Exploring the role of meteorological factors in predicting incident pulmonary tuberculosis: A time-series study in eastern China

Zhongqi Li  
Nanjing Medical University  https://orcid.org/0000-0001-8152-5033

Hongqiu Pan  
the Third Hospital of Zhenjiang City

Qiao Liu  
Nanjing Medical University

Huan Song  
Nanjing Medical University

Jianming Wang  (jmwang@njmu.edu.cn)  
Nanjing Medical University

Research Article

Keywords: Pulmonary tuberculosis, Meteorological factors, Time series, Predicting

Posted Date: August 18th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-60497/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published on November 5th, 2020. See the published version at https://doi.org/10.1186/s40249-020-00771-7.
Exploring the role of meteorological factors in predicting incident pulmonary tuberculosis: A time-series study in eastern China

Zhongqi Li\textsuperscript{1†}, Hongqiu Pan\textsuperscript{2†}, Qiao Liu\textsuperscript{1†}, Huan Song\textsuperscript{1}, Jianming Wang\textsuperscript{1*}

1Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, 211166, People’s Republic of China

2Department of Tuberculosis, the Third Hospital of Zhenjiang City, Zhenjiang, 212005 China

†These authors contributed equally to this work.

Correspondence: Jianming Wang. Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, 101 Longmian Ave. Nanjing, 211166, China. Tel: +86-25-86868438. Email: jmwang@njmu.edu.cn
Abstract

Background: Many studies have compared the performance of time-series models in predicting pulmonary tuberculosis (PTB). Few studies regarding the role of meteorological factors in predicting PTB are available. This study aims to explore whether incorporating meteorological factors can improve the performance of time series models in predicting pulmonary tuberculosis (PTB).

Methods: We collected the monthly number of PTB cases registered in three cities of China from 2005 to 2018, and data of six meteorological factors in the same period. We constructed three time-series models, including the autoregressive integrated moving average (ARIMA) model, the ARIMA with exogenous variables (ARIMAX) model, and the recurrent neural network (RNN) model. The construction of the ARIMA model did not incorporate meteorological factors, while the construction of ARIMAX and RNN models incorporated meteorological factors. The mean absolute percentage error (MAPE) and root mean square error (RMSE) were used to compare the performance of models in predicting the monthly number of PTB cases in 2018.

Results: Both the cross-correlation analysis and spearman rank correlation test showed that PTB was related to meteorological factors in the three cities. The prediction performance of both ARIMA and RNN models was improved after incorporating the meteorological factors. The MAPEs of the ARIMA, ARIMAX, and RNN models were 12.536%, 11.957%, and 12.360% in Xuzhou, 15.568%, 11.155%, and 14.087% in Nantong, and 9.700%, 9.660% and 12.501% in Wuxi, respectively. The RMSEs of the three models were 36.194, 33.956 and 34.785 in Xuzhou, 34.073, 25.884 and 31.828 in Nantong, and 19.545, 19.026 and 26.019 in Wuxi, respectively.
Conclusions: Our study revealed a possible link between PTB and meteorological factors. Taking meteorological factors into consideration may increase the accuracy of time series models in predicting PTB.

Key words: Pulmonary tuberculosis; Meteorological factors; Time series; Predicting
Background

Tuberculosis (TB) is a chronic communicable disease that severely threatens human health, ranking among the top ten causes of death worldwide. The World Health Organization (WHO) estimated that approximately 10.0 million people fell ill with TB around the world in 2018. Meanwhile, there were an estimated 1.2 million TB deaths among HIV-negative people and 251,000 TB deaths among HIV-positive people. In order to curb the TB epidemic, the WHO set a goal of reducing the morbidity and mortality of TB by 90% and 95%, respectively, between 2015 and 2035[1]. Accurately predicting the trend of the epidemic can help foresee the possible peaks and provide a reference for the prevention and control of TB.

A time series is formed by recording the development process of a random event in the order of time. Time series analysis plays a vital role in predicting trends by identifying the law of health-related events changing with time. The autoregressive integrated moving average (ARIMA) model is the most classic time series analysis model and has been widely applied to predict various infectious diseases like hepatitis B[2], hemorrhagic fever with renal syndrome[3], coronavirus disease 2019[4] and hand, foot and mouth disease[5]. The ARIMA with exogenous variables (ARIMAX) exhibits superior prediction performance by adding other event-related factors as input variables[6, 7]. Another commonly used time series analysis model is the artificial neural network (ANN), which is designed to simulate the way the human brain analyzes and processes information. It has been applied to construct time series models to forecast human diseases[8-10]. The recurrent neural network (RNN) is a specific ANN with the ability to transfer information across time steps, as it can remember the previous information and apply it to the current output calculation. The
ability to model temporal dependencies makes it particularly appropriate to analyze
time series, which consists of a sequence of points that are not independent[11, 12].

Time series analysis has been used to predict TB morbidity or mortality, but most
were conducted in one city or one region and based on one or two models without
incorporating meteorological factors in the model[13-16]. Our previous studies have
observed that the incidence of TB exhibited seasonal fluctuations, indicating a
potential relationship with meteorological factors[17]. Thus, in the current study, we
performed a time series analysis in the three cities of Jiangsu Province, China, and
applied three types of models, including ARIMA, ARIMAX, and RNN, to explore
whether the inclusion of meteorological factors can improve the performance of
modeling.

Methods

Study area
Jiangsu Province is located on the eastern coast of China, with an area of 107,200
square kilometers. It governs 13 cities and has a permanent population of 80.7 million
at the end of 2019. We randomly selected one city from northern, central, and
southern Jiangsu, respectively, and finally included Xuzhou, Nantong, and Wuxi as
the study sites.

Data collection
All newly diagnosed TB cases in China are registered in an online surveillance system
operated by the Center for Disease Control and Prevention. We extracted the monthly
reported number of pulmonary TB (PTB) cases in the study sites between 2005 and
2018. We also collected meteorological factors in the same area at the same time from the China Meteorological Data Network (http://www.nmic.cn/). These meteorological factors included monthly average temperature (MAT), monthly average atmospheric pressure (MAP), monthly average wind speed (MAS), monthly average relative humidity (MAH), monthly precipitation (MP), and monthly sunshine time (MST).

Construction of the ARIMA model

As described in our previous studies[17], we constructed a seasonal ARIMA model, which is expressed as ARIMA \((p, d, q)(P, D, Q)s\). The \(p\), \(d\), and \(q\) represent the autoregressive order, the number of ordinary differences, and the moving average order, respectively. The \(P\), \(D\), and \(Q\) represent the seasonal autoregressive order, the number of seasonal differences, and the seasonal moving average order, respectively. The \(s\) represents the length of a periodic pattern (\(s = 12\) in this study). The number of PTB cases predicted at time \(t\) (\(Y_t\)) was determined by the formula:

\[
Y_t = \frac{\theta_q(B)\Theta_Q(B^s)a_t}{\Phi_P(B^s)\Phi_P(B)(1-B)^d(1-B^s)^D},
\]

where \(\theta_q(B)\) is the operator of moving average, \(\Theta_Q(B^s)\) is the operator of seasonal moving average, \(\phi_p(B)\) is the operator of autoregressive, \(\Phi_P(B^s)\) is the operator of seasonal autoregressive, \((1 - B)^d\) is the component of ordinary difference, \((1 - B^s)^D\) is the component of seasonal difference, \(a_t\) is the white noise and \(Y_t\) is the predicted variable[18, 19]. Based on the monthly number of PTB, we constructed the ARIMA model for each city. First, we applied the ordinary difference and seasonal difference to make the series stationary. Second, by referring to the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the stationary series, we initially identified the values of the parameters (\(p\), \(q\), \(P\) and \(Q\)) to establish alternative ARIMA models. Third, we determined the optimal ARIMA model according to three criteria: (a) the
smaller the normalized Bayesian information criterion (BIC) value of the model was, the better the model would be; (b) the residual series of the model was proved to be a white noise by the Ljung-Box test; (c) the parameter estimation showed that the parameters of the model were significant[17]. Finally, we selected the optimal ARIMA model to predict PTB cases in 2018.

