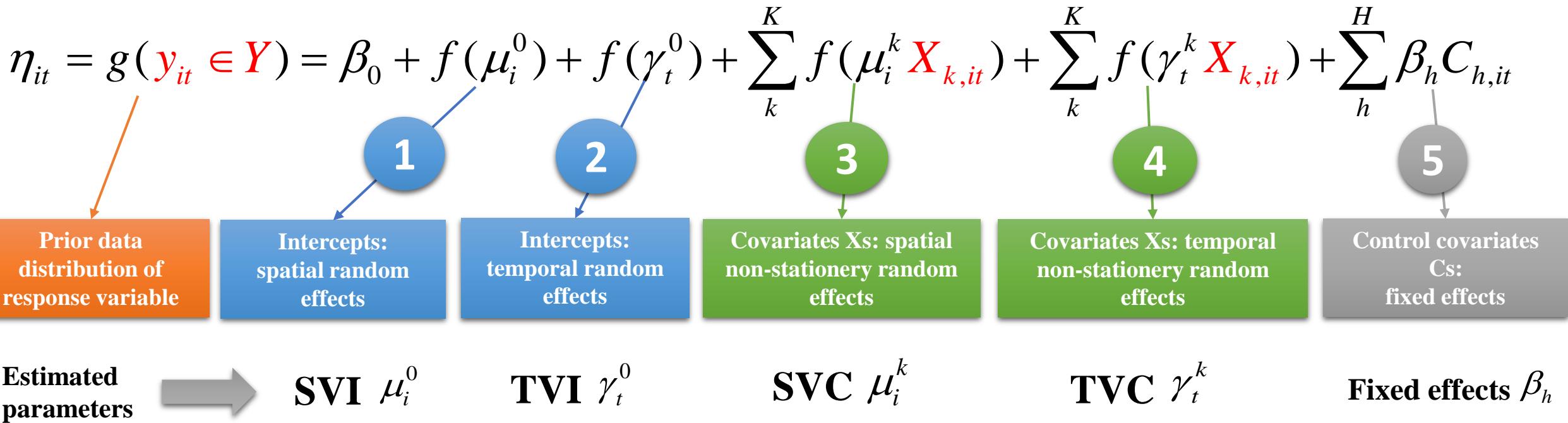
☆ 99 被引用次数 : 9 相关文章 所有 2 个版本

1. Background——1.5 Comparisons between GWR and Bayesian SVC

- **Key literature review:**
 - (1) Bayesian SVC produced more accurate estimates of the regression coefficients than GWR, in the presence of moderate and strong **collinearity**, using point-referenced public health data (Wheeler & Calder 2007, Wheeler & Waller 2009).
 - (2) GWR models were inferior to the SVC models, and GWR might be not robust to **collinearity** among the covariates, supported by two ecological data sets (Finley, 2011).
 - (3) GWR has been shown to be more robust **in larger samples** (Paez, Farber, and Wheeler 2011; Fotheringham and Oshan 2016).
 - (4) However, the interpretation of the **bandwidth parameter is not transferrable** between the two frameworks because the **correlation kernel** used to specify spatial mixed effects in the SVC simply behaves differently from the **data-borrowing kernel** of a GWR (Wolf et.al, 2017).
- **Conclusion:** ultimately, the choice of which local model to employ to study spatial nonstationary process might come down to personal preference, access to software and familiarity with the broad statistical framework in which each model sits (Wolf et.al, 2017).

2. Bayesian STVC——2.1 Theory of a general STVC model

- The **Spatiotemporally Varying Coefficients (STVC)** model is a recent proposed Bayesian local regression approach for detecting both spatial and temporal nonstationary relationships between the dependent and various independent variables (process heterogeneity) , which is an essential development and extension of the local analysis methods (Song and Shi, et.al. 2019).



η_{it} Structured additive predictor; Y : target dependent variable;
 $g(\cdot)$ Likelihood function; $f(\cdot)$ Latent Gaussian model function;
 \mathbf{X}_k Main/core independent variables ($1-K$) with spatiotemporal nonstationary assumption;
 \mathbf{C}_h Other control independent variables ($1-H$) with stationary assumption.

