Forecasting incidence of infectious diarrhea using random forest in Jiangsu rovince, China

Xinyu Fang
Nanjing Medical University
Wendong Liu
Jiangsu Province Center for Disease Control and Prevention
Jing Ai
Jiangsu Province Center for Disease Control and Prevention
Ying Wu
Jiangsu Province Center for Disease Control and Prevention
Yingying Shi
Jiangsu Province Center for Disease Control and Prevention
Wenqi Shen
Jiangsu Province Center for Disease Control and Prevention
Changjun Bao (bao2000_cn@163.com)
Jiangsu Province Center for Disease Control and Prevention https://orcid.org/0000-0002-0546-1338
He Mike
Columbia University Mailman School of Public Health

Research article

Keywords: Infectious diarrhea; Forecasting; Random forest

Posted Date: January 7th, 2020

DOI: https://doi.org/10.21203/rs.2.15862/v2

Version of Record: A version of this preprint was published at BMC Infectious Diseases on March 14th, 2020. See the published version at https://doi.org/10.1186/s12879-020-4930-2.

Abstract

Background: Infectious diarrhea can lead to considerable disease burden around the world. Thus, the accurate prediction of infectious diarrhea epidemic is crucial for public health authorities. This study aimed to develop an optimal random forest (RF) model considering meteorological factors to predict morbidity of infectious diarrhea in Jiangsu Province, China. Methods: A RF model was constructed and compared with the classical autoregressive integrated moving average (ARIMA)/X models. Morbidity and meteorological data from 2012–2016 were used for model construction and the rest data in 2017 were used for testing. Results: The RF model considered atmosphere pressure, precipitation, relative humidity and their lagged terms, 1-4 weeks' lag morbidity and the time variable as predictors. Meanwhile, a univariate model ARIMA(1,0,1)(1,0,0) 52 (AIC=−575.92, BIC=−558.14) and a multivariable model ARIMAX(1,0,1)(1,0,0) 52 with 0-1 week's lag precipitation (AIC=−578.58, BIC=−578.13) were developed as benchmark models . The RF model outperformed the ARIMA/X models with a mean absolute percentage error (MAPE) of approximately 20% . The performance of the ARIMAX model was similar to that of the ARIMA model with MAPE approximately as high as 30%. Conclusions: The RF model well fitted the dynamic of the infectious diarrhea epidemic and achieved ideal prediction accuracy. It comprehensively combined meteorological factors and their hysteresis effects. It also integrated the autocorrelation and seasonality of morbidity. The RF model could be used to predict the epidemic level, and has good potential of practical application.

Background

Infectious diarrhea is one of the crucial causes of morbidity and mortality in infants and young population. This is a global major public health issue, especially in developing countries[1]. In 2015, diarrheal diseases led to an estimated 688 million illnesses and 499,000 deaths among children under the age of 5 years [2]. In the past decade, the morbidity has also increased in various regions in China[3]. Thus, an accurate forecast of infectious diarrhea based on predictive models is crucial for public health authorities to clearly understand the epidemic characteristics, track the seasonal updates in advance and select main response actions, such as disease surveillance and deployment of emergency supplies [4].

The autoregressive integrated moving average (ARIMA) model is widely used in diarrhea incidence prediction as classical method, however, it has some limitations at the same time [4-7]. For example, Yang et al.[4] used the ARIMA model in the early warning systems of diarrhea, and they achieved a poor fit. Several studies have shown that meteorological factors are associated with diarrhea and can be used to predict its incidence[8-9]. Yan et al.[7] developed a multivariable ARIMA (ARIMAX) model considering temperature and rainfall and only achieved high short-term predictive accuracy. This might be because the ARIMA/X models assume linear relationships between the independent and dependent variables. However, meteorological factors were shown to be non-linearly associated with the epidemic of infectious diarrhea[9-10].

The RF model is a new regression method and could address the limitations of ARIMA/X models in the prediction of diarrhea incidence[11-14]. It can effectively extract non-linear relationships from data. The RF model uses independent variables to create classification and regression trees (CARTs) and each constituent tree is trained on a potentially non-linear regression space. The RF model might have good predictive stability for actual instable morbidity. In the RF model, the training set for each tree is randomly selected from the data, and the final predicted value is the average of all the CARTs outputs. The RF model has been widely used for infectious diseases prediction such as predicting the West Nile virus infection and Bovine viral diarrhea[12-13]. Notably, Michael et al.[14] showed that the RF model has advantages over the ARIMA model in predicting avian influenza H5N1 outbreaks. However, no study has used the RF model to predict infectious diarrhea morbidity.

This study aimed to develop an optimal RF model to predict infectious diarrhea epidemic with meteorological factors in Jiangsu Province, China. Meanwhile, the performance of the RF model was compared with that of the ARIMA/X models. The model can be used to develop an early warning system for infectious diarrhea to facilitate preventive strategies more effectively.

Methods

Study area

Jiangsu Province, located on the eastern-coast of China (latitude 30°45'–35°20'N and longitude 116°18'–121°57'E), has an area of 102,600 km^2 and a population of approximately 80 million. It has a typical temperate subtropical monsoon climate with a mild temperature, moderate rainfall and distinct four-season pattern.