Construction of the ARIMAX model

The ARIMAX model adds exogenous variables based on ARIMA model, which can be described by the formula: 

$$ Y_t = \frac{\theta_q(B)\Theta_Q(B^s)\alpha_t}{\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D} + X, $$

where $X$ represents the external regressor, which can be univariate or multivariate. The other parameters are consistent with the ARIMA model[18]. Based on the monthly number of PTB cases and six meteorological factors, we constructed the ARIMAX model for each city. First, we constructed the optimal ARIMA model for each meteorological factor, obtained residual series of optimal ARIMA models, and ensured that they were all white noise. Second, we used the cross-correlation function (CCF) to analyze the residual series of PTB cases and meteorological factors to evaluate the correlation between them at different lag times. Third, we tried different combinations of significant meteorological factors as external variables into the optimal ARIMA model to construct alternative ARIMAX models. Finally, we determined the optimal ARIMAX model according to three criteria: (a) the normalized BIC value was smaller than the optimal one; (b) the residual series of the model was proved to be a white noise by the Ljung-Box test; (c) the performance of the model in predicting PTB cases in 2018.
Construction of the RNN model

The ANN usually consists of the input layer, hidden layer, and output layer. The layers of traditional ANN are fully connected, but the neurons in each layer are not connected. The difference of RNN is that it adds the connection between the neurons in the hidden layer (Figure 1A). Figure 1B shows the unfolding diagram of the forward propagation of the RNN[20], where $x_t$ represents the input at time $t$, $h_t$ represents the hidden state at time $t$, $h_t = \text{sigmoid}(W \ast h_{t-1} + U \ast x_t)$, $W$ represents the weight of the input, $U$ represents the weight of the input at the moment, $y_t$ represents the output at time $t$, $y_t = \text{softmax}(V \ast h_t)$, and $V$ represents the weight of the output. Therefore, the input of the hidden layer of the RNN includes not only the output of the input layer but also the previous output of the hidden layer, making the model have memory function. We divided data into the training set, testing set, and predicting set. We trained each RNN model for three times and compared their performance on the testing set to determine the optimal RNN model. For each RNN model, we set the learning rate as 0.05, 0.1, and 0.2, and the dimensions of the hidden layer as 3, 5, and 10, respectively, and identified the appropriate training epochs through the epoch-error plot. By comparing the performance of the model on the testing set, we determined the most suitable parameters for each RNN model. First, we normalized the original data to convert all values to intervals [0, 1], using the formula: $X' = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$, where $X$ is the original value, $X_{\text{max}}$ is the maximum value of original data, $X_{\text{min}}$ is the minimum value of original data, and $X'$ is the normalized value after conversion. Second, we used the number of PTB cases in the previous one, two, three, six, and twelve months as the input of the training set, respectively and the number of PTB cases in the current month as the output of the training set to construct five different RNN models
(RNN1-RNN5), which did not incorporate meteorological factors. We compared the
performance of five RNN models on the testing set and selected the best one to
incorporate meteorological factors into it. Third, we used the Spearman rank
correlation test to evaluate the correlation between PTB cases in the current month
and meteorological factors one, two, and three months ago, respectively. Fourth, we
incorporated the significant meteorological factors into the best model of
RNN1-RNN5 to construct another four RNN models (RNN6-RNN9). Finally, we
compared the performance of nine RNN models on the testing set to determine the
optimal one and applied it to predict PTB cases in 2018.

Evaluating the performance of three models

The diagnostic indicators, including mean absolute percentage error (MAPE) and root
mean square error (RMSE), were used to evaluate the performance of the three
models: \[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|X_i - \hat{X}_i| \times 100}{X_i}, \quad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{X}_i)^2},
\] where $X_i$ is the actual value at time $i$, $\hat{X}_i$ is the output value of the model at time $i$ and $n$ is the
number of samples.

Statistical software

We used SPSS 25.0 (Nanjing Medical University, Nanjing, China) to construct the
ARIMA and ARIMAX models, and the package of “rnn” in R 3.6.3 (https://www.r-project.org/) to construct the RNN model. The significant level was set
at 0.05.
Results

The ARIMA model

The monthly number of PTB cases showed a long-term downward trend and seasonal fluctuations, with the peak in March to April, and the trough in December to January (in Xuzhou) or January to February (in Nantong and Wuxi) (Figure 2). Therefore, we applied one ordinary difference and one seasonal difference to make the series stationary \((d = D =1)\). Then, we initially identified the parameters of the ARIMA model \((p, q, P, \text{and } Q)\) to construct alternative models for each city according to the ACF and PACF plots of the stationary series (Figure S1, A1-A3, and B1-B3). We determined the optimal ARIMA model as ARIMA \((1,1,1)(0,1,1)_{12}\) for Xuzhou, and ARIMA \((0,1,1)(0,1,1)_{12}\) for Nantong and Wuxi, since they had the smallest normalized BIC, residual series were proved to be white noise, and parameters were all significant \((P <0.05)\) (Table S1, C1-C3, and D1-D3 of Figure S1). PTB cases in 2018 were predicted by the optimal ARIMA model and listed in Table 1.

