SUPPLEMENTARY MATERIAL

ARTICLE

Spatiotemporal disparities in regional public risk perception of COVID-19 using Bayesian Spatiotemporally Varying Coefficients (STVC) series models across Chinese cities

JOURNAL

International Journal of Disaster Risk Reduction

CONTENTS

SUPPLEMENTARY MATERIAL 1

Supplementary Note 1. Principal component analysis for COVID-19 regional public attention 2

Supplementary Note 2. Selection of urban socioeconomic covariates 2

Supplementary Note 3. Bayesian random-effect contribution percentage (RCP) index........ 4

Supplementary Note 4. Implementation and evaluation of Bayesian STVC series models 5

Supplementary Note 5. Temporal trends of regional public attention at the provincial level 8

Supplementary Note 6. Spatiotemporal associations of provincial-level public attention with population mobility 9

Supplementary Reference 10

Supplementary Note 1. Principal component analysis for COVID-19 regional public attention

The Baidu Index from the Baidu search engine reflects the collective regional public attention during the COVID-19 pandemic over China[1,2]. According to the rankings in the search volume during the study period, we collected data regarding twelve representative search terms reflective of the evolving process of COVID-19-induced events from the original disease outbreak in China to the widespread pandemic in the worldwide range. Furthermore, we adopted a classical data dimension reduction method, the principal component analysis (PCA), to process the 12 search keywords listed above to extract a single indicator reflective of the daily primary public attention to COVID-19 across 366 cities in China. PCA is a widely known multivariate approach that converts different correlated variables into a few linearly uncorrelated variables named principal components, among which, the first principal component contains the most information about the dataset[3,4]. PCA is performed in the R software environment for statistical computing and graphics.

Supplementary Table 1 summarizes the importance of the first six components in this case. We found that the first principal component (Comp.1) was solely capable of explaining as high as 80.20% variance of all 12 search keywords, which satisfied our data dimension reduction requirement. Hence, we calculated the principal component score of Comp.1 for each space-time unit on the strength of the PCA loading matrix and the observed values of 12 Baidu Index terms. This new PCA-based dimensionality reduction indicator is renamed as the 'Composite Baidu Index' to characterize the overall situation of China's daily public attention to COVID-19 in each city, and further set as the target variable of interest under the regression modeling frame.

Supplementary Table 1. Importance of components in the principal component analysis (PCA)

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	574.20	240.12	90.01	83.72	46.75	42.72
Proportion of variance	80.20%	14.03%	1.97%	1.71%	0.53%	0.44%
Cumulative proportion	80.20%	94.22%	96.20%	97.90%	98.43%	98.88%

Supplementary Note 2. Selection of urban socioeconomic covariates

The regional socioeconomic conditions could reflect disparities in collective searching behaviours among

Chinese citizens residing in different regions, which were therefore selected as spatially control covariates in modeling city-level public attention to COVID-19. We collected 20 urban socioeconomic factors from the China City Statistical Yearbook (http://www.stats.gov.cn/tjsj/). The Variance Inflation Factor (VIF) index was adopted to measure the multicollinearity, which refers to a correlation among these urban socioeconomic factors[5]. Usually, VIF < 10 indicates that the multicollinearity of one variable is acceptable[6].

Supplementary Table 2 summarizes the VIF values of 20 unban socioeconomic factors. In this case, three factors, namely, Gross Regional Product (GDP) per capita (SE2, VIF = 6.35), employment population density of primary industry (SE4, VIF = 1.66), and average salary of employees (SE18, VIF = 3.28) met the screening criteria of VIF < 10, which were set as the city-level control covariates to be added into the spatiotemporal non-stationary regressions.