Data sources

In China, infectious diarrhea (excluding cholera, dysentery, typhoid and paratyphoid) is an intestinal infectious disease with diarrhea and/or vomiting as the main symptom. It has been listed as a legal Class C infectious disease [3]. An infectious diarrhea case, clinically diagnosed or etiologically confirmed in any hospitals or healthcare institutions throughout the country, must be reported timely and directly to the National Notifiable Disease Surveillance System (NNDSS) [15] (http://www.cdpc.chinacdc.cn). In this study, the weekly number of infectious diarrhea cases in Jiangsu Province during 2012–2017 were downloaded from the NNDSS, including both clinically diagnosed cases and etiologically confirmed cases.

The demographic data were collected from the Jiangsu provincial statistics department. The weekly meteorological variables were calculated based on the daily data obtained from the Jiangsu Meteorological Service Center. The data included atmospheric pressure, mean temperature, maximum temperature, minimum temperature, precipitation, relative humidity and sunshine duration.

ARIMA/X model

ARIMA model, namely the Box–Jenkins model, is widely used for time series analysis[16]. The seasonal ARIMA, which incorporates seasonal variation and is developed from the ARIMA model, performs better in the presence of an obvious seasonal pattern[17-18]. It is denoted as ARIMA(p,d,q)$(P,D,Q)_S$, where p, d and q indicate the orders of general auto-regression (AR), differencing and moving average (MA) terms; P, D and Q are orders of seasonal AR, differencing and MA terms, respectively; and s is the seasonal periodicity (s=52 weeks in this study)[18].

The ARIMA model is fitted in three essential steps as follows:

First, the augmented Dickey–Fuller test is implemented to detect whether the original time series is stationary (statistical properties such as mean and variance are all constant over time). If not, logarithmic transformation or difference is adopted to achieve stability.

Second, ARIMA models are established for the stationary time series, the one with the minimum Akaike information criterion (AIC) and Bayesian information criterion (BIC) values is considered as the optimal model. Then, the model parameters are estimated using the conditional least square method.

Third, to verify the adequacy of the ARIMA model, the Box–Ljung test is conducted to test whether the residual series is white noise sequence. White noise sequence is a purely random time series without autocorrelation, and useful information cannot be extracted from it for model fitting. If not, the model should be reestablished. Finally, prospective prediction is performed using the optimal model.

Based on the optimal ARIMA model, the multivariate ARIMA model including meteorological factors as external regressors[19] is further developed, denoted as ARIMAX model.

In this study, the ARIMA/X models were used as a reference to evaluate the performance of the RF model. The cross-correlation analysis was used to identify the lagged associations (1-4 weeks' lag [20-21]) between the meteorological factors and the incidence of infectious diarrhea.

RF Model

RF model is an ensemble machine learning method proposed by Breiman [11]. It creates multiple CARTs, each tree is trained on a bootstrap sample of the original training data with a randomly selected subset of input variables, and it takes the average of the outputs of the CARTs as the final prediction. One of its most important features is calculating the variable importance, which measures the association between a given variable and the prediction accuracy by the percentage increase in the mean square-error (%IncMSE).

The RF model fitting consists of four essential steps [14]:

First, the bootstrap sampling method is used to randomly select sample units from the original training data to create multiple CARTs.

Second, the bootstrap sampling method is also used to select candidate variables for each CART. In this study, the related meteorological variables were chosen as the predictors. Meanwhile, 1-4 weeks' lag morbidity and time variable were incorporated into the RF model to consider the effect of autocorrelation and seasonality of the dependent variable, respectively.

Third, the average of outputs from all the CARTs are calculated as the final predictive value.

Fourth, the importance of each variable is assessed by the decrease in accuracy.

Model evaluation

Three models were fitted in this study, the RF model with meteorological variables, univariate ARIMA model and multivariate ARIMAX model. The data subset for the period 2012–2016 were used as the training set to fit models, and data from 2017 were used as the testing set to evaluate the forecasting accuracies. Root mean square error (RMSE) and mean absolute percentage error (MAPE) were selected to evaluate the performance of each model, which were calculated as follows: (see Calculations in the Supplemental Files)

where n is number of real data or predicted values, is the real data, and is the predicted value.

Statistics analysis

All analyses were performed in R (version3.5.1). Seasonal decomposition was carried out to elucidate the temporal pattern of infectious diarrhea. The RF model were fitted using the “randomForest” package, and the ARIMA/X models were fitted using the “Forecast” package.

Results

General description

A total of 102,020 cases were detected during 2012–2017 in Jiangsu Province, China, reaching an average annual incidence of 21.40 per 100,000. As shown in Figure 1, the incidence exhibited an increasing long-term trend in 6 years. Moreover, there was a distinct seasonality, i.e., two incidence peaks were observed in each year: a higher winter peak from December to February and a lower summer peak from July to September. The descriptive statistics for the meteorological variables were summarized in Table 1.

