# Supplementary materials

## S1. Datasets of online public attention to COVID-19

The Baidu Index from the Baidu search engine reflects the collective public attention during the COVID-19 pandemic over China1,2. We selected 12 Baidu Index terms of COVID-19 based on those ranking high in search volume from 31 December 2019 to 18 March 2020, which are recommended by an official Baidu Keyword Mining platform (http://stool.chinaz.com). These final identified keywords are: ‘不明原因肺炎 (Pneumonia of unknown cause, in Chinese)’, ‘全国新冠肺炎疫情实时动态 (Real-time updates on the COVID-19 across the country, in Chinese)’, ‘无症状感染者 (Asymptomatic carrier, in Chinese)’, ‘新冠肺炎 (COVID-19, in Chinese)’, ‘新冠肺炎的症状有哪些症状 (What are the symptoms of COVID-19, in Chinese)’, ‘新冠肺炎最新消息 (The latest on COVID-19, in Chinese)’, ‘疫情地图 (Epidemic map, in Chinese)’, ‘COVID-19’, ‘2019-ncov’, ‘sars’, ‘ncp (novel coronavirus pneumonia)’, and ‘SARS-CoV-2’.

Furthermore, we adopted a classical data dimension reduction method, the principal component analysis (PCA), to process the above 12 search keywords to extract a single indicator to represent the daily primary public attention to COVID-19 across Chinese 367 cities. PCA is a famous multivariate approach that converts different correlated variables into a few linearly uncorrelated variables named principal components, and in this conversion, the first principal component contains the most information about the dataset3,4. PCA is performed in the R software environment for statistical computing and graphics. **Table S1** 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 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 modelling frame.

**Table S1**. 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% |

## S2. **Urban socioeconomic factors**

The regional socioeconomic conditions could reflect disparities in collective searching behaviour among Chinese cities, thereby we chose them as spatially control covariates in modelling 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 is adopted to measure the multicollinearity, which refers to a correlation among these urban socioeconomic factors5. Usually, VIF < 10 indicates that the multicollinearity of one variable is acceptable6. **Table S2** summarizes the VIF values for 20 unban socioeconomic factors. In this case, three factors, namely, 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 into the spatiotemporal nonstationary regressions.

**Table S2**. 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 |

## S3. Random-effect Contribution Percentage (RCP) index

Using the local coefficients estimated by Bayesian STVC and STIVC models, we can get the relative contribution of different factors on public attention in each city over the study period. However, it is not always easy to draw a pithy conclusion to characterize the overall contribution of factors from these spatiotemporal parameters. To meet that need, an evaluation index named random-effect contribution percentage (RCP) is proposed for Bayesian random effect regressions to qualify the relative contribution of one or more random effect components (LGMs) in total variations of the target variable. In other words, the RCP index is an evaluation tool in identifying various factors’ contribution on a space-time scale. Here, the RCP index is used as a measure of how much China’s areal public attention to COVID-19 across space and time can be explained by different random-effect components, e.g., temporal, spatial, and spatiotemporal interacting nonstationarity of covariates.

The basic theory of RCP is similar to the intraclass correlation coefficient (ICC)7 or the variance partition coefficient (VPC)8, which is defined as,

, (s1)

where  is a percentage value ranging in [0,1],  is the sum of the variance of all the implemented random-effects,  is the variance of the unexplained random effect (residual term), and  is the sum of the variance of our interested random-effect components (LGMs). The improvement of this index lies in that, we can synchronously obtain the credible intervals (CIs) of the RCP index by sampling from the joint posterior of hyperparameters for each random-effect component8, by taking advantage of the flexible Bayesian modelling environment of R-INLA package in R9,10. In this case, we constructed the summary (quantile) statistics using the sampled RCPs of 100,000 times to guarantee a convergent result.

Practically,  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. Under our Bayesian STVC modelling system, these random effects can be either the spatiotemporal heterogeneity of the intercepts (fitted by STVI and STIVI) or the spatiotemporal nonstationarity of each explanatory factor’s coefficients (fitted by STVC and STIVC). For instance, in this case,  can be set as  for a total of *K* factor’s spatial contribution at the city level,  for a total of *K* factor’s national-level temporal contribution, or  for a total of *M* factor’s provincial-level spatiotemporal contribution.