The ARIMAX model

The time series plots of six meteorological factors in three cities between 2005 and 2017 are shown in Figure S2. The optimal ARIMA models of MAT, MAP, MAS, MAH, MP and MST were ARIMA \((0,0,0)(0,1,1)_{12}\), ARIMA \((0,0,0)(0,1,1)_{12}\), ARIMA \((1,0,0)(2,1,0)_{12}\), ARIMA \((1,0,1)(0,1,1)_{12}\), ARIMA \((0,0,1)(0,1,1)_{12}\), ARIMA \((1,1,1)(1,1,0)_{12}\), ARIMA \((1,1,1)(1,1,1)_{12}\), ARIMA \((0,0,0)(0,1,1)_{12}\) and ARIMA \((1,0,1)(0,1,1)_{12}\) for Xuzhou, ARIMA \((1,0,1)(0,1,1)_{12}\), ARIMA \((0,0,1)(0,1,1)_{12}\), ARIMA \((1,0,1)(0,1,1)_{12}\), ARIMA \((0,0,1)(0,1,1)_{12}\), ARIMA \((0,0,0)(2,1,0)_{12}\), ARIMA \((0,0,1)(0,1,1)_{12}\), ARIMA \((0,1,1)(0,1,1)_{12}\), ARIMA \((0,1,2)(0,1,1)_{12}\), ARIMA \((0,0,0)(0,1,1)_{12}\) and ARIMA \((1,1,1)(0,1,1)_{12}\) for Nantong, and ARIMA \((1,0,1)(0,1,1)_{12}\), ARIMA \((0,0,1)(0,1,1)_{12}\), ARIMA \((0,0,0)(0,1,1)_{12}\) and ARIMA \((1,1,1)(0,1,1)_{12}\) for Wuxi. We then estimated the correlation between PTB
and each meteorological factor at different lag times. The CCF plots showed that PTB was positively correlated with MAS (2 months lag), MAH (1 month lag) and MP (2 months lag), and negatively correlated with MST (1 month lag) in Xuzhou, PTB was positively correlated with MAT (0 month lag), MAP (1 month lag) and MAS (2 months lag) in Nantong, and PTB was positively correlated with MST (0 month lag), and negatively correlated with MAH (0 month lag) \((P < 0.05)\) in Wuxi (Figure 3). We incorporated different combinations of significant meteorological factors as external variables into the optimal ARIMA model to construct alternative ARIMAX models (Table 2). Finally, we determined the optimal ARIMAX model as ARIMA \((1,1,1)(0,1,1)_{12}\) with MP (2 months lag) for Xuzhou, ARIMA \((0,1,1)(0,1,1)_{12}\) with MAP (1 month lag) for Nantong and ARIMA \((0,1,1)(0,1,1)_{12}\) with MAH (0 month lag) for Wuxi. PTB cases in 2018 were predicted by the optimal ARIMAX model and listed in Table 1.

The RNN model

We compared the MAPE of each RNN model with different parameters on the testing set to identify the appropriate parameters. The RNN5 model had the smallest MAPE on the testing set in each city (Table 3). The number of PTB cases in the current month in Xuzhou has positively correlated with MAS one month ago \((P < 0.01)\), MAS two months ago \((P < 0.01)\) and MAS three months ago \((P < 0.01)\), and negatively correlated with MST two months ago \((P < 0.01)\), MAT three months ago \((P < 0.01)\), MP three months ago \((P < 0.05)\) and MST three months ago \((P < 0.01)\). The number of PTB cases in the current month in Nantong has negatively correlated with MAS one month ago \((P < 0.05)\), MAH one month ago \((P < 0.05)\), MAS two months ago \((P < 0.01)\), MAH two months ago \((P < 0.01)\), MAS three months ago \((P < 0.01)\) and MAH...
three months ago (P <0.05). The number of PTB cases in the current month in Wuxi
has positively correlated with MAT one month ago (P <0.01), MAS one month ago (P
<0.01), MST one month ago (P <0.05), MAS two months ago (P <0.01) and MAS
three months ago (P <0.01), and negatively correlated with MAP one month ago (P
<0.01), MAH one month ago (P <0.05), MAT three months ago (P <0.05) and MAH
three months ago (P <0.05) (Table S2). Then, we constructed the RNN6-RNN9
models by incorporating significant meteorological factors into the RNN5 model. The
detailed composition of the nine RNN models is listed in Table S3. We determined the
optimal RNN model as RNN8 for Xuzhou, and RNN7 for Nantong and Wuxi, since
they had the smallest MAPE on the testing set after three times training (Table 3).
Figure S3 showed the epoch-error plots of optimal RNN models after three times of
training. The downward trend of errors of models on the training set was no longer
significant after reaching the set number of epochs, indicating that the training epochs
were appropriate. Finally, we chose the RNN8 model after the first training in Xuzhou,
and RNN7 model after the second training in Nantong and Wuxi (Table 3). PTB cases
in 2018 were predicted by the optimal RNN model and listed in Table 1.

Evaluating the performance of three models

As shown in Table 4, the ARIMAX model is slightly superior to ARIMA and RNN in
Xuzhou, the ARIMAX model is significantly superior to ARIMA and RNN in
Nantong, and the ARIMAX model is slightly superior to ARIMA and significantly
superior to RNN in Wuxi. Generally speaking, the ARIMAX model showed the best
performance.
Discussion

In this study, we explored the role of meteorological factors in predicting PTB in three cities of China by constructing the ARIMA, ARIMAX, and RNN models. The prediction ability of the model has been improved by adding meteorological factors. The ARIMAX model (ARIMA with meteorological factors) showed the best performance. To our knowledge, this is the first time-series study to construct different models in different cities to explore the role of meteorological factors in predicting PTB.

Although the notification rate of TB has declined with an annual rate of 3% between 2005 and 2017[13], China still had about 866,000 new cases notified in 2018, second only to India[1]. Accurately forecasting the future trend of TB epidemic can help policymakers to implement effective interventions and distribute healthcare resources appropriately. Previous studies have explored models, such as ARIMA[13, 16], X12-ARIMA[16], and ARIMA-generalized regression neural network (GRNN), in predicting TB[13]. However, few models considered seasonal variation characteristics, socioeconomic levels, and meteorological factors[15, 21-23]. Therefore, we divided the study areas into three regions according to the geographical location and economic level and then compared the performance of different models with or without adding meteorological factors in predicting PTB in the Chinese population.

The ARIMA model, also known as the Box-Jenkins model[24], can analyze various types of time series data and is a commonly used model in time series analysis[2-5]. Different from the ARIMA model, which is a univariate time series model, the ARIMAX model can deal with multivariate time series data. It adds other variables
related to the target series as input variables to improve the prediction accuracy. A
time-series study in Guangzhou of China showed that the ARIMA model with the
imported cases and the minimum temperature as input variables was superior to a
single ARIMA model in forecasting dengue transmission[18]. Another time-series
study in Abidjan of Cote d’Ivoire also indicated that including rainfall as an input
variable can increase the accuracy of the ARIMA model in predicting influenza[25].
However, when we incorporated two or more meteorological factors into the ARIMA
model, its prediction performance did not continuously increase, which may be
attributed to the high collinearity between these meteorological factors.

Considering that both ARIMA and ARIMAX were linear regression models, we
applied the RNN model, which has a strong nonlinear fitting ability. It can recognize
the relationship between variables without any restrictions and has a memory function.
This memory function makes the RNN model not only takes the current data as input
but also applies its long-term experience as input[12]. When constructing the RNN
model, some parameters need to be determined artificially. Besides, since the initial
weights and thresholds are random when training the RNN model, even for the same
training set, the output of the model on the testing set will not be precisely the same.
Therefore, we trained each RNN model with different parameters and compositions
three times and compared their performance on the testing set to determine the
optimal RNN model. Finally, we found that the prediction performance of the RNN
model was improved after incorporating meteorological factors.

The possible link between PTB and meteorological factors may be attributed to the
following reasons. First, the temperature can affect indoor and outdoor activities of
TB patients and susceptible people. For example, in hot summer and cold winter, people tend to stay indoors, which will increase the probability of mycobacterium tuberculosis (M.tb) transmission[26]. Second, high wind speed can dilute the concentration of environmental M.tb, thereby reducing the risk of infection. Airflow is usually formed from high-pressure areas to low-pressure areas, so the correlation between PTB and atmospheric pressure may be related to wind speed, but further exploration is needed[27]. Third, high relative humidity and abundant precipitation can provide an appropriate living environment for M.tb[27, 28]. Continuous exposure to dry air may decrease the production of protective mucus on the respiratory tract surface, thereby weakening its resistance to the pathogen[29]. Fourth, the large amount of ultraviolet light provided by long-term sunshine not only restricts the growth of M.tb but also promotes the synthesis of vitamin D, which can protect people from TB to some extent[27].

