Models

Model S1 Bayesian spatiotemporal model

We established a Bayesian spatiotemporal model to analyze the spatiotemporal pattern of TB incidence in 95 districts and counties in Jiangsu Province from 2011 to 2021. A logit-link Bayesian Beta regression model was used to analyze TB incidence [1], incorporating fixed effects associated with clinical factors, and separated spatial and temporal random effects. Specifically, let $y(s,t)$, denote TB incidence in year t over the district or county s with $t = 1, 2, ..., 12$ standing for the year 2011 to 2021, and $s = 1, ..., 95$ for the 95 counties/district in Jiangsu province, China. The incidence was defined as the number of reported cases per 1,000 residents, with values ranging from 0 to 1. Suppose that $y(s,t)$ follows a Beta law with mean $\mu(s,t) \in (0,1)$ varying in time across counties/districts and a constant precision parameter $\phi > 0$. Namely,

$$\text{logit}\,\mu(s,t) = \mathbf{x}^{\mathrm{T}}(\mathrm{s},\mathrm{t})\beta + \delta_t + u_s + \epsilon(s,t).$$

Here, the vector $\mathbf{x}(s,t)$ denotes the clinical factors (age, delay time, delay rate, confirmed rate, admission rate). The $\epsilon(s,t)$ item is an unstructured random effect in the model. We selected the first order auto-regressive (AR(1)) model and Besag-York-Mollie2 (BYM2) model to reveal the overall temporal random effect δ_t and spatial random effect u_s [2].

The auto-regressive (AR) model is a linear predictive modeling technique that uses previous observations to predict current ones by employing AR parameters as coefficients [3]. Specifically, current period values are modeled as a sum of past outcomes multiplied by a numeric factor. In this study, we employ an auto-regressive model of order 1 (denoted by AR(1)) for the random temporal effect δ_t, namely, we suppose that $\delta_t, t = 1,2, \ldots$, is a

stationary time series with mean zero, satisfying

$$\delta_t = a\delta_{t-1} + \varepsilon_t,$$

where ε_t is a normal white noise process with zero mean and constant variance, and a is a numeric constant by which we multiply the lagged variable δ_{t-1}, which is interpreted as the part of the previous value that remains in the present.

The Besag-York-Mollie2 (BYM2) model is considered here to account for the residual spatial random effect (u_s) such that geographically close areas have similar incidence rates. It is represented as [4]

$$u_s = \frac{1}{\sqrt{\tau_b}}\left(\sqrt{1-\phi}v_{*s} + \sqrt{\phi}u_{*s}\right),$$

where $\tau_b > 0$ is the precision parameter controlling the marginal variance contribution of the weighted sum of u_{*s} and v_{*s}. It intricately combines the scaled intrinsic conditional autoregressive (CAR) model u_{*s} with unit generalized variance, with unstructured (namely independent and identically distributed) spatial random effect v_{*s} satisfying $v_{*s} \sim N(0,1)$. The mixing parameter $\phi \in [0,1]$, is a spatial smoothing parameter, measuring the proportion of the marginal variance explained by the structured random effect.

We implemented the model in R version 4.3.1 and utilized Integrated Nested Laplace Approximation (INLA) to estimate the posterior distributions of the model parameters. The hyperparameters were assigned the default prior distributions of INLA.

Model S2 Space-time contribution variation partitioning

Let $x(s,t)$ be one of the socioeconomic status, meteorological conditions, air pollution, and geographical information in year t over county/district s. We took the spatial average of $x(\mathrm{s},\mathrm{t}), s \in \mathcal{S}$, denoted as $x(\cdot,t)$, smoothing out the spatial variation and thus remaining the temporal dynamics during the study period. Similarly, the time-averaged part $x(s,\cdot)$ remains only the spatial variation of the factor between counties/districts. This process yields a decomposition of $x(s,t)$ into three components: the temporal part $x(\cdot,t)$, the spatial part $x(s,\cdot)$, and the residual component, which is defined as $r(s,t) = x(s,t) - x(\cdot,t) - x(s,\cdot)$. Hence,

$$x(s,t) = x(\cdot,t) + x(s,\cdot) + r(s,t).$$

This decomposition is analogous to the two-factor analysis of variance [5]. Below, we developed an innovative approach to quantify the interpreted contribution of each term to the variability of TB incidence.