SVI: spatial varying intercepts
 TVI: temporal varying intercepts
 SVC: spatial varying coefficients
 TVC: temporal varying coefficients

2. Bayesian STVC——2.1 Theory of a general STVC model

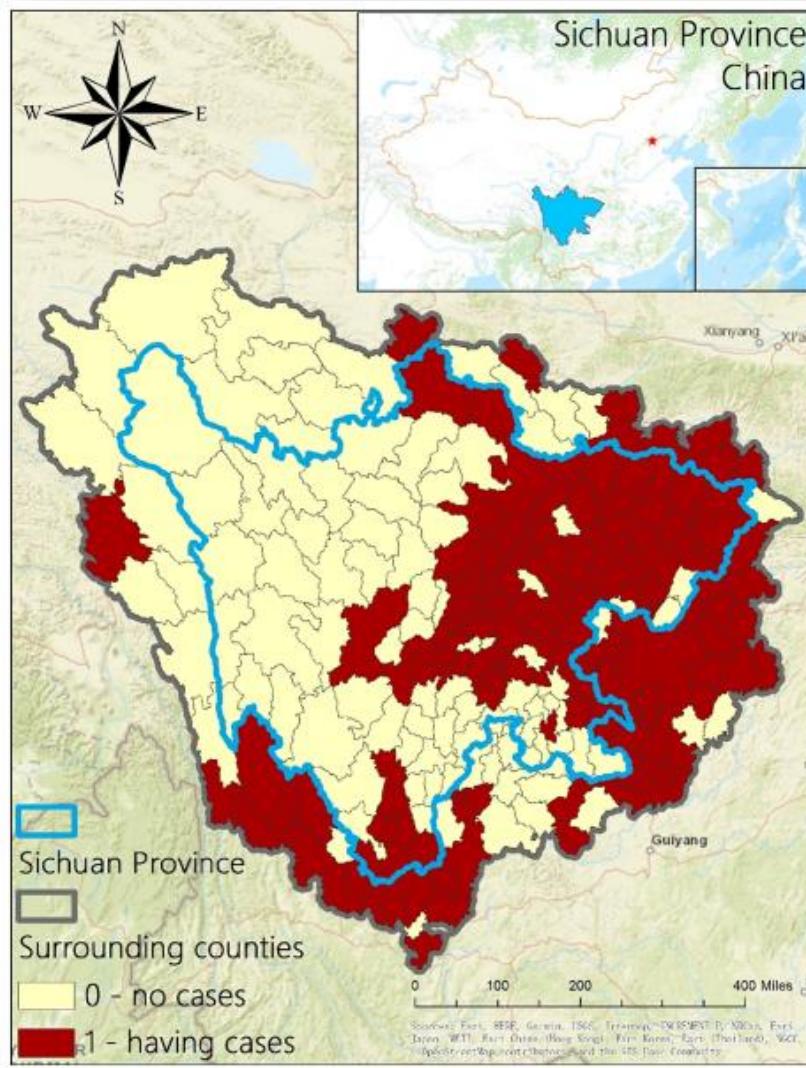
- For spatial autocorrelation sub-level model, **conditional autoregressive f (CAR)** prior is considered for all the spatial random effects, i.e., components 1 and 3.
- For temporal autocorrelation sub-level model, **random walk f (RW)** prior is considered for all the temporal random effects, i.e., components 2 and 4 .
- The CAR prior model assumes that the disease occurrence risk in a spatial area is derived from nearby geographical neighbors (**spatial autocorrelation**). *That is, everything is related to everything else, but near things are more related than distant things -Tobler.*

$$\mu_i^0, \mu_i^k \sim CAR \quad \mu_i \mid \mu_{j \neq i} \sim N\left(\frac{1}{m_i} \sum_{i \sim j} \mu_i, \frac{\sigma^2}{m_i}\right)$$