The association between PTB and meteorological factors varied across regions[27], which may be partially attributed to socioeconomic differences or analytic methods. TB is a poverty-related infectious disease[1]. Different economic levels may lead to an uneven distribution of socioeconomic factors that affect the risk of TB, such as food and nutrition security, living conditions, community environment, and medical resources[23, 30]. The inconsistency between analytical methods may be due to their different requirements for the data. The Spearman rank correlation test has no special requirements for the distribution of variables and has a wide range of applications. However, if there is a long-term trend in both time series, it will lead to a biased correlation. The cross-correlation analysis can evaluate the correlation between time series at different lag times without the influence of long-term trends. Besides, the
exposure-response relationship between TB and meteorological factors might be nonlinear. For example, as mentioned earlier, TB may benefit from extremely high or extremely low temperatures and relative humidity. Both the Spearman rank correlation test and the cross-correlation analysis can perform linear correlation analyses between time series but have limitations in solving nonlinear relationships.

Our study has several limitations. First, the ARIMA, ARIMAX, and RNN models are all short-term prediction models; continuous data collection to update models is essential for maintaining their prediction performance. Second, we incorporated all combinations of significant meteorological factors into the ARIMA model to construct the ARIMAX model, but we only incorporated four combinations of meteorological factors into the RNN model. Although the performance of the RNN model in this study is inferior to the ARIMAX model, its prediction performance needs further exploration. Third, we just qualitatively evaluated the linear correlation between PTB and meteorological factors based on monthly data. Considering that this relationship may be nonlinear and has lag time, we intend to apply the distributed lag nonlinear model to quantitatively evaluate it based on weekly or daily data in the future study.

Conclusions

The prediction performance of both ARIMA and RNN model was improved after incorporating meteorological factors, and the ARIMAX model (ARIMA with meteorological factors) had the best performance, indicating a potential link between PTB and meteorological factors. Taking meteorological factors into consideration may increase the accuracy of time series models in predicting the trend of PTB.
Abbreviations

TB: tuberculosis; PTB: pulmonary tuberculosis; M.tb: mycobacterium tuberculosis;
ARIMA: autoregressive integrated moving average; ARIMAX: autoregressive
integrated moving average with exogenous variables; ANN: artificial neural network;
RNN: recurrent neural network; MAPE: mean absolute percentage error; RMSE: root
mean square error; MAT: monthly average temperature; MAP: monthly average
atmospheric pressure; MAS: monthly average wind speed; MAH: monthly average
relative humidity; MP: monthly precipitation; MST: monthly sunshine time; ACF:
autocorrelation function; PACF: partial autocorrelation function; BIC: Bayesian
information criterion; CCF: cross-correlation function.

Acknowledgements

Not applicable.

Authors’ contributions

ZL, HP, and JW conceived, initiated, and led the study. ZL, HP, QL and HS collected
the data. ZL, HP, QL, and JW analyzed the data with input from all the authors. ZL
and JW prepared the manuscript. All authors reviewed and approved the manuscript.

Funding

This study was funded by the National Natural Science Foundation of China
Availability of data and materials
Please contact author for data requests.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no conflict of interest.

Author details
1Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, 211166, People’s Republic of China.

2Department of Tuberculosis, the Third Hospital of Zhenjiang City, Zhenjiang,
416 References

1. WHO: Global Tuberculosis Report 2019. In. Geneva: World Health Organization; 2019.
2. Wang YW, Shen ZZ, Jiang Y: Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China. *PloS one* 2018, 13(9):e0201987.
3. Liu Q, Liu X, Jiang B, Yang W: Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. *BMC infectious diseases* 2011, 11:218.
4. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M: Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in brief* 2020, 29:105340.
5. Liu L, Luan RS, Yin F, Zhu XP, Lu Q: Predicting the incidence of hand, foot and mouth disease in Sichuan province, China using the ARIMA model. *Epidemiology and infection* 2016, 144(1):144-151.
6. Zhang X, Zhang T, Pei J, Liu Y, Li X, Medrano-Gracia P: Time Series Modelling of Syphilis Incidence in China from 2005 to 2012. *PloS one* 2016, 11(2):e0149401.
7. Molina LL, Angon E, Garcia A, Moralejo RH, Caballero-Villalobos J, Perea J: Time series analysis of bovine venereal diseases in La Pampa, Argentina. *PloS one* 2018, 13(8):e0201739.
8. Guan P, Huang DS, Zhou BS: Forecasting model for the incidence of hepatitis A based on artificial neural network. *World journal of gastroenterology* 2004, 10(24):3579-3582.
9. Cunha GB, Luitgards-Moura JF, Naves EL, Andrade AO, Pereira AA, Milagre ST: [Use of an artificial neural network to predict the incidence of malaria in the city of Canta, state of Roraima]. *Revista da Sociedade Brasileira de Medicina Tropical* 2010, 43(5):567-570.
10. Baquero OS, Santana LMR, Chiaravalloti-Neto F: Dengue forecasting in Sao Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. *PloS one* 2018, 13(4):e0195065.
11. Kim ZM, Oh H, Kim HG, Lim CG, Oh KJ, Choi HJ: Modeling long-term human activeness using recurrent neural networks for biometric data. *BMC medical informatics and decision making* 2017, 17(Suppl 1):57.
12. Lipton ZC: A Critical Review of Recurrent Neural Networks for Sequence Learning. *Computer Science* 2015.
13. Wang H, Tian CW, Wang WM, Luo XM: Time-series analysis of tuberculosis from 2005 to 2017 in China. Epidemiology and infection 2018, 146(8):935-939.

14. Zheng YL, Zhang LP, Zhang XL, Wang K, Zheng YJ: Forecast model analysis for the morbidity of tuberculosis in Xinjiang, China. PloS one 2015, 10(3):e0116832.

15. Kumar V, Singh A, Adhikary M, Daral S, Khokhar A, Singh S: Seasonality of tuberculosis in delhi, India: a time series analysis. Tuberculosis research and treatment 2014, 2014:514093.

16. Liao Z, Zhang X, Zhang Y, Peng D: Seasonality and Trend Forecasting of Tuberculosis Incidence in Chongqing, China. Interdisciplinary sciences, computational life sciences 2019, 11(1):77-85.

17. Li Z, Wang Z, Song H, Liu Q, He B, Shi P, Ji Y, Xu D, Wang J: Application of a hybrid model in predicting the incidence of tuberculosis in a Chinese population. Infection and drug resistance 2019, 12:1011-1020.

18. Jing QL, Cheng Q, Marshall JM, Hu WB, Yang ZC, Lu JH: Imported cases and minimum temperature drive dengue transmission in Guangzhou, China: evidence from ARIMAX model. Epidemiology and infection 2018, 146(10):1226-1235.

19. Gharbi M, Quenel P, Gustave J, Cassadou S, La Ruche G, Girdary L, Marrama L: Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. BMC infectious diseases 2011, 11:166.

20. LeCun Y, Bengio Y, Hinton G: Deep learning. Nature 2015, 521(7553):436-444.

21. Wang KW, Deng C, Li JP, Zhang YY, Li XY, Wu MC: Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. Epidemiology and infection 2017, 145(6):1118-1129.