For a given covariate, we denote M_0 and M_{t+s+r} as the marginal log-likelihoods obtained from the Bayesian Beta regression models, which include the basic clinical factors and an additional term $x(s,t)$, respectively. A similar understanding applies for M_t, M_s and M_{t+s} with additional terms of temporal, spatial, and spatial and temporal terms accordingly. The proportion of variation explained by the temporal, spatial, and residual components is derived from the relative reduction in marginal log-likelihoods between nested models. Specifically, the improvement in model fit obtained by adding the temporal term alone ($M_t - M_0$) and on top of the spatial term ($M_{t+s} - M_s$) was averaged to represent the temporal contribution (ω_t). Similarly, the improvement from adding the spatial term alone ($M_s - M_0$)

and on top of the temporal term ($M_{t+s} - M_t$) was averaged to yield the spatial contribution (ω_s). The remaining improvement from the full model relative to the spatiotemporal model ($M_{t+s+r} - M_{t+s}$) was defined as the residual contribution (ω_r). Each contribution was normalized by the total improvement ($M_{t+s+r} - M_0$), ensuring that $\omega_t + \omega_s + \omega_r = 1$:

$$\begin{cases} \omega_t = \frac{(M_t - M_0) + (M_{t+s} - M_s)}{2 \times (M_{t+s+r} - M_0)}, \\ \omega_s = \frac{(M_s - M_0) + (M_{t+s} - M_t)}{2 \times (M_{t+s+r} - M_0)}, \\ \omega_r = \frac{M_{t+s+r} - M_{t+s}}{M_{t+s+r} - M_0}, \end{cases} \quad \text{(S.1)}$$

where $M_0, M_t, M_s, M_{t+s}, M_{t+s+r}$ represent the log-likelihoods obtained obtained from the corresponding Bayesian Beta regression models with the basic, temporal, spatial, spatiotemporal, and full specifications, respectively.

Model S3 XGBoost model

We used the robust Extreme Gradient Boosting (XGBoost) machine learning algorithm to identify influencing factors most strongly correlated with TB incidence [6]. In XGBoost, feature importance ranking was determined using three methods: gain, frequency, and cover [7]. Gain, which served as the primary indicator, evaluated the importance of a feature based on its impact on tree branching. Frequency represented the occurrence count of a feature across all constructed trees and offered a simplified version of the gain. Cover measured the relative representation of the observations of a feature. In our study, we used the gain metric to assess the feature importance, providing a detailed understanding of the contribution of influencing factors to TB incidence.

XGBoost employs an iterative approach that integrates multiple weak learning models to produce a strong learner [8]. The XGBoost algorithm utilizes decision trees to define the target function, represented as [6]:

$$\hat{y}_i = \sum_{m=1}^{M} f_m(x_i), f_m \in F,$$

where $\hat{y}_i$ represents the predictive value (TB incidence) associated with the i-th observation, x_i denotes the vector of input variables, M represents the number of trees, f_m represents each independent regression tree. Compared to the original gradient boosting framework proposed by Friedman [9], XGBoost introduces the regularized objective to the loss function. The regularized objective for optimizing the m-th iteration:

$$L^m = \sum_{i=1}^{n} l(y_i, \hat{y}_i^m) + \sum_{j=1}^{m} \Omega(f_j), \quad \text{(S.1)}$$

where n represents the number of samples, $l(\,)$ represents the differentiable loss function

that measures the difference between the predicted $\hat{y}_i{}^m$ and the target y_i, and $\Omega(\,)$ denotes the term that inhibits model complexity. Here,