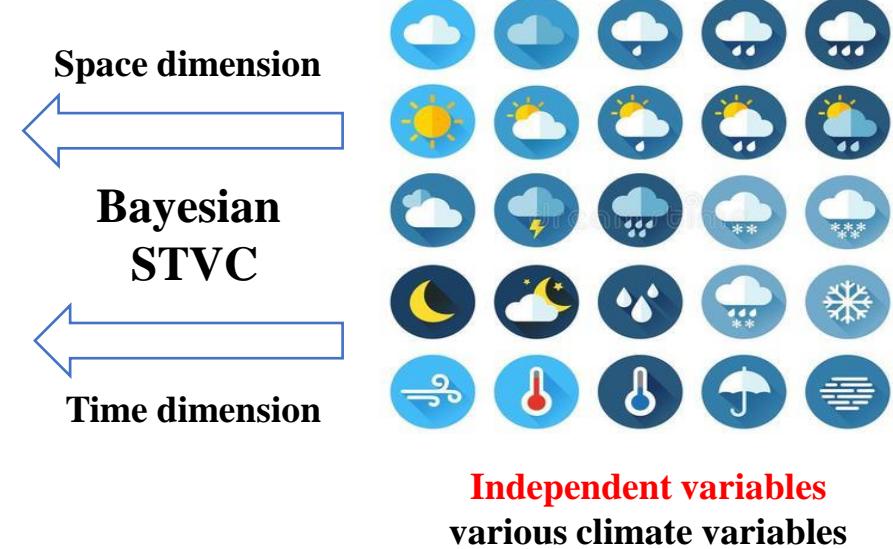
- Similar to CAR, the RW prior model assumes that the temporal variation of disease occurrence risk is influenced by adjacent time points (**temporal autocorrelation**), which is expressed as a smoothly varying nonlinear curve.

$$\gamma_t^0, \gamma_t^k \sim RW \quad \pi(\gamma_t \mid \sigma_\gamma^2) \propto \exp\left(-\frac{1}{2\sigma_\gamma^2} \sum_{t=2}^T (\gamma_t - \gamma_{t-1})^2\right)$$

3. Case study I : explore HFMD-climate nonstationary relationships at local spatiotemporal scales



- Pediatric **hand, foot, and mouth disease (HFMD)** has generally been found to be associated with climate.
- However, knowledge about how this association varies spatiotemporally is very limited, especially when considering the influence of local socioeconomic conditions.
- We use data of monthly county-level HFMD occurrence and data of related climate and socioeconomic variables in Sichuan, China from 2009 to 2011 for our experiments.



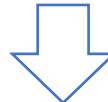
3.1 Disease relative risk assessment indicator OR

- **Odds ratio (OR)** is widely used to identify risk factors for a binary disease outcome variable in classical epidemiology.
- Generally, *OR* indicator can be obtained directly by a logistic regression model.

$$OR_x = e^{\beta} \quad \ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_n x_{n,i} \quad OR > 1, OR < 1, OR = 1$$

- where β is the regression coefficient of the potential risk factor x .

Global *OR*
(stationery assumption)



Original covariates maps
to show **fake** risks

Spatialization/
Downscaling
issue



GWR, SVC

Local spatial *ORs*
(non-stationery assumption)



**True covariate-disease risk
associations maps**

Q: How to obtain local *ORs* to represent local risk at both spatial and temporal scales?

3.2 Spatiotemporal local ORs estimated by STVC model

$$\eta_{it} = \ln\left(\frac{P_{it}}{1-P_{it}}\right) = \beta_0 + f(\mu_i^0) + f(\gamma_t^0) + \sum_k^K f(\mu_i^k X_{k, it}) + \sum_k^K f(\gamma_t^k X_{k, it}) + \sum_h^H \beta_h C_{h, it} \dots$$

Varying coefficients

$\mu_{k,i}$ is the **spatial varying coefficients (SVC)** that represents the spatial covariate-disease risk associations.

$\gamma_{k,t}$ is the **temporal varying coefficients (TVC)** that represents the temporal covariate-disease risk associations.

The **local spatial $OR_{k,i}$ value** of each risk factor is obtained by

$$OR_{k,i} = e^{\mu_{k,i}}$$

The **local temporal $OR_{k,t}$ value** of each risk factor is obtained by

$$OR_{k,t} = e^{\gamma_{k,t}}$$

Local spatial and temporal ORs

3.3 Model implementation

- To evaluate the performance of the STVC model, we compared it with the other four regression models.