Supplementary Table 2. China's urban socioeconomic factors (SE1- SE20) and their Variance Inflation Factor (VIF) values

Abbreviation	Socioeconomic factors	Unit	VIF
SE1	Gross Regional Product (GDP)	Yuan	24.68
SE2	GDP per capita	Yuan	6.35
SE3	Above-scale total industrial density	Number/km2	19.06
SE4	Population density of primary industry employees	Person/km2	1.66
SE5	Population density of second industry employees	Person/km2	367.39
SE6	Population density of tertiary industry employees	Person/km2	249.09
SE7	Number of mobile phone users at year-end	household/km2	198.29
SE8	Internet broadband access users	household/km2	78.90
SE9	Local general public budget revenue per capita	Yuan	124.55
SE10	Local general public budget spending per capita	Yuan	24.38
SE11	Total investment in fixed assets per capita	Yuan	58.81
SE12	Total retail sales of consumer goods per capita	Yuan	42.37
SE13	Junior high school student density	Person/km2	261.35
SE14	Primary school student density	Person/km2	129.21
SE15	Hospital density	Number/km2	26.67
SE16	Hospital beds per capita	Number/Person	184.24
SE17	Urban worker population density	Person/km2	1147.36
SE18	Average employee salary	Yuan	3.28
SE19	Year-end financial institutions balance per capita	Yuan	235.70
SE20	Year-end financial institutions loan balance per capita	Yuan	152.79

Supplementary Note 3. Bayesian random-effect contribution percentage (RCP) index

An evaluation index named random-effect contribution percentage (RCP) is proposed for Bayesian-based random effect regressions including Bayesian STVC series models to qualify the relative contribution (percentages) of one or more random effect components in total variations of the target variable. The basic theory of Bayesian RCP is similar to Intraclass Correlation Coefficient (ICC)[7] or Variance Partition Coefficient (VPC)[8].

The Bayesian RCP is defined as,

$$\rho = \frac{\sum \sigma_r^2}{\sum \sigma_o^2 + \sigma_\varepsilon^2} \times 100\% \tag{s5}$$

where ρ is a percentage value ranging in [0,1], σ_o^2 is the sum of the variance of all the implemented random effects, σ_ε^2 is the variance of the unexplained random effect (i.e., the residual term), and σ_r^2 is the sum of the variance of our interested random-effect components (i.e., LGMs).

Practically, σ_r^2 is alternative depending on our actual needs, which can be one single random effect's variance of a specific factor, or a sum of the variance of different-sourced random effects. Regarding Bayesian STVC series models, these random effects can be either the spatiotemporal heterogeneity of the intercepts fitted by STVI and STIVI models, or the spatiotemporal nonstationarity of each explanatory factor's coefficients fitted by STVC and STIVC models. In this case, σ_r^2 can be set as $\sigma_{\mu_{ik}}^2$ for a total of *K* factor's spatial contribution at the city level, $\sigma_{\gamma_{ik}}^2$ for a total of *K* factor's national-level temporal contribution, or $\sigma_{\delta_{(i)tm}}^2$ for a total of *M* factor's provincial-level spatiotemporal interaction contribution.

An essential improvement of this Bayesian RCP index lies in that we can synchronously obtain its credible intervals (CIs) by sampling from the joint posterior of hyperparameters for each random-effect component[8] within the flexible Bayesian modeling environment using R-INLA package in R[9,10].

In this case, we constructed the summary (quantile) statistics using the sampled RCPs of 100,000 times to guarantee a convergent output.

Supplementary Note 4. Implementation and evaluation of Bayesian STVC series models

According to the Bayesian STVC modeling framework for analyzing China's city-specific online public attention and public risk perception to COVID-19 (**Fig. 1**), we implemented five types of Bayesian spatiotemporal regressions, including two spatiotemporal stationary regressions, i.e., STVI (model 1) and STIVI (model 2), and three spatiotemporal nonstationary regressions, i.e., STVC (model 3), SSH-type STIVC (model 4) model, and TSH-type STIVC (model 5). It should be noted that five spatiotemporal models estimate various informative parameters from different viewpoints, and each model is useful for the analysis of China's COVID-19 case. Here, we introduce their modeling paradigms, and compare their modeling performances concerning three aspects, namely, model fitness, model complexity, and model predictive ability.