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \parallel w\parallel^2,$$

where T refers to the number of a tree's leaf nodes, γ and λ refer to the regularisation parameters. The predicted $\hat{y}_i^m$ at the m-th iteration can be expressed as $\hat{y}_i^m = \hat{y}_i^{m-1} + f_m(x_i)$, then Equation S. 1 can be written as:

$$L^m = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{m-1} + f_m(x_i)\right) + \Omega\left(f_j\right) + \text{const},$$

where const represents $\sum_{j=1}^{m-1}\Omega\left(f_j\right)$. Since $\sum_{j=1}^{m-1}\Omega\left(f_j\right)$ has been determined by the tree of the $(m-1)$-th iteration, it can be taken as a constant. Then, by taking the Taylor expansion of the loss function to the second-order, the objective function can be written as follows:

$$L^m = \sum_{i=1}^{n}\left[l\left(y_i, \hat{y}_i^{m-1}\right) + g_i f_m(x_i) + \frac{1}{2}h_i f_m^2(x_i)\right] + \Omega(f_m),$$

where $g_i = \partial_{\hat{y}_i^{m-1}} l\left(y_i, \hat{y}_i^{m-1}\right)$ and $h_i = \partial^2_{\hat{y}_i^{m-1}} l\left(y_i, \hat{y}_i^{m-1}\right)$ are the first and the second-order derivatives on the loss function, respectively.

Trees are split into left and right nodes. I_L represents the sets of cases for the left node, while I_R represents the sets of cases for the right node. Letting $I = I_L \cup I_R$, then we can calculate the loss reduction after splitting the node as follows:

$$\text{Gain}=\frac{1}{2}\left[\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i+\lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i+\lambda} - \frac{\left(\sum_{i\in I} g_i\right)^2}{\sum_{i\in I} h_i+\lambda}\right] - \gamma.$$

Define K as a random variable denoting the feature responsible for infinitesimal units of improvement in accuracy in model performance, where $K = 1, 2, \ldots, 11$ correspond to the eleven features considered in the model. The probability $q_i = P(K = i)$ is defined as the gain

to the i-th feature, which reflects its share of the total importance.

In parallel, define D be a random variable denoting the component associated with that unit of improvement, taking values s, t, and r, which correspond to the contributions of the spatial, temporal, and residual components, respectively. The total contribution can be decomposed as:

$$\sum_{i=1}^{11} P(K=i) = \sum_{i=1}^{11} \sum_{d\in\{s,t,r\}} P(K=i, D=d) = \sum_{i=1}^{11} \sum_{d\in\{s,t,r\}} P(D=d|K=i)P(K=i)$$
$$= \sum_{i=1}^{11} \sum_{d} \omega_{id} q_i = 1,$$

where $\omega_{id} = P(D=d|K=i)$ corresponds to the proportion of decomposition attributable to d-th component of i-th feature, and the product $\omega_{id} q_i = P(K=i, D=d)$ therefore represents the joint contribution of this component to the overall model improvement, integrating gain-based importance with decomposition proportion.

Supplementary Tables and Figures

Table S1: Descriptive statistics of both response and explanatory variables during 2011—2021, Jiangsu province, China.

Factors	Variables	Unit	Mean	(Min, Max)
	TB Incidence	per thousand people	0.39	(0.14, 0.79)
Clinical	Age	year	50	(33, 66)
	Delay Time	day	42	(7, 359)
	Delay Rate	%	37.22	(4.45, 90.83)
	Confirmed Rate	%	38.8	(10.56, 73.95)
	Admission Rate	%	91.34	(79.67, 98.90)
Socioeconomic	GDP	thousand yuan	781.6	(31.10, 4748.10)
	Health Technician	per thousand people	7.19	(0.67, 58.39)
Meteorological	Wind Speed	m/s	0.39	(0.01, 1.56)
	Precipitation	mm	1.05	(0.02, 5.22)
	Temperature	℃	15.93	(15.38, 17.70)
	Relative Humidity	%	71.19	(49.54, 83.89)
Air pollution	PM_1	$\mu g/m^3$	29.73	(13.22, 48.24)
	$PM_{2.5}$	$\mu g/m^3$	53.21	(25.20,89.98)
	PM_{10}	$\mu g/m^3$	88.93	(45.03, 148.59)
	O_3	$\mu g/m^3$	101.64	(76.36, 121.97)
Geographical	NDVI	/	0.46	(0.01, 0.64)

In the following, we describe the explanatory variables in detail.