Model 1: Ordinary ecological logistic regression model

$$\eta_{it} = \beta_0 + \sum_k^m \beta_k C_k + \sum_j^n \beta_j SE_j$$

Model 2: Temporal ecological logistic regression model

$$\eta_{it} = \beta_0 + \sum_k^m \beta_k C_k + \sum_j^n \beta_j SE_j + \gamma_t + \varphi_t$$

Model 3: Spatial ecological logistic regression model

$$\eta_{it} = \beta_0 + \sum_k^m \beta_k C_k + \sum_j^n \beta_j SE_j + \mu_i + \nu_i$$

Model 4: Spatiotemporal ecological logistic regression model

$$\eta_{it} = \beta_0 + \sum_k^m \beta_k C_k + \sum_j^n \beta_j SE_j + \mu_i + \nu_i + \gamma_t + \varphi_t$$

Model 5: STVC model

$$\eta_{it} = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \xi_i + \psi_t + \sum_k^m f(\mu_{k,i} X_{k,it}) + \sum_k^m f(\gamma_{k,t} X_{k,it}) + \sum_j^n \beta_j SE_j$$

3.4 Model evaluation results under three cross-validation experiments

- The STVC model outperformed these models in the Bayesian model evaluation.

Table 1

Model evaluation results of the five alternative Bayesian models in three cross-validation experiments.

	10% Cross-validation			15% Cross-validation			20% Cross-validation		
	DIC	pD	LS	DIC	pD	LS	DIC	pD	LS
Model 1	8010.98	6.02	0.5750	7597.18	6.02	0.5774	7203.89	6.01	0.5817
Model 2	7324.76	35.86	0.5258	6936.34	35.77	0.5272	6584.97	35.52	0.5318
Model 3	6523.59	190.52	0.4673	6225.77	188.67	0.4722	5896.36	187.99	0.4752
Model 4	6125.04	196.21	0.4394	5840.08	195.11	0.4436	5537.17	194.02	0.4469
Model 5	5191.31	408.24	0.3741	4962.17	301.11	0.3779	4761.72	321.22	0.3848
	PA(1)	PA(0)	PA(1,0)	PA(1)	PA(0)	PA(1,0)	PA(1)	PA(0)	PA(1,0)
Model 1	55.18%	75.77%	64.73%	70.33%	67.70%	69.08%	70.33%	71.65%	70.99%
Model 2	61.45%	81.06%	70.54%	73.44%	70.42%	72.01%	73.46%	75.48%	74.48%
Model 3	76.14%	71.87%	74.16%	80.82%	72.23%	76.74%	79.48%	74.71%	77.07%
Model 4	78.55%	73.26%	76.10%	80.82%	75.32%	78.21%	79.22%	77.78%	78.49%
Model 5	80.48%	81.06%	80.75%	83.93%	79.67%	81.91%	83.66%	82.89%	83.27%

Note: (model 1: ordinary; model 2: temporal; model 3: spatial; model 4: spatiotemporal; model 5: STVC). DIC: deviance information criterion; pD: effective number of parameters; LS: logarithmic score; PA(1): prediction accuracy for have-cases regions; PA(0): prediction accuracy for no-cases regions; PA(1,0): prediction accuracy for all-cases regions.

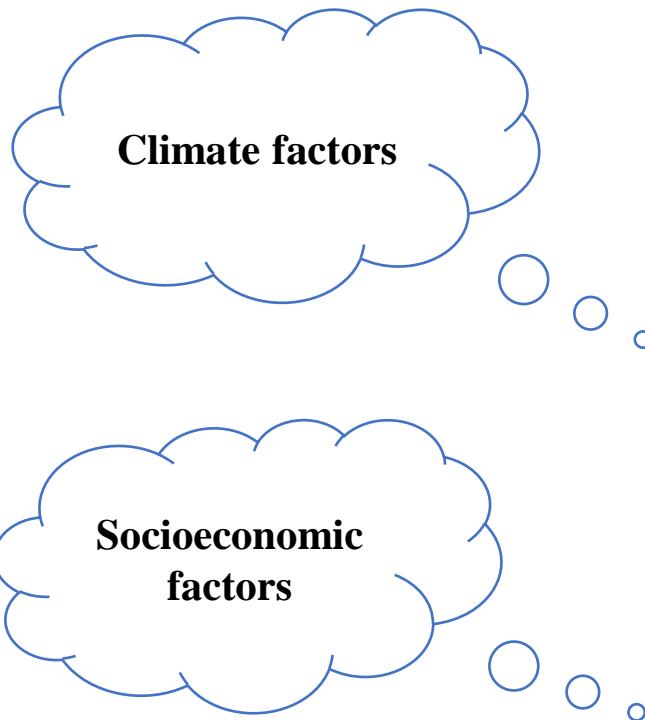
- Cross-validation experiments showed that the STVC model achieved the best average prediction accuracy (81.98%), compared with ordinary (68.27%), temporal (72.34%), spatial (75.99%) and spatiotemporal (77.60%) ecological models.