The formulas of model 1 (STVI) and model 2 (STIVI) are,

$$\eta_{it} = \sum\nolimits_{l=1}^{L} \beta_l X_{itl} + f_S\left(\xi_i\right) + f_T\left(\psi_t\right) + \varepsilon_{it} \tag{s1}$$

$$\eta_{it} = \sum\nolimits_{l=1}^{L} \beta_l X_{itl} + f_S\left(\xi_i\right) + f_{ST}\left(\omega_{z(i)t}\right) + \varepsilon_{it} \tag{s2}$$

where β_l denotes the overall coefficient of the l-th covariate X with L=7 and qualifies the stationary impacts of daily reported disease cases (X1 and X2), daily population mobility (X3 and X4), and urban socioeconomic conditions (X5-X7) on the space-time outcomes of city-level public attention y_{it} across China.

Model 3 (STVC), model 4 (SSH-type STIVC), and model 5 (TSH-type STIVC) are normally formulated by removing the spatiotemporal random effect of intercepts to ensure the fitted local coefficients of each covariate having noticeable and reasonable spatial-temporal variations[11-13]. The customized model 3 and model 4 are formulated using equations (s3) and (s4). The formulation of model 5 is the same as described in equations (7) and (8) of the main body.

$$\eta_{it} = \sum\nolimits_{k=1}^{K} f_S\left(\mu_{ik} SX_{itk}\right) + \sum\nolimits_{m=1}^{M} f_T\left(\gamma_{tm} TX_{itm}\right) + \varepsilon_{it} \tag{s3}$$

$$\eta_{it} = \sum\nolimits_{k=1}^{K} f_S\left(\mu_{ik} SX_{itk}\right) + \sum\nolimits_{m=1}^{M} f_{ST}\left(\delta_{z(i)tm} TX_{itm}\right) + \varepsilon_{it} \tag{s4}$$

Here, spatial covariates SX included all explanatory factors (X1-X7) that were under the spatial nonstationary assumption. In contrast, temporal covariates TX only included those main explanatory factors (X1-X4), which were given temporal or spatiotemporal interaction nonstationary assumptions.

Furthermore, Deviance Information Criterion (DIC)[14] and Watanabe Akaike Information Criterion

($WAIC$)[15] are used to evaluate the Bayesian model fitness, and both are the smaller, the better. Meanwhile, the Bayesian model complexity is evaluated with effective parameters, P_{DIC} and P_{WAIC} (higher values indicate a more complex model). Lastly, the Bayesian model predictive ability is evaluated by the Logarithmic Score (LS) that is the smaller, the better[16]. **Supplementary Table 3** summarizes the performances of the implemented five models based on evaluation indicators DIC, $WAIC$, P_{DIC}, P_{WAIC}, and LS.

Supplementary Table 3. Evaluation for five implemented spatiotemporal regressions (Bayesian STVC series models)

Model	Name	DIC	$WAIC$	P_{DIC}	P_{WAIC}	LS
model 1	STVI	332249	332271	440	454	5.84
model 2	STIVI	320024	320111	1152	1182	5.63
model 3	STVC	334924	335203	1557	1663	5.87
model 4	STIVC (SSH-type)	313669	313528	3142	2523	5.50
Model 5	STIVC (TSH-type)	296767	298004	3404	3948	5.21

NOTE, DIC: Deviance Information Criterion; $WAIC$: Watanabe-Akaike Information Criterion; P_{DIC}: DIC-calculated effective number of parameters; P_{WAIC}: $WAIC$-calculated effective number of parameters; LS: Logarithmic Score; STVI: Spatiotemporally Varying Intercepts model; STIVI: Spatiotemporally Interacting Varying Intercepts model; STVC: Spatiotemporally Varying Coefficients model; STIVC (SSH-type): Spatiotemporally Interacting Varying Coefficients model with the spatiotemporal interaction setting at the SSH scale; STIVC (TSH-type): STIVC model with the spatiotemporal interaction setting at the TSH scale. SSH: spatial stratified heterogeneity, TSH: temporal stratified heterogeneity.

Some informative findings on comparison, selection, and utilization of Bayesian STVC series models are summarized here.

(i) In general, from model 1 to model 5, the model complexity (P_{DIC} and P_{WAIC}) constantly increased but with the apparent improvement of both model fitting degree (DIC and $WAIC$) and predictive ability (LS), due to the constantly refined spatiotemporal assumptions for fitting local coefficients of observable factors, instead of just for fitting local intercepts (unobservable).