22. Shakhgeldyan KI, Geltser BI, Emtseva ED, Sereda VG, Mazelis LS, Sakharova OB: [The Analysis of Social Economic Factors Impact on Epidemic Situation with Tuberculosis in Regions of the Russian Federation]. Problemy sotsial'noi gigieny, zdravookhraneniia i istorii meditsiny 2018, 26(2):78-82.

23. de Castro DB, de Seixas Maciel EMG, Sadahiro M, Pinto RC, de Albuquerque BC, Braga JU: Tuberculosis incidence inequalities and its social determinants in Manaus from 2007 to 2016. International journal for equity in health 2018, 17(1):187.

24. Box GE, Jenkins GM, Reinsel GC, Ljung GM: Time series analysis: forecasting and control: John Wiley & Sons; 2015.

25. N’Gattia A K, Coulibaly D, Nzussouo NT, Kadjo HA, Cherif D, Traore Y, Kouakou BK, Kouassi PD, Ekra KD, Dagnan NS et al: Effects of climatological parameters in modeling and forecasting seasonal influenza
transmission in Abidjan, Cote d'Ivoire. *BMC public health* 2016, 16:972.

26. Liao CM, Hsieh NH, Huang TL, Cheng YH, Lin YJ, Chio CP, Chen SC, Ling MP: Assessing trends and predictors of tuberculosis in Taiwan. *BMC public health* 2012, 12:29.

27. Zhang Y, Liu M, Wu SS, Jiang H, Zhang J, Wang S, Ma W, Li Q, Ma Y, Liu Y et al: Spatial distribution of tuberculosis and its association with meteorological factors in mainland China. *BMC infectious diseases* 2019, 19(1):379.

28. Cao K, Yang K, Wang C, Guo J, Tao L, Liu Q, Gehendra M, Zhang Y, Guo X: Spatial-Temporal Epidemiology of Tuberculosis in Mainland China: An Analysis Based on Bayesian Theory. *International journal of environmental research and public health* 2016, 13(5).

29. Xiao Y, He L, Chen Y, Wang Q, Meng Q, Chang W, Xiong L, Yu Z: The influence of meteorological factors on tuberculosis incidence in Southwest China from 2006 to 2015. *Scientific reports* 2018, 8(1):10053.

30. Castro DB, Pinto RC, Albuquerque BC, Sadahiro M, Braga JU: The Socioeconomic Factors and the Indigenous Component of Tuberculosis in Amazonas. *PloS one* 2016, 11(6):e0158574.
Figure legends

Figure 1. The recurrent neural network (RNN).

A: The structure diagram of the RNN; B: The unfolding diagram of the forward propagation of the RNN.

\( X_t \): the input at time \( t \); \( h_t \): the hidden state at time \( t \); \( y_t \): the output at time \( t \); \( W \): the weight of the input; \( U \): the weight of the input at the moment; \( V \): the weight of the output.

Figure 2. The monthly pulmonary tuberculosis cases in three cities between 2005 and 2017.

Figure 3. The cross-correlation function plots of residual series of pulmonary tuberculosis and meteorological factors.

A: PTB and MAT; B: PTB and MAP; C: PTB and MAS; D: PTB and MAH; E: PTB and MP; F: PTB and MST; 1: Xuzhou; 2: Nantong; 3: Wuxi.

PTB: pulmonary tuberculosis; MAT: monthly average temperature; MAP: monthly average atmospheric pressure; MAS: monthly average wind speed; MAH: monthly average relative humidity; MP: monthly precipitation; MST: monthly sunshine time.
Table 1. The monthly number of pulmonary tuberculosis cases in the three cities in 2018 predicted by the ARIMA, ARIMAX, and RNN models.

| Month      | Xuzhou city |          | Nantong city |          | Wuxi city |          |
|------------|-------------|----------|--------------|----------|-----------|----------|
|            | Observation | ARIMA    | ARIMAX       | RNN      | Observation | ARIMA    | ARIMAX       | RNN      | Observation | ARIMA    | ARIMAX       | RNN      |
| January    | 237         | 239      | 227          | 258      | 193       | 145      | 177          | 194      | 158         | 137      | 135          | 141      |
| February   | 188         | 259      | 249          | 249      | 177       | 149      | 163          | 202      | 112         | 143      | 141          | 170      |
| March      | 300         | 323      | 304          | 227      | 261       | 216      | 258          | 209      | 194         | 174      | 175          | 177      |
| April      | 273         | 291      | 277          | 285      | 221       | 195      | 202          | 219      | 188         | 177      | 184          | 139      |
| May        | 271         | 288      | 283          | 282      | 251       | 196      | 212          | 206      | 206         | 178      | 181          | 194      |
| June       | 273         | 293      | 286          | 285      | 230       | 196      | 204          | 224      | 183         | 202      | 204          | 188      |
| July       | 229         | 242      | 237          | 250      | 204       | 167      | 186          | 200      | 218         | 183      | 187          | 193      |
| August     | 248         | 257      | 237          | 242      | 166       | 195      | 205          | 218      | 206         | 198      | 200          | 214      |
| September  | 169         | 255      | 247          | 215      | 170       | 189      | 170          | 207      | 203         | 195      | 199          | 192      |
| October    | 208         | 206      | 238          | 192      | 151       | 134      | 125          | 178      | 150         | 160      | 170          | 177      |
| November   | 193         | 209      | 243          | 180      | 149       | 146      | 150          | 183      | 191         | 183      | 181          | 199      |
| December   | 241         | 206      | 240          | 201      | 162       | 195      | 212          | 190      | 177         | 178      | 181          | 179      |
Table 2. The alternative ARIMAX models of three cities.

| City   | Model                          | Normalized BIC value | \( P^* \) | MAPE(%)\# |
|--------|--------------------------------|----------------------|---------|----------|
| Xuzhou | ARIMA (1,1,1)(0,1,1)_{12}      | 8.857                | 0.861   | 12.536   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAS2 | 8.595                | 0.714   | 14.053   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAH1 | 8.467                | 0.399   | 24.094   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MP2  | 8.617                | 0.356   | 11.957   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MST1 | 8.593                | 0.767   | 17.616   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAS2+MAH1 | 8.609            | 0.338   | 25.020   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAS2+MP2 | 8.658            | 0.691   | 17.221   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAS2+MST1 | 8.679            | 0.902   | 17.336   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAS2+MAH1+MST1 | 8.560        | 0.431   | 20.679   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAP1 | 8.183                | 0.777   | 11.155   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAP1+MAS2 | 8.323            | 0.730   | 16.289   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAP1+MAS2+MP2 | 8.700        | 0.371   | 20.711   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAP1+MAS2+MP2+MST1 | 8.755        | 0.427   | 23.012   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAP1+MAS2+MP2+MST1 | 8.755        | 0.851   | 17.209   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAP1+MAS2+MP2+MST1 | 8.692        | 0.241   | 39.174   |
|        | ARIMA (1,1,1)(0,1,1)_{12}+MAP1+MAS2+MP2+MST1 | 8.831        | 0.581   | 17.437   |
| Nantong| ARIMA (0,1,1)(0,1,1)_{12}      | 8.609                | 0.433   | 15.568   |
|        | ARIMA (0,1,1)(0,1,1)_{12}+MAT0 | 8.288                | 0.981   | 16.768   |
|        | ARIMA (0,1,1)(0,1,1)_{12}+MAP1 | 8.183                | 0.777   | 11.155   |
|        | ARIMA (0,1,1)(0,1,1)_{12}+MAS2 | 8.323                | 0.730   | 16.289   |
|        | ARIMA (0,1,1)(0,1,1)_{12}+MAT0+MAP1 | 8.340        | 0.836   | 14.992   |
|        | ARIMA (0,1,1)(0,1,1)_{12}+MAT0+MAS2 | 8.419            | 0.965   | 16.971   |
|        | ARIMA (0,1,1)(0,1,1)_{12}+MAT0+MAS2 | 8.314            | 0.766   | 11.904   |
|        | ARIMA (0,1,1)(0,1,1)_{12}+MAT0+MAP1+MAS2 | 8.470        | 0.892   | 13.055   |
| Wuxi   | ARIMA (0,1,1)(0,1,1)_{12}      | 6.933                | 0.176   | 9.700    |
|        | ARIMA (0,1,1)(0,1,1)_{12}+MAH0 | 6.845                | 0.119   | 9.660    |
|        | ARIMA (0,1,1)(0,1,1)_{12}+MST0 | 6.818                | 0.068   | 10.514   |
|        | ARIMA (0,1,1)(0,1,1)_{12}+MAH0+MST0 | 7.003            | 0.088   | 9.738    |