Clinical factor. We integrate TB individual data from the Jiangsu Provincial CDC by year and county/district to obtain the following indicators: (1) age, representing the average onset age of patients, (2) admission rate, signifying the proportion of patients admitted to the hospital, (3) confirmed rate refers to the proportion of patients who have been confirmed with TB among all reported TB cases, and the confirmed cases typically include those diagnosed with smear-positive TB, Polymerase Chain Reaction (PCR)-positive TB, and other variants of TB determined through pathological examination of lung lesion specimens, (4) delay time indicates the average number of days patients delay treatment, and (5) delay rate denotes the proportion of patients with diagnostic delay. Here, we define a delay as the interval between the day of onset and the day of diagnosis exceeding 28 days, following the previous studies on the delay of consultation [10] and the delay of diagnosis (as the number of days from

consultation to diagnosis more than 14 days [11]).

Socioeconomic status. The socioeconomic status at the county/district level are collected from the Statistical Yearbooks of the thirteen prefecture-level cities in Jiangsu, including healthcare technical personnel per thousand people, the number of healthcare technical personnel divided by the total population in thousandth, GDP per capita, and the GDP (adjusted with the Consumer Price Index in 2021) divided by the total permanent population.

Air pollution. The average annual data of particulate matter $\leqslant$ 2.5 μm in aerodynamic diameter ($PM_{2.5}$), $\leqslant$ 10 μm (PM_{10}), $\leqslant$ 1 μm (PM_1), and ozone (O_3) are extracted from the China High Air Pollutants (CHAP) dataset [12-14]. The spatial resolution is 1 km for $PM_{2.5}$, PM_{10}, PM_1, and 10 km for O_3.

Meteorological conditions. We obtain the data from the ERA5-Land Dataset and download monthly concentrations of four meteorological factors, including temperature, precipitation, wind speed, and relative humidity [15]. All the meteorological factors are measured at a spatial resolution of 9 km.

Geographical information. We use the 8 km spatial resolution Normalized Difference Vegetation Index (NDVI) dataset from the National Earth System Science Data Center of China, a standardized index in remote sensing for vegetation monitoring and analysis [16].

153 **Table S2**: Bayesian estimation of clinical factors of TB risk in Jiangsu province, 2011—2021.

Variables	Posterior Mean (95% CI)	OR (95% CI)
Age	0.030 (0.022 to 0.037)	1.030 (1.022 to 1.037)
Delay time	0.001 (0.000 to 0.002)	1.001 (1.000 to 1.002)
Delay rate	−0.003 (−0.005 to −0.001)	0.997 (0.995 to 0.999)
Admission rate	−0.004 (−0.012 to 0.005)	0.996 (0.982 to 1.005)
Confirmed rate	−0.015 (−0.017 to −0.012)	0.985 (0.983 to 0.988)

154

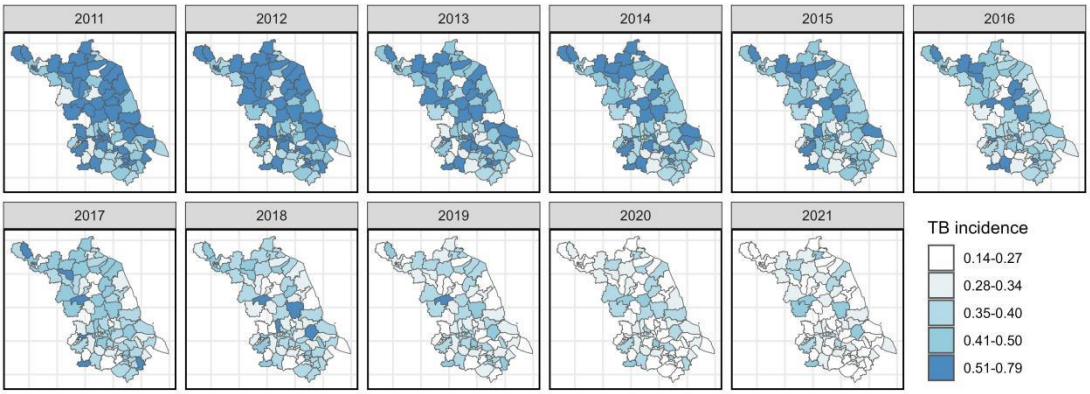
Table S3: Hyperparameters and model performances for XGBoost models.