(ii) STIVI (model 2) proved to be better than STVI (model 1) in both modeling fitness and prediction ability. Likewise, STIVC (model 4 and model 5) proved to be better than STVC (model 3), suggesting the effectiveness to incorporate the space-time stratification theory (SSH and TSH) to construct the spatiotemporal interaction random effects for both observable factors and the intercept.

(iii) The newly proposed STIVC models (model 4 and model 5) turned out to be the best in model fitness and predictive ability in spite of their highest complexity. Model 5 has better performances than model 4, suggesting that the TSH-type STIVC may be more suitable for prediction research. However, if we are interested in the time differences within different regions, such as the daily changes of COVID-19 public risk perception, the SSH-type STIVC model is the only option.

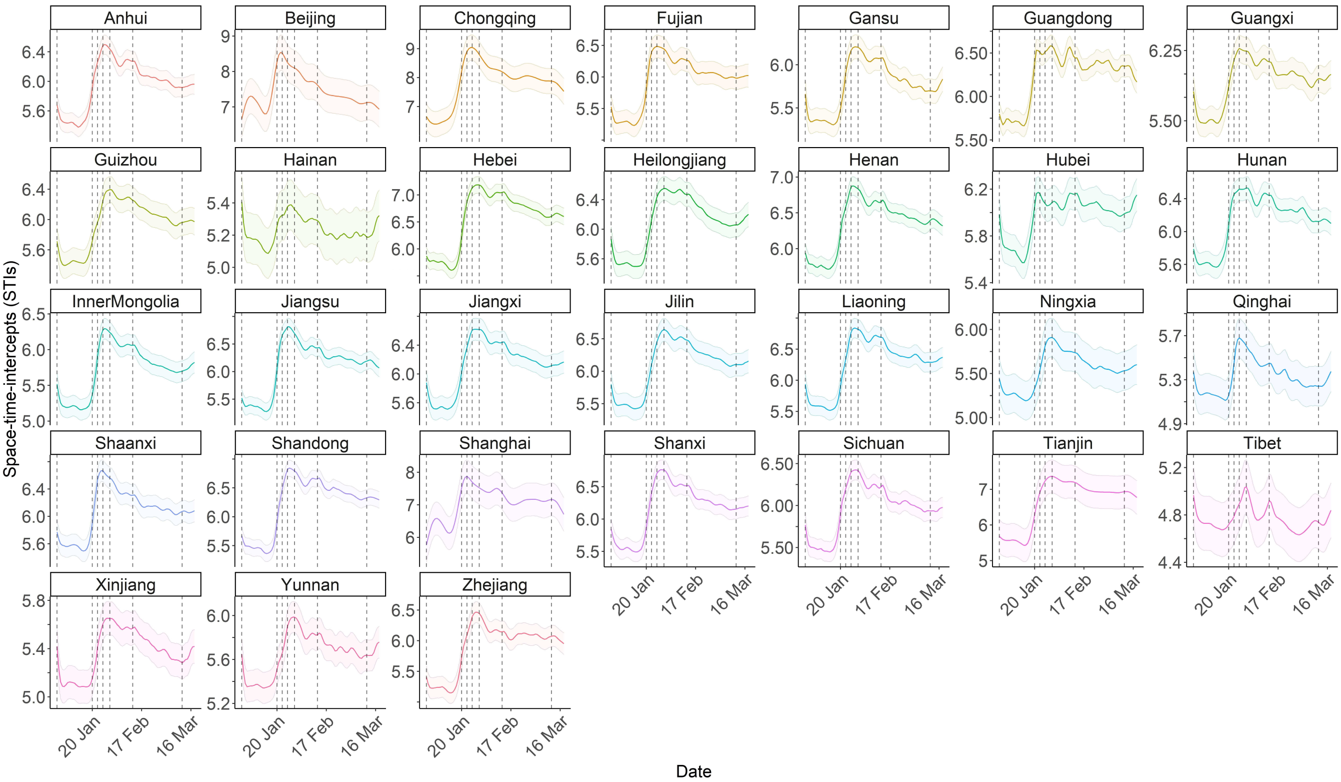
(iv) Among two types of STIVC models, the standard STIVC (SSH-type) model is a more practical model because it identifies the daily temporal heterogeneity within each province, and the extended STIVC (TSH-type) model is only used as a supplement for producing multi-period public risk perception maps. The extended STIVC model sacrifices time precision, which is generally not undesirable in the field of real-time dynamic risk management. Also, given the data collected on emerging infectious diseases or other emergencies on a daily basis, it is also not easy to classify risks at different stages of development.