Compared to mainstream methods in identifying factors’ overall contributions, such as Random Forest11,12 and GeoDetector (*q* statistics)13,14, the advantages of the RCP index lie in two aspects arising from the Bayesian hierarchical modelling frame. On the one hand, in addition to detecting the overall contribution, it can detect the differential contributions of each factor, e.g., in time and space dimensions using STVC-similar models. On the other hand, the Bayesian 95% CIs of contribution for each component can be contained to access evaluation uncertainties.

## S4. Cluster and outlier mapping using Local Moran’s I statistic

The cluster and outlier analysis for the COVID-19 spatial risk perception map was achieved in the software of geospatial analytics ArcGIS. Given a polygon map in shapefile format, the Cluster and Outlier Analysis tool outputs statistically significant hot spots, cold spots, and two types of spatial outliers based on the principle of Anselin Local Moran's I statistic, which is given as15,16,

, . (s2)

where  is the COVID-19 risk perception value within each city (e.g., space-coefficients of reported disease cases estimated from Bayesian STVC and STIVC models), is the mean value of , and  is the spatial weight between city *i* and *j*, with *n* being the total number of cities across the study area of China.

Regarding the concept of spatial risk perception to COVID-19, there are four types of informative zones identified in these cluster and outlier maps, namely, the normal response areas (High-High Cluster), the potential public panic areas (High-Low Outlier), the potential low public literacy areas (Low-High Outlier), and the stable areas (Low-Low Cluster). Those Not Significant areas in maps only indicate that the spatial risk perception among these cities were not statistically significant (> 90% confidence) to form a cluster or an outlier.

## S5. Bayesian STIVC model considering temporal stratified heterogeneity

The previous used Bayesian STIVC model based on the spatial stratified heterogeneity (SSH) theory has been successfully used to estimate both provincial-level spatiotemporal interactive effects as well as spatial residual effects at city level. However, such SSH-based STIVC model estimates the spatiotemporal interaction effect by sacrificing spatial scale, that is, from city to province level in this case. In practice, we may want to keep the spatial scale constant and know their differences between typical periods, and we have data at many time points, which can be divided into different stages reasonably. In order to satisfy this demand, instead of SSH, we introduce the temporal stratified heterogeneity (TSH) to development a new variant of Bayesian STIVC model to estimate space-time interactive coefficients, which are kept to the smallest spatial scale but with disparities between stages.

For each city *i* in China, daily observations marked by  are available. These temporal observations can be further divided into three stages supported by previous time scale results (**Fig. 1b** and **Fig. 4a**), labelled as . This TSH-based STIVC model for fitting the associations between China’s city-level public attention and potential influencing factors is defined as,

,

, , . (s3)

In equation (s3),  denotes that the daily observation *t* belongs to the upper temporal stage *q*. Model estimates three categories of coefficients, i.e., . For *M* explanatory factors  (X1-X4) with both space and time variations, the space-time-coefficients (STCs)  is estimated at the city level across various periods, and the time-coefficients (TCs)  is estimated at the overall national level. As for *D* spatial covariates  (X5-X7) without daily variations, space-coefficients (SCs)  is estimated at the city level. The prior LGMs  for parameters STCs, TCs, and SCs are the same as defined in equations (1) - (4). Here, the space-time interaction component  is defined as , where  denotes the Kronecker product,  denotes a  spatial structure matrix following an iCAR prior model, and  denotes an identity matrix of  following an exchangeable prior distribution that is given for the temporal stratified random effect17,18. This type of space-time interaction assumes that the structured spatial patterns are different from time to time at the TSH level. We used this TSH-based STIVC model to estimate local parameters  to present the temporal dynamics of spatial public risk perception maps of COVID-19 across Chinese cities (**Extended Data Fig. 4**).