Model	Hyperparameters	RMSE	R^2
With all factors included	nrounds: 100 max_depth: 5 eta: 0.05 gamma: 0.05 colsample_bytree: 0.7 min_child_weight: 5 subsample: 0.5	0.0865	0.5572
With external indirect factors included	nrounds: 100 max_depth: 5 eta: 0.01 gamma: 0.05 colsample_bytree: 0.6 min_child_weight: 5 subsample: 0.8	0.1087	0.4736

Note: XGBoost model was built by splitting the data into a 70% training set and a 30% testing set, using a 10-fold cross-validation grid search to optimize parameters. External indirect factors included socioeconomic, meteorological, air pollution, and geographical factors.

Table S4: Relative importance of external indirect factors based on the partitioned contributions of temporal, spatial, and residual components to variations in TB incidence.

Factors	Variables	Time (%)	Space (%)	Remaining (%)	Relative Importance (%)
Socioeconomic	GDP	3.09	0.55	2.69	6.33
	Health Technicians	5.47	3.52	7.81	16.80
	Total	8.56	4.07	10.50	23.13
Meteorological	Wind Speed	0.87	1.15	0.89	2.91
	Precipitation	0.69	0.83	0.88	2.40
	Temperature	1.63	2.97	2.41	7.01
	Humidity	1.54	1.59	1.27	4.40
	Total	4.73	6.53	5.54	16.71
Air Pollution	PM_1	1.32	0.95	0.84	3.11
	$PM_{2.5}$	1.19	0.88	0.97	3.04
	PM_{10}	3.67	2.45	3.07	9.19
	O_3	16.86	11.17	13.29	41.32
	Total	23.04	15.45	18.17	56.66
Geographical	NDVI	0.79	1.69	1.04	3.52



164

165 **Figure S1**: TB incidence for the spatiotemporal analysis in Jiangsu province, 2011—2021.

166

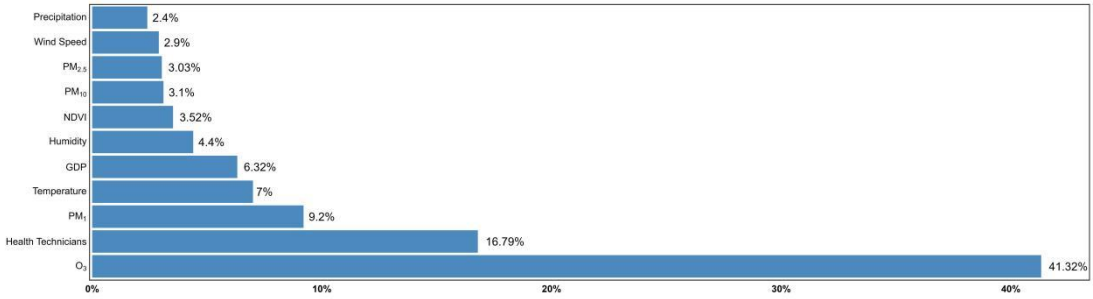


Figure S2: Relative importance of exposure to socioeconomic status, meteorological conditions, air pollution, and geographical information in explaining TB incidence using the XGBoost model.