3.5 Global scale risk indicator (traditional epidemiologic OR)

- Both climate and socioeconomic factors were found to be associated with HFMD occurrence in the study area.

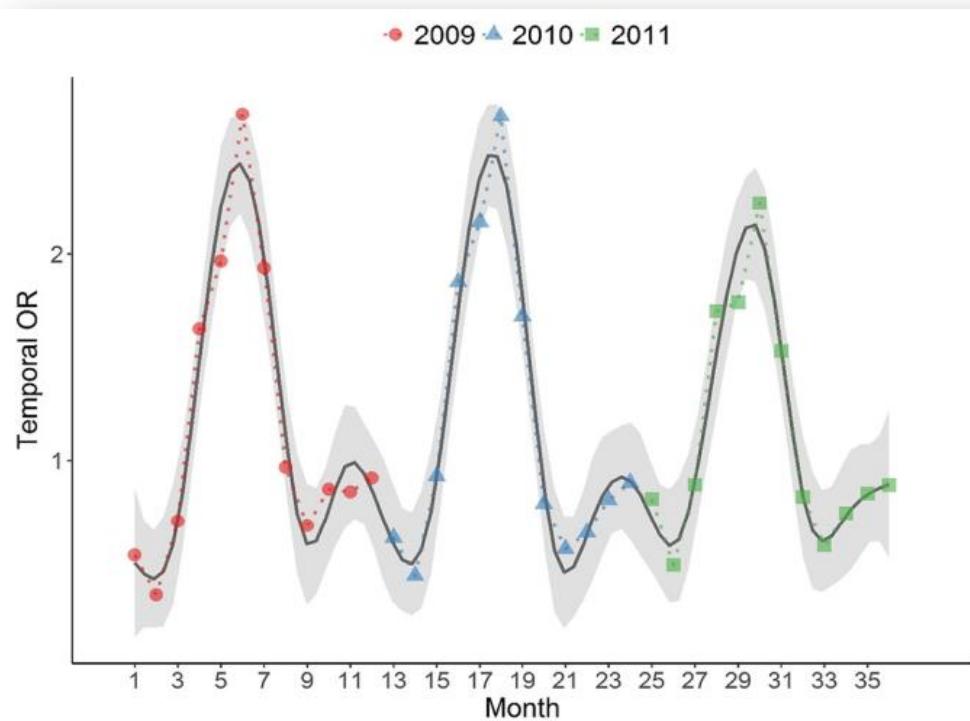
Table 2

Odds ratio (OR) values and estimated posterior parameters of the climate and socioeconomic factors on HFMD occurrence.



Covariates	OR	Mean	SD	0.025 CI	0.975 CI
Temperature	2.59	0.9508	0.2858	0.3980	1.5201
Relative humidity	1.35	0.2978	0.1401	0.0232	0.5729
Sunshine hours	1.06	0.0621	0.1207	-0.1749	0.2988
Air pressure	0.95	-0.0423	0.2439	-0.5203	0.4374
Wind speed	0.65	-0.4259	0.1250	-0.6720	-0.1812
Precipitation	0.99	-0.0010	0.0007	-0.0023	0.0004
Enterprise number density	3.43	1.2326	0.2641	0.7166	1.7532
Per capita household savings	2.44	0.8921	0.2068	0.4850	1.2971
Per capita gross domestic product (GDP)	1.41	0.3465	0.1672	0.0183	0.6748
Children gender ratio	1.35	0.3015	0.1152	0.0761	0.5281
Per capita industrial output values	0.71	-0.3486	0.2210	-0.7832	0.0844