## S6. Model implementation and evaluation

In this case, we mainly developed four Bayesian spatiotemporal regressions for analyzing China’s online public attention and city-specific risk perception to COVID-19, including two global-scale spatiotemporal stationary regressions, i.e., STVI (model 1) and STIVI (model 2) models, and two local-scale spatiotemporal nonstationary regressions, i.e., STVC (model 3) and STIVC (model 4) models. It should be noted that four models estimate different types of informative parameters from different viewpoints and each model is useful for the analysis of China’s COVID-19 case (**Extended Data Fig. 1**). Here, we focus on comparing their modelling performances from three aspects, namely, model fitness, model complexity, and model predictive ability.

The formulas of STVI and STIVI models are,

, and (s4)

, (s5)

where  denotes the overall coefficient of the *l*-th covariate with *L*=7, which qualifies the linear numerical 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  in China.

The finally developed spatiotemporal nonstationary regressions were customized by removing the spatiotemporal random effects of intercepts to ensure that the fitted local coefficients of each covariate have noticeable spatial and temporal variations19-21. Such STVC and STIVC models without a spatiotemporal intercept term are formulated as,

, and (s6)

, (s7)

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

**Table S3** summarizes the performances of the above implemented Bayesian regressions using *DIC*, *WAIC*, *PDIC*, *PWAIC*, and *LS*. Among five evaluation indicators, Deviance Information Criterion (*DIC*)22 and Watanabe Akaike Information Criterion (*WAIC*)23 are used to evaluate the model fitness of Bayesian regression, and both are the smaller, the better. Meanwhile, the model complexity is evaluated by using effective parameters, *PDIC* and *PWAIC* (higher values indicate a more complex model). Lastly, the model predictive ability is evaluated by using the Logarithmic Score (*LS*) that is the smaller, the better24.

**Table S3**. Evaluation of Bayesian spatiotemporal regressions for China’s COVID-19 public attention case.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Name | *DIC* | *WAIC* | *PDIC* | *PWAIC* | *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 | 313669 | 313528 | 3142 | 2523 | 5.50 |

*DIC*: Deviance Information Criterion; *WAIC*: Watanabe-Akaike Information Criterion; : *DIC*-calculated effective number of parameters; : *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: Spatiotemporally Interacting Varying Coefficients model.

Some informative findings are summarized here.

1. From model 1 to model 4, the model complexity (*PDIC* and *PWAIC*) constantly increases 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, not just for fitting local intercepts.

2. The newly developed STIVC (model 4) regression turned out as the best model in fitness and predictive ability, despite that it had the highest complexity.

3. STIVI (model 2) is better than STVI (model 1) in modelling fitness and prediction ability. Likewise, STIVC (model 4) turns to be better than STVC (model 3), suggesting the effectiveness to incorporate the SSH theory to construct the space-time interaction random effects.

4. This case study further proves that a local-scale spatiotemporal nonstationary regression (STVC/STIVC) model usually has better model performance than a global-scale spatiotemporal stationary regression (STVI/STIVI) model, due to the critical estimation of the spatiotemporal nonstationarity of a series of observed factors19-21.

## 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 Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. Variable selection using random forests. *Pattern recognition letters* **31**, 2225-2236, doi:<https://doi.org/10.1016/j.patrec.2010.03.014> (2010).

12 Breiman, L. Random forests. *Machine learning* **45**, 5-32 (2001).

13 Wang, J. F. *et al.* Geographical detectors‐based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. *International Journal of Geographical Information Science* **24**, 107-127 (2010).

14 Wang, J., Zhang, T. & Fu, B. A measure of spatial stratified heterogeneity. *Ecological Indicators* **67**, 250-256, doi:<https://doi.org/10.1016/j.ecolind.2016.02.052> (2016).

15 Anselin, L. Local indicators of spatial association—LISA. *Geographical analysis* **27**, 93-115 (1995).

16 Scott, L. M. & Janikas, M. V. in *Handbook of applied spatial analysis* 27-41 (Springer, 2010).

17 Knorr‐Held, L. Bayesian modelling of inseparable space‐time variation in disease risk. *Statistics in medicine* **19**, 2555-2567 (2000).

18 Adin, A., Goicoa, T. & Ugarte, M. D. Online relative risks/rates estimation in spatial and spatio-temporal disease mapping. *Computer methods and programs in biomedicine* **172**, 103-116 (2019).

19 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).

20 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).

21 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).

22 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).

23 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).

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