References

1. Fabrizi E, Ferrante MR, Trivisano C. Bayesian Beta regression models for the estimation of poverty and inequality parameters in small areas. In: Pratesi M, editor. Analysis of Poverty Data by Small Area Estimation. Hoboken (NJ): Wiley; 2016. Chapter 16. doi: 10.1002/9781118814963.ch16.

2. Amsalu E, Liu M, Li Q, Wang X, Tao L, Liu X, et al. Spatial-temporal analysis of tuberculosis in the geriatric population of China: an analysis based on the Bayesian conditional autoregressive model. Arch Gerontol Geriatr, 2019; 83, 328−337. doi: 10.1016/j.archger.2019.05.011.

3. Mao Q, Zhang K, Yan W, Cheng C. Forecasting the incidence of tuberculosis in China using the seasonal autoregressive integrated moving average (SARIMA) model. J Infect Public Health, 2018; 11(5), 707−712. doi: 10.1016/j.jiph.2018.04.009.

4. Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH. Penalising model component complexity: a principled, practical approach to constructing priors. Statist Sci, 2017; 32(1), 1−28. doi: 10.1214/16-STS576.

5. Berrios J, Mosher E, Benzo S, Grajeda C, Baggili I. Factorizing 2FA: forensic analysis of two-factor authentication applications. Forensic Sci Int Digit Investig, 2023; 45(Suppl), 301569. doi: 10.1016/j.fsidi.2023.301569.

6. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min. New York (NY): ACM Press; 2016. p. 785−794.

7. Ma L, Fu T, Blaschke T, Li M, Tiede D, Zhou Z, et al. Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. ISPRS Int J Geo-Inf, 2017; 6(2), 51. doi: 10.3390/ijgi6020051.

8. Piraei R, Afzali SH, Niazkar M. Assessment of XGBoost to estimate total sediment loads in rivers. Water Resour Manage, 2023; 37, 5289−5306. doi: 10.1007/s11269-023-03606-w.

9. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat, 2001; 29(5), 1189−1232.

10. Fu L, Wang Y, Zhu W, Wang W. Consultation delay and influencing factors among pulmonary tuberculosis patients in Huzhou City from 2008 to 2018. Chin J Dis Control Prev, 2021; 25(2), 235−239. doi: 10.16462 /j.cnki.zhjbkz.2021.02.022.

11. Yang Q, Tong Y, Yin X, et al. Delays in care seeking, diagnosis and treatment of patients with pulmonary tuberculosis in Hubei, China. Int Health, 2020; 12(2), 101−106.

doi: 10.1093/inthealth/ihz036.

12. Wei J, Huang W, Li Z, Xue W, Peng Y, Sun L, et al. Estimating 1-km-resolution $PM_{2.5}$ concentrations across China using the space-time random forest approach. Remote Sens Environ, 2019; 231, 111221. doi: 10.1016/j.rse.2019.111221.

13. Wei J, Li Z, Li K, Dickerson RR, Pinker RT, Wang J, et al. Full-coverage mapping and spatiotemporal variations of ground-level ozone (O_3) pollution from 2013 to 2020 across China. Remote Sens Environ, 2022; 270, 112775. doi: 10.1016/j.rse.2021.112775.

14. Wei J, Li Z, Xue W, Sun L, Fan T, Liu L, et al. The ChinaHighPM_{10} dataset: generation, validation, and spatiotemporal variations from 2015 to 2019 across China. Environ Int, 2021; 146, 106290. doi: 10.1016/j.envint.2020.106290.

15. Muñoz-Sabater J, Dutra E, Agustí-Panareda A, Albergel C, Arduini G, Balsamo G, et al. ERA5-land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst Sci Data, 2021; 13(9), 4349−4383. doi: 10.5194/essd-13-4349-2021.

16. Matas-Granados L, Pizarro M, Cayuela L, Domingo D, Gómez D, García MB. Long-term monitoring of NDVI changes by remote sensing to assess the vulnerability of threatened plants. Biol Conserv, 2022; 265, 109428. doi: 10.1016/j.biocon.2021.109428.