3.6 Temporal disease distribution and disease-climate relations (temporally local ORs)



Temporal local ORs of disease itself

X1 temperature

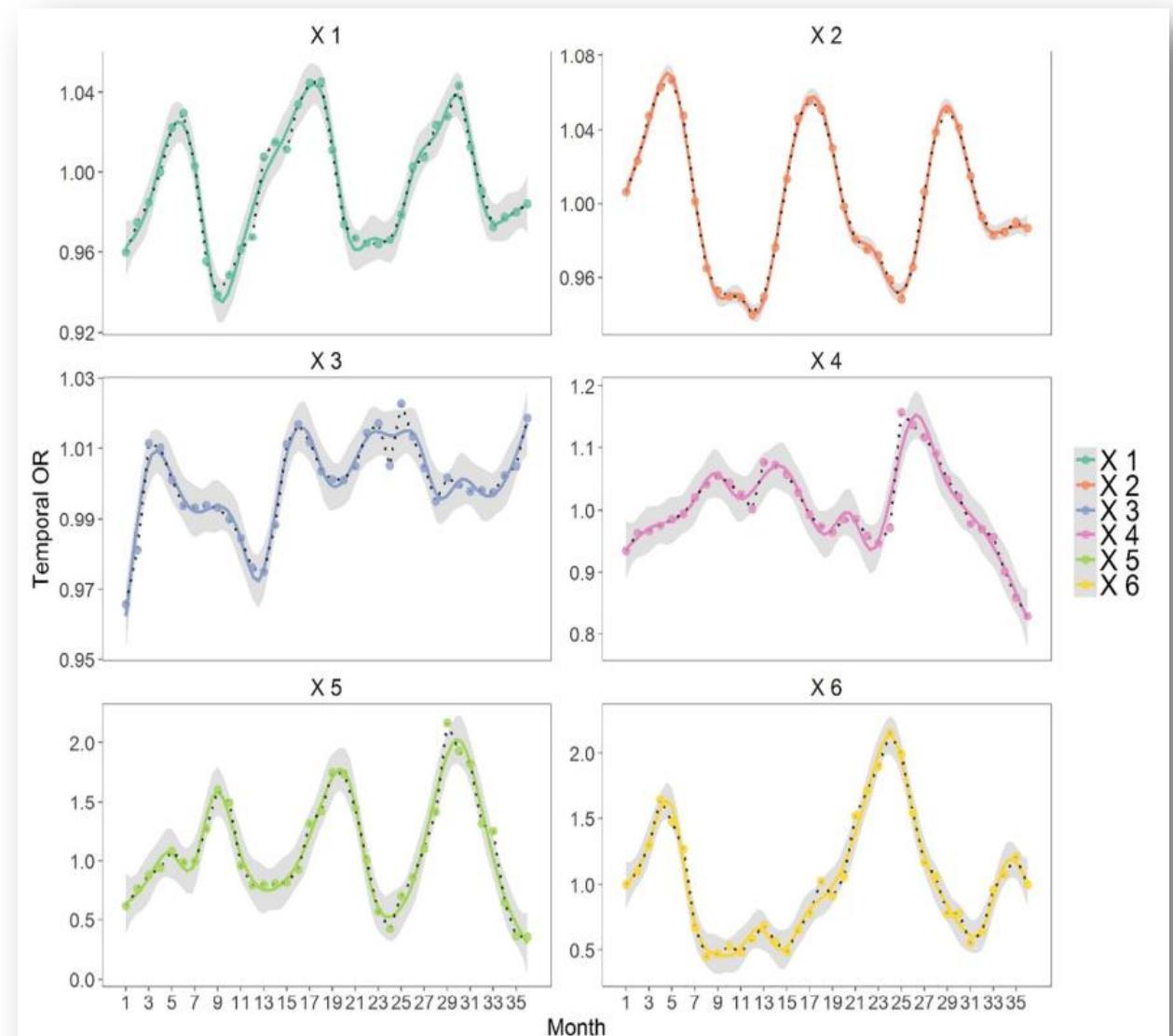