BIC: Bayesian information criterion; MAPE: mean absolute percentage error; MAT: monthly average temperature; MAP: monthly average atmospheric pressure; MAS: monthly average wind speed; MAH: monthly average relative humidity; MP: monthly precipitation; MST: monthly sunshine time; 0: 0-month lag; 1: 1-month lag; 2: 2-months lag.

\(^*\): The Ljung-Box test.

\(^\#\): The MAPE of the model in predicting the monthly number of PTB cases in 2018.
Table 3. The alternative RNN models of three cities.

| City      | Model                        | Learning rate | Dimensions of the hidden layer | Number of Epochs | MAPE(%)<sup>a</sup> | MAPE(%)<sup>b</sup> | MAPE(%)<sup>c</sup> |
|-----------|------------------------------|---------------|--------------------------------|------------------|----------------------|----------------------|----------------------|
| Xuzhou    | RNN1                         | 0.05          | 3                              | 500              | 16.143               | 15.989               | 16.455               |
|           | RNN2                         | 0.05          | 3                              | 500              | 13.416               | 13.304               | 14.405               |
|           | RNN3                         | 0.2           | 3                              | 150              | 13.082               | 11.952               | 12.068               |
|           | RNN4                         | 0.05          | 3                              | 600              | 10.331               | 10.329               | 10.399               |
|           | RNN5                         | 0.05          | 5                              | 600              | 8.451                | 8.250                | 8.541                |
|           | RNN6 (RNN5+MAS1)             | 0.05          | 3                              | 1000             | 7.361                | 7.327                | 7.330                |
|           | RNN7 (RNN5+MAS2+MST2)        | 0.05          | 3                              | 800              | 6.381                | 6.313                | 6.423                |
|           | RNN8 (RNN5+MAT3+MAS3+MP3+MST3) | 0.05         | 5                              | 600              | 4.776                | 4.887                | 4.973                |
|           | RNN9 (RNN5+MAS1+MAS2+MST2+MAT3+MAS3+MP3+MST3) | 0.05     | 10                             | 600              | 5.749                | 5.403                | 5.902                |
| Nantong   | RNN1                         | 0.05          | 3                              | 500              | 21.910               | 21.987               | 21.781               |
|           | RNN2                         | 0.2           | 5                              | 80               | 16.924               | 17.805               | 16.306               |
|           | RNN3                         | 0.2           | 3                              | 150              | 13.818               | 14.255               | 13.864               |
|           | RNN4                         | 0.2           | 3                              | 150              | 12.780               | 12.839               | 12.799               |
|           | RNN5                         | 0.2           | 5                              | 100              | 11.380               | 11.442               | 11.243               |
|           | RNN6 (RNN5+MAS1+MAH1)        | 0.05          | 5                              | 1000             | 9.185                | 8.823                | 8.383                |
|           | RNN7 (RNN5+MAS2+MAH2)        | 0.05          | 5                              | 1000             | 8.580                | 8.263                | 8.520                |
|           | RNN8 (RNN5+MAS3+MAH3)        | 0.05          | 10                             | 800              | 8.868                | 8.787                | 8.687                |
|           | RNN9 (RNN5+MAS1+MAH1+MAS2+MAH2+MAS3+MAH3) | 0.05   | 10                             | 800              | 8.789                | 9.213                | 9.187                |
| Wuxi      | RNN1                         | 0.1           | 10                             | 150              | 23.758               | 23.807               | 23.765               |
|           | RNN2                         | 0.05          | 5                              | 400              | 19.933               | 19.537               | 20.168               |
|           | RNN3                         | 0.05          | 10                             | 250              | 18.228               | 17.841               | 18.585               |
|           | RNN4                         | 0.05          | 10                             | 400              | 17.147               | 17.395               | 17.310               |
|           | RNN5                         | 0.05          | 5                              | 600              | 14.097               | 13.925               | 13.949               |
|           | RNN6 (RNN5+MAT1+MAP1+MAS1+MAH1+MST1) | 0.05  | 3                              | 1500             | 13.010               | 13.393               | 13.037               |
|           | RNN7 (RNN5+MAS2)             | 0.1           | 5                              | 800              | 12.622               | 12.364               | 12.797               |
|           | RNN8 (RNN5+MAT3+MAS3+MAH3)   | 0.05          | 10                             | 1000             | 12.713               | 13.060               | 12.944               |
|           | RNN9 (RNN5+MAT1+MAP1+MAS1+MAH1+MST1+MAS2+MAT3+MAS3+MAH3) | 0.1     | 3                              | 1000             | 12.810               | 12.798               | 13.458               |

MAPE: mean absolute percentage error; MAT: monthly average temperature; MAP: monthly average atmospheric pressure; MAS: monthly average wind speed; MAH: monthly average relative humidity; MP: monthly precipitation; MST: monthly sunshine time; 1: 1 month ago; 2: 2 months ago; 3: 3 months ago.

<sup>a</sup>: the MAPE of the model on the testing set after the first training. <sup>b</sup>: the MAPE of the model on the testing set after the second training. <sup>c</sup>: the MAPE of the model on the testing set after the third training.
Table 4. Evaluating the performance of ARIMA, ARIMAX, and RNN models in predicting the monthly number of pulmonary tuberculosis cases in three cities in 2018.

| City    | Diagnostic indicator | ARIMA | ARIMAX | RNN   |
|---------|----------------------|-------|--------|-------|
| Xuzhou  | MAPE(%)              | 12.536| 11.957 | 12.360|
|         | RMSE                 | 36.194| 33.956 | 34.785|
| Nantong | MAPE(%)              | 15.568| 11.155 | 14.087|
|         | RMSE                 | 34.073| 25.884 | 31.828|
| Wuxi    | MAPE(%)              | 9.700 | 9.660  | 12.501|
|         | RMSE                 | 19.545| 19.026 | 26.019|

MAPE: mean absolute percentage error; RMSE: root mean square error.
Figure 1. The recurrent neural network (RNN).