(v) Compared with TSH, SSH is a more common scenario in the real-world geospatial research, due to geographical divisions such as administrative divisions are generally established and standardized. Geospatial big data pays more attention to regional differences, especially the time differences within the region that the standard STIVC model can directly estimate. Therefore, the standard STIVC (SSH-type) model may be a more practical model for tackling real-world issues, especially those induced by major emergencies.

(vi) This case study once again verifies that local scale spatiotemporal nonstationary regression models (STVC and STIVC) generally have better model performance than global scale spatiotemporal stationary regression models (STVI and STIVI), even without considering the spatiotemporal effects of intercepts[11-13], highlighting the critical role of the observed factors' spatiotemporal nonstationarity in regression family.

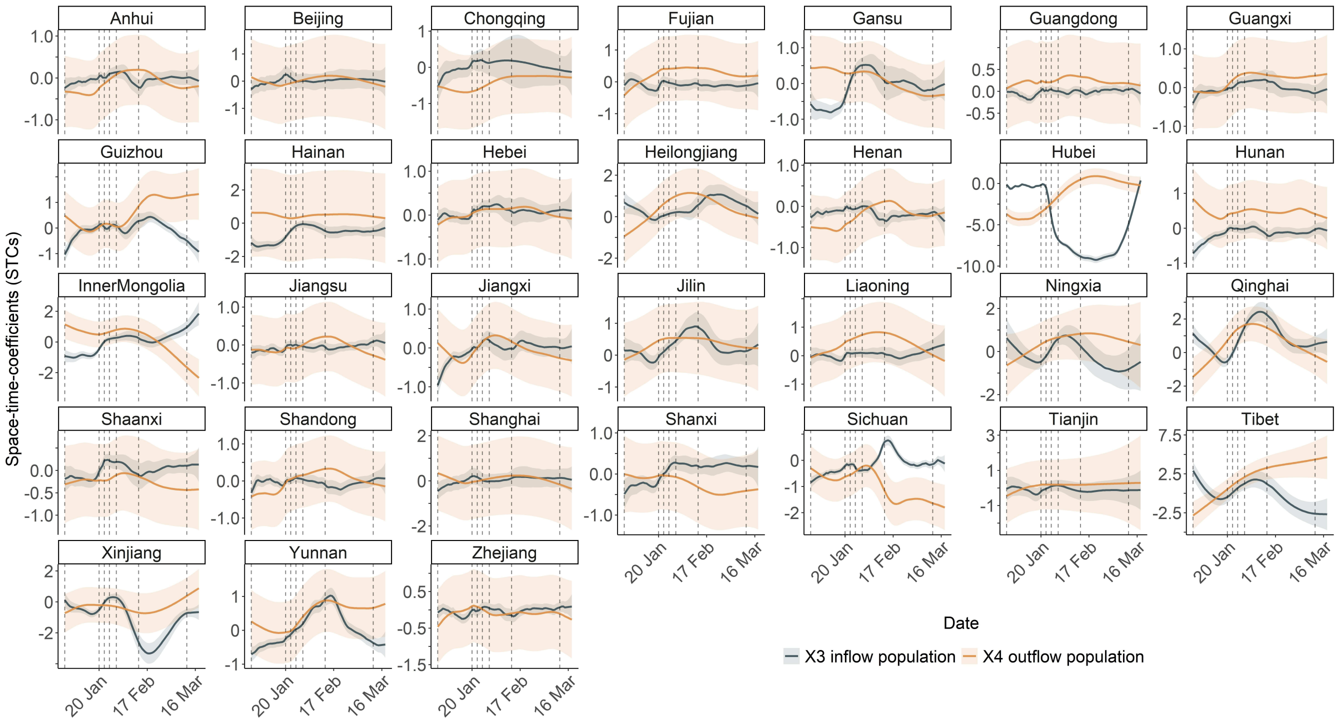
(vii) If further research goal is predicative dominated, we would recommend the STIVC model that generally performs better than the STVC model, especially considering the spatiotemporal nonstationary interactions at the TSH scale.