X2 relative humidity

X3 precipitation

X4 air pressure

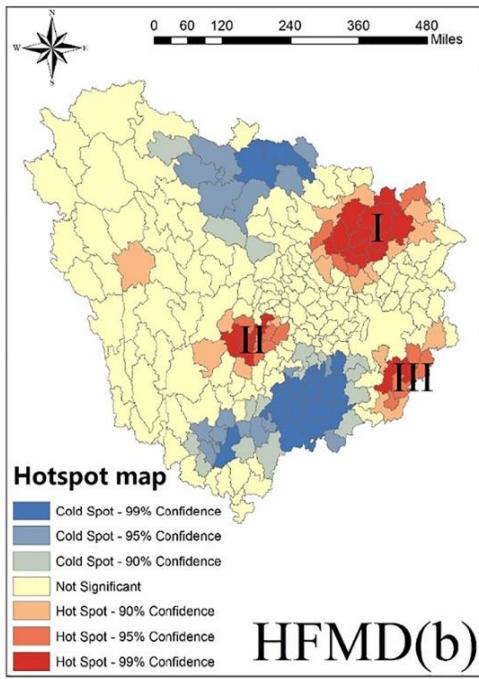
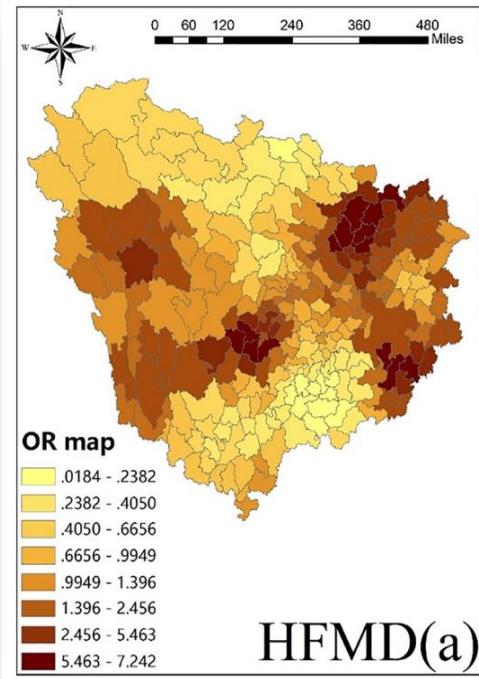
X5 wind speed