A: The structure diagram of the RNN; B: The unfolding diagram of the forward propagation of the RNN.

$X_t$: the input at time $t$; $h_t$: the hidden state at time $t$; $y_t$: the output at time $t$; $W$: the weight of the input; $U$: the weight of the input at the moment; $V$: the weight of the output.
Figure 2. The monthly pulmonary tuberculosis cases in three cities between 2005 and 2017.
Figure 3. The cross-correlation function plots of residual series of pulmonary tuberculosis and meteorological factors.

A: PTB and MAT; B: PTB and MAP; C: PTB and MAS; D: PTB and MAH; E: PTB and MP; F: PTB and MST; 1: Xuzhou; 2: Nantong; 3: Wuxi.

PTB: pulmonary tuberculosis; MAT: monthly average temperature; MAP: monthly average atmospheric pressure; MAS: monthly average wind speed; MAH: monthly average relative humidity; MP: monthly precipitation; MST: monthly sunshine time.
Supplementary files

Figure S1. The ACF and PACF plots.
A: The ACF plots of the monthly number of PTB cases after one ordinary difference and one seasonal difference in the three cities; B: The PACF plots of the monthly number of PTB cases after one ordinary difference and one seasonal difference in the three cities; C: The ACF plots of the residual series of the optimal ARIMA models of the three cities; D: The PACF plots of the residual series of the optimal ARIMA models of the three cities; 1: Xuzhou; 2: Nantong; 3: Wuxi.
ACF: autocorrelation function; PACF: partial autocorrelation function.

Figure S2. The time series plots of six meteorological factors in three cities between 2005 and 2017.

Figure S3. The epoch-error plots of the optimal RNN models of three cities after three times of training.
A: Xuzhou; B: Nantong; C: Wuxi; 1: the first training; 2: the second training; 3: the third training.
Supplementary Tables

Table S1. The alternative ARIMA models of the three cities.

| City   | Model                        | Normalized BIC value | $p^*$  | The parameters were all significant |
|--------|------------------------------|----------------------|--------|-------------------------------------|
|        | Xuzhou                       |                      |        |                                     |
|        | ARIMA (0,1,1)(0,1,1)$_{12}$  | 8.863                | 0.316  | Yes                                 |
|        | ARIMA (1,1,0)(0,1,1)$_{12}$  | 8.882                | 0.276  | Yes                                 |
|        | ARIMA (1,1,1)(0,1,1)$_{12}$  | 8.857                | 0.861  | Yes                                 |
|        | ARIMA (0,1,1)(1,1,0)$_{12}$  | 8.936                | 0.653  | Yes                                 |
|        | ARIMA (1,1,0)(1,1,0)$_{12}$  | 8.955                | 0.498  | Yes                                 |
|        | ARIMA (1,1,1)(1,1,0)$_{12}$  | 8.952                | 0.804  | No                                  |
|        | ARIMA (0,1,1)(1,1,1)$_{12}$  | 8.912                | 0.146  | Yes                                 |
|        | ARIMA (1,1,0)(1,1,1)$_{12}$  | 8.928                | 0.106  | Yes                                 |
|        | ARIMA (1,1,1)(1,1,1)$_{12}$  | 8.910                | 0.549  | Yes                                 |
|        | Nantong                      |                      |        |                                     |
|        | ARIMA (0,1,1)(0,1,1)$_{12}$  | 8.609                | 0.433  | Yes                                 |
|        | ARIMA (1,1,0)(0,1,1)$_{12}$  | 8.679                | 0.106  | Yes                                 |
|        | ARIMA (1,1,1)(0,1,1)$_{12}$  | 8.632                | 0.508  | Yes                                 |
|        | ARIMA (2,1,0)(0,1,1)$_{12}$  | 8.661                | 0.519  | Yes                                 |
|        | ARIMA (2,1,1)(0,1,1)$_{12}$  | 8.668                | 0.325  | No                                  |
|        | ARIMA (0,1,1)(1,1,0)$_{12}$  | 8.643                | 0.739  | Yes                                 |
|        | ARIMA (1,1,0)(1,1,0)$_{12}$  | 8.719                | 0.125  | Yes                                 |
|        | ARIMA (1,1,1)(1,1,0)$_{12}$  | 8.667                | 0.300  | Yes                                 |
|        | ARIMA (2,1,0)(1,1,0)$_{12}$  | 8.701                | 0.669  | Yes                                 |
|        | ARIMA (2,1,1)(1,1,0)$_{12}$  | 8.696                | 0.313  | No                                  |
|        | ARIMA (0,1,1)(1,1,1)$_{12}$  | 8.649                | 0.367  | No                                  |
|        | ARIMA (1,1,0)(1,1,1)$_{12}$  | 8.721                | 0.078  | No                                  |
|        | ARIMA (1,1,1)(1,1,1)$_{12}$  | 8.666                | 0.452  | No                                  |
|        | ARIMA (2,1,0)(1,1,1)$_{12}$  | 8.702                | 0.455  | No                                  |
|        | ARIMA (2,1,1)(1,1,1)$_{12}$  | 8.703                | 0.305  | No                                  |
|        | Wuxi                         |                      |        |                                     |
|        | ARIMA (0,1,1)(0,1,1)$_{12}$  | 6.933                | 0.176  | Yes                                 |
|        | ARIMA (1,1,0)(0,1,1)$_{12}$  | 7.088                | 0.005  | Yes                                 |
|        | ARIMA (1,1,1)(0,1,1)$_{12}$  | 6.958                | 0.256  | No                                  |

BIC: Bayesian information criterion.

*: The Ljung-Box test.
### Table S2. The spearman rank correlation coefficients between the monthly number of PTB cases and meteorological factors in the three cities.

| Variable | MAT1 | MAP1 | MAS1 | MAH1 | MP1 | MST1 | MAT2 | MAP2 | MAS2 | MAH2 | MP2 | MST2 | MAT3 | MAP3 | MAS3 | MAH3 | MP3 | MST3 |
|----------|------|------|------|------|-----|------|------|------|------|------|-----|------|------|------|------|------|-----|------|
| PTBα     | 0.035 | -0.115 | 0.536* | -0.025 | 0.004 | -0.153 | -0.115 | 0.017 | 0.519* | -0.083 | -0.075 | -0.244* | -0.227* | 0.137 | 0.450* | -0.133 | -0.178* | -0.262* |
| PTBβ     | 0.068 | -0.090 | -0.177* | -0.199* | -0.054 | 0.114 | -0.043 | -0.007 | -0.219* | -0.229* | -0.117 | 0.098 | -0.103 | 0.061 | -0.240* | -0.184* | -0.114 | 0.042 |
| PTBγ     | 0.216* | -0.297* | 0.461* | -0.164* | 0.099 | 0.168* | 0.011 | -0.132 | 0.398* | -0.155 | 0.053 | -0.160* | 0.035 | 0.425* | -0.193* | -0.065 | 0.060 |

MAT: monthly average temperature; MAP: monthly average atmospheric pressure; MAS: monthly average wind speed; MAH: monthly average relative humidity; 
MP: monthly precipitation; MST: monthly sunshine time; 1: 1 month ago; 2: 2 months ago; 3: 3 months ago.

α: the number of PTB cases in the current month in Xuzhou.