Supplementary Note 5. Temporal trends of regional public attention at the provincial level



Supplementary Fig. 1 | Differentiated temporal trends of regional public attention to COVID-19 in 31 provinces during the nationwide pandemic wave of China. The key parameter Space-Time-Intercepts (STIs) with 95% CIs is estimated by the STIVI model. Vertical dashed lines represent those iconic COVID-19 events in China (**Fig. 2a**).

Supplementary Note 6. Spatiotemporal associations of provincial-level public attention with population mobility



Supplementary Fig. 2 | Spatiotemporal associations between daily public attention and population mobility (X3 and X4) in 31 provinces of China. The key parameter Space-time-coefficients (STCs) of X3 and X4 with 95% CIs is estimated by the STIVC model. Vertical dashed lines represent those iconic COVID-19 events in China (**Fig. 2a**).

Supplementary Reference

1 Xu, C., Zhang, X. & Wang, Y. Mapping of health literacy and social panic via web search data during the COVID-19 public health emergency: infodemiological study. *Journal of Medical Internet Research* **22**, e18831 (2020).

2 Gong, X., Han, Y., Hou, M. & Guo, R. Online Public Attention During the Early Days of the COVID-19 Pandemic: Infoveillance Study Based on Baidu Index. *JMIR public health and surveillance* **6**, e23098 (2020).

3 Mahmoudi, M. R., Heydari, M. H., Qasem, S. N., Mosavi, A. & Band, S. S. Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries. *Alexandria Engineering Journal* **60**, 457-464 (2021).

4 Joshua, V., Grace, J. S., Emmanuel, J. G., Satish, S. & Elangovan, A. Spatial mapping of COVID-19 for Indian states using Principal Component Analysis. *Clinical Epidemiology and Global Health* **10**, 100690 (2021).

5 Kim, J. H. Multicollinearity and misleading statistical results. *Korean journal of anesthesiology* **72**, 558 (2019).

6 O'brien, R. M. A caution regarding rules of thumb for variance inflation factors. *Quality & quantity* **41**, 673-690 (2007).

7 Goldstein, H., Browne, W. & Rasbash, J. Partitioning variation in multilevel models. *Understanding statistics: statistical issues in psychology, education, and the social sciences* **1**, 223-231 (2002).

8 Wang, X., Ryan, Y. Y. & Faraway, J. J. *Bayesian Regression Modeling with INLA*. (Chapman and Hall/CRC, 2018).

9 Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)* **71**, 319-392, doi:https://doi.org/10.1111/j.1467-9868.2008.00700.x (2009).

10 Rue, H. *et al.* Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application* **4**, 395-421, doi:https://doi.org/10.1146/annurev-statistics-060116-054045 (2017).

11 Song, C., Shi, X. & Wang, J. Spatiotemporally Varying Coefficients (STVC) model: A Bayesian local regression to detect spatial and temporal nonstationarity in variables relationships. *Annals of GIS* **26**, 277-291 (2020).

12 Zhang, X. *et al.* Socioeconomic and Environmental Impacts on Regional Tourism across Chinese Cities: A Spatiotemporal Heterogeneous Perspective. *ISPRS International Journal of Geo-Information* **10**, 410 (2021).

13 Song, C. *et al.* Spatial and Temporal Impacts of Socioeconomic and Environmental Factors on Healthcare Resources: A County-Level Bayesian Local Spatiotemporal Regression Modeling Study of Hospital Beds in Southwest China. *International Journal of Environmental Research and Public Health* **17**, 5890 (2020).

14 Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)* **64**, 583-639, doi:https://doi.org/10.1111/1467-9868.00353 (2002).

15 Watanabe, S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**, 3571-3594 (2010).

16 Held, L., Schrödle, B. & Rue, H. in *Statistical modelling and regression structures* (eds Thomas Kneib & Gerhard Tutz) 91-110 (Springer, 2010).