X6 sunshine hours



Temporal local ORs of disease-climate relationships

3.7 Spatial disease distribution and disease-climate relations (spatially local ORs)



Spatial local ORs of disease itself

X1 temperature

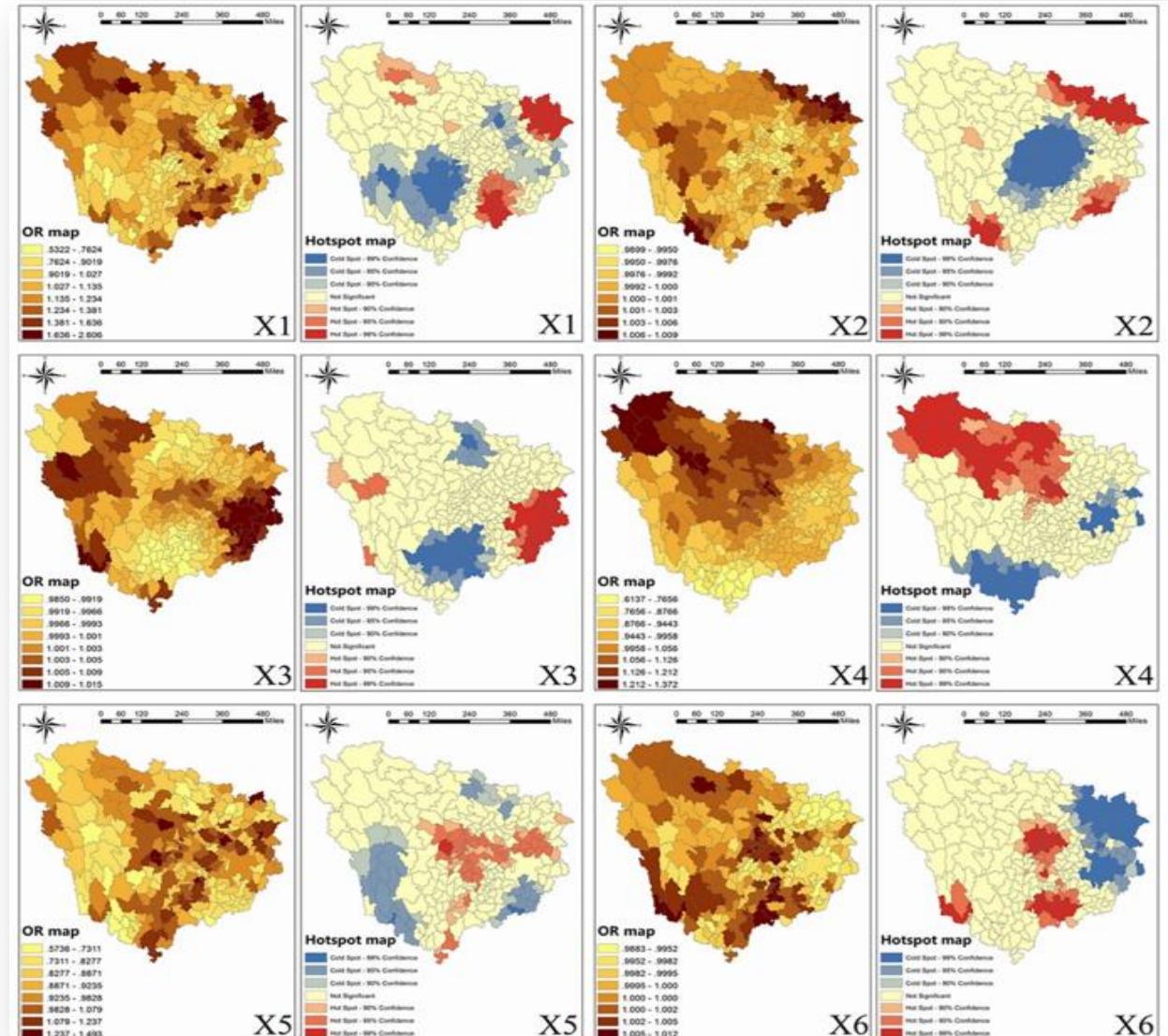
X2 relative humidity

X3 precipitation

X4 air pressure

X5 wind speed

X6 sunshine hours



Spatial local ORs of disease-climate relationships

3.8 Conclusion and innovation

I. Regarding space-time analysis methods development:

- The **Bayesian STVC** model is firstly and officially proposed for exploring the local-scale spatiotemporal nonstationary dependent-independent variables relationships, as well as to further improve model fitness and prediction accuracy.

II. Regarding local-scale epidemiologic risk assessment:

- The STVC model with logistic prior distribution was able to spatialize the risk indicator odds ratio (OR) into **local ORs** to represent spatial and temporal varying disease-climate relationships.

III. Regarding hand, foot and mouth disease (HFMD) propagation:

- We detected local temporal **nonlinear seasonal trends** and **spatial hot spots** for both disease occurrence and disease-climate associations over 36 months in Sichuan, China.
- Among the six representative climate variables, **temperature** (OR = 2.59), **relative humidity** (OR = 1.35), and **wind speed** (OR = 0.65) were not only overall related to the increase of HFMD occurrence, but also demonstrated spatiotemporal variations in their local associations with HFMD.

5. Further studies of Bayesian STVC-family modeling

Limitations of Bayesian STVC:

Overfitting, extreme parameter problem.

High computational burden, high model complexity for large sample size.

No off-the-shelf software for Bayesian STVC modeling.

...

Thanks.

Homepage: <https://chaosong.blog/>

E-mail: songc345@163.com

This work was supported by the National Natural Science Foundation of China (grant number 41701448), the Young Scholars Development Fund of Southwest Petroleum University (grant number 201699010064), a grant from State Key Laboratory of Resources and Environmental Information System, and the State Key Laboratory of Remote Sensing Science.