β: the number of PTB cases in the current month in Nantong.

γ: the number of PTB cases in the current month in Wuxi.

*: $P < 0.05$.

#: $P < 0.01$. 


| Model | Training set | Testing set | Predicting set |
|-------|--------------|-------------|---------------|
|       | Input | Output | Input | Output | Input | Output |
| RNN1  | First group: NPTBC in Jan. 2005; Last group: NPTBC in Nov. 2016. | First group: NPTBC in Feb. 2005; Last group: NPTBC in Dec. 2016. | First group: NPTBC in Dec. 2016; Last group: NPTBC in Nov. 2017. | First group: NPTBC in Dec. 2017; Last group: NPTBC in Nov. 2018. |
| RNN2  | First group: NPTBC from Jan. to Feb. 2005; Last group: NPTBC from Oct. to Nov. 2016. | First group: NPTBC in Mar. 2005; Last group: NPTBC in Dec. 2016. | First group: NPTBC from Nov. to Dec. 2016; Last group: NPTBC from Oct. to Nov. 2017. | First group: NPTBC from Nov. to Dec. 2017; Last group: NPTBC from Oct. to Nov. 2018. |
| RNN3  | First group: NPTBC from Jan. to Mar. 2005; Last group: NPTBC from Sep. to Nov. 2016. | First group: NPTBC in Apr. 2005; Last group: NPTBC in Dec. 2016. | First group: NPTBC from Oct. to Dec. 2016; Last group: NPTBC from Sep. to Nov. 2017. | First group: NPTBC from Oct. to Dec. 2017; Last group: NPTBC from Sep. to Nov. 2018. |
| RNN4  | First group: NPTBC from Jan. to Jun. 2005; Last group: NPTBC from Jun. to Nov. 2016. | First group: NPTBC in Jul. 2005; Last group: NPTBC in Dec. 2016. | First group: NPTBC from Jul. to Dec. 2016; Last group: NPTBC from Jun. to Nov. 2017. | First group: NPTBC from Jul. to Dec. 2017; Last group: NPTBC from Jun. to Nov. 2018. |
| RNN5  | First group: NPTBC from Jan. to Dec. 2005; Last group: NPTBC from Dec. 2015 to Nov. 2016. | First group: NPTBC in Jan. 2006; Last group: NPTBC in Dec. 2016. | First group: NPTBC from Jan. to Dec. 2016; Last group: NPTBC from Dec. 2016 to Nov. 2017. | First group: NPTBC from Jan. to Dec. 2017; Last group: NPTBC from Dec. 2017 to Nov. 2018. |
| RNN6  | First group: NPTBC from Jan. to Dec. 2005 and MF in Dec. 2005; Last group: NPTBC from Dec. | First group: NPTBC in Jan. 2006; Last group: NPTBC in Dec. 2016. | First group: NPTBC from Jan. to Dec. 2016 and MF in Dec. 2016; Last group: NPTBC from Dec. | First group: NPTBC from Jan. to Dec. 2017 and MF in Dec. 2017; Last group: NPTBC from Dec. |
|   | First group: NPTBC from Jan. to Dec. 2005 and MF in Nov. 2005; Last group: NPTBC from Dec. 2015 to Nov. 2016 and MF in Oct. 2016. | First group: NPTBC in Jan. 2006; Last group: NPTBC in Dec. 2016. | First group: NPTBC from Jan. to Dec. 2016 and MF in Oct. 2016; Last group: NPTBC from Dec. 2016 to Nov. 2017 and MF in Sep. 2017. | First group: NPTBC from Jan. to Dec. 2017 and MF from Oct. to Dec. 2017; Last group: NPTBC from Dec. 2017 to Nov. 2018 and MF from Sep. to Nov. 2018. |
|---|---|---|---|---|
| RNN7 | 2015 to Nov. 2016 and MF in Nov. 2016. | 2016 to Nov. 2017 and MF in Nov. 2017. | 2016 to Nov. 2017 and MF in Nov. 2018. | 2017 to Nov. 2018 and MF in Nov. 2018. |
| RNN8 | First group: NPTBC from Jan. to Dec. 2005 and MF in Oct. 2005; Last group: NPTBC from Dec. 2015 to Nov. 2016 and MF in Sep. 2016. | First group: NPTBC in Jan. 2006; Last group: NPTBC in Dec. 2016. | First group: NPTBC from Jan. to Dec. 2016 and MF in Oct. 2016; Last group: NPTBC from Dec. 2016 to Nov. 2017 and MF in Sep. 2017. | First group: NPTBC from Jan. to Dec. 2017 and MF in Oct. 2017; Last group: NPTBC from Dec. 2017 to Nov. 2018 and MF in Sep. 2018. |
| RNN9 | First group: NPTBC from Jan. to Dec. 2005 and MF from Oct. to Dec. 2005; Last group: NPTBC from Dec. 2015 to Nov. 2016 and MF from Sep. to Nov. 2016. | First group: NPTBC in Jan. 2006; Last group: NPTBC in Dec. 2016. | First group: NPTBC from Jan. to Dec. 2016 and MF from Oct. to Dec. 2016; Last group: NPTBC from Dec. 2016 to Nov. 2017 and MF from Sep. to Nov. 2017. | First group: NPTBC from Jan. to Dec. 2017 and MF from Oct. to Dec. 2017; Last group: NPTBC from Dec. 2017 to Nov. 2018 and MF from Sep. to Nov. 2018. |

NPTBC: number of PTB cases; MF: meteorological factors significantly correlated with PTB.
Figure S1. The ACF and PACF plots.

A: The ACF plots of the monthly number of PTB cases after one ordinary difference and one seasonal difference in the three cities; B: The PACF plots of the monthly number of PTB cases after one ordinary difference and one seasonal difference in the three cities; C: The ACF plots of the residual series of the optimal ARIMA models of the three cities; D: The PACF plots of the residual series of the optimal ARIMA models of the three cities; 1: Xuzhou; 2: Nantong; 3: Wuxi.

ACF: autocorrelation function; PACF: partial autocorrelation function.
Figure S2. The time series plots of the six meteorological factors in the three cities between 2005 and 2017.
Figure S3. The epoch-error plots of the optimal RNN models of the three cities after three times of training.

A: Xuzhou; B: Nantong; C: Wuxi; 1: the first training; 2: the second training; 3: the third training.
The recurrent neural network (RNN). A: The structure diagram of the RNN; B: The unfolding diagram of the forward propagation of the RNN. $X_t$: the input at time $t$; $h_t$: the hidden state at time $t$; $y_t$: the output at time $t$; $W$: the weight of the input; $U$: the weight of the input at the moment; $V$: the weight of the output.
Figure 2

The monthly pulmonary tuberculosis cases in three cities between 2005 and 2017.
Figure 3

The cross-correlation function plots of residual series of pulmonary tuberculosis and meteorological factors. A: PTB and MAT; B: PTB and MAP; C: PTB and MAS; D: PTB and MAH; E: PTB and MP; F: PTB and MST; 1: Xuzhou; 2: Nantong; 3: Wuxi. PTB: pulmonary tuberculosis; MAT: monthly average temperature; MAP: monthly average atmospheric pressure; MAS: monthly average wind speed; MAH: monthly average relative humidity; MP: monthly precipitation; MST: monthly sunshine time.