Table 1. Summary of weekly meteorological factors in Jiangsu Province, 2012–2017

Variable	Min	P25	Median	P75	Max
nospheric pressure (Pa)	998.58	1007.02	1015.38	1022.56	1032.09
lean temperature (°C)	-2.19	7.39	17.13	23.67	32.65
ximum temperature (°C)	1.25	12.36	22.38	27.60	37.41
nimum temperature (°C)	-4.77	3.59	13.08	20.63	28.24
Relative humidity (%)	45.93	68.06	74.69	80.40	91.88
Precipitation (mm)	0.00	3.53	11.94	30.12	59.66
nshine duration (hour)	2.25	27.71	37.50	48.72	82.01

Correlation analysis

As shown in Table 2, atmospheric pressure and precipitation were significantly associated with the 0–2 weeks’ lag and 0–3 weeks’ lag infectious diarrhea morbidity, respectively. Meanwhile, relative humidity was related to synchronous morbidity (r_s=−0.13, *P*=0.02). Temperature variables as well as sunshine duration were not correlated with the incidence.

Table2. Cross correlation coefficients between infectious diarrhea and meteorological factors in Jiangsu Province, 2012-2017

›spheric ssure Pa)	Mean temperature (℃)	Maximum temperature (℃)	Minimum temperature (℃)	Relative humidity (%)	Precipitation (mm)	Sunshine duration (hour)
21**	−0.10	−0.09	−0.11	−0.13*	−0.23**	0.07
17**	−0.06	−0.06	−0.07	−0.08	−0.22**	0.05
.12*	−0.02	−0.01	−0.02	−0.04	−0.14*	0.03
).08	0.03	0.03	0.03	−0.02	−0.12*	0.04
).04	0.08	0.08	0.08	0.04	−0.08	0.05

Note: *P<0.05, **P<0.01

Note: *P<0.05, **P<0.01

Model fitting

ARIMA/X model

The original time series of infectious diarrhea morbidity was stationary (Dickey−Fuller=−4.26, P<0.01). Univariate ARIMA models were developed. The best-fitting ARIMA model was determined to be ARIMA(1,0,1)(1,0,0)$_{52}$, with a minimum AIC=−575.92 and a minimum BIC=−558.14. The Ljung−Box test (χ^2=0.01, P=0.93) suggested that the residual series of the model was white noise sequence.

Then, meteorological factors were added as covariates into the optimal ARIMA model to establish multivariate ARIMAX models. Finally, ARIMAX(1,0,1)(1,1,0)$_{52}$ with 0−1 week's lag precipitation was identified as the optimal ARIMAX model, with a minimum AIC=−578.58 and a minimum BIC=−578.13 (Ljung−Box test: χ^2=0.00548, P=0.10).

RF model

A RF model was constructed, with predictors atmosphere pressure, precipitation and their lagged terms, relative humidity, 1−4 week's lag morbidity and time variable. Figure 2 showed that the lagged dependent terms were the most imperative among all predictors. Atmospheric pressure and its lagged terms were shown to be the most vital metrological factors, followed by lagged precipitation.

Prediction performance comparison

Table 3 compared the RF model and the ARIMA/X models, and the predicting outputs were shown in Figure 3. The RF model with meteorological variables outperformed the ARIMA/X models in both model fitting and prospective stages in terms of RMSE and MAPE. The values predicted by RF matched the actual values very well, with a MAPE of approximately 20%. The performance of the ARIMAX model was similar to that of the ARIMA model with a high MAPE of approximately 30%.

Table 3. Performance of RF and ARIMA/X model

Model	RMSE		MAPE (%)	
	Training set	Testing set	Training set	Testing set
RF	0.04	0.31	6.88	20.89
ARIMAX(1,0,1)(1,0,0)$_{52}$	0.08	0.46	13.64	28.06
ARIMA(1,0,1)(1,0,0)$_{52}$	0.08	0.45	13.78	28.53

Discussion

The incidence of infectious diarrhea in Jiangsu Province exhibited a long-term gradual growth trend. Mathematical models for predicting are urgently required to reinforce integrated management for monitoring, control and prevention of infectious diarrhea. We constructed a RF model with meteorological factors, which had good accuracy for predicting the incidence of infectious diarrhea with a MAPE of approximately 20%. It could accurately estimate the seasonal fluctuation of this disease. The model may be used as an important tool by public health authorities.

The RF model is more suitable than ARIMA/X method for predicting infectious diarrhea epidemic in the study region. The performance of the ARIMAX model was similar to that of the ARIMA model, which suggested that the introduction of meteorological factors did not significantly optimize the prediction accuracy of the ARIMA model. This finding was consistent with those of other studies[3-5]. The RF model provided a meaningfully better fit to the data in terms of RMSE and MAPE. Compared with the ARIMA/X models, the prediction error of the RF model decreased by approximately 50% and 30% in the training and testing sets, respectively. This is because the RF model can better fit non-linear relationships. Moreover, compared with the ARIMAX model, the RF model is not influenced by multicollinearity, mainly because of the random selection of variables for each tree in RF[11]. The meteorological factors and their lagged terms were incorporated into the models if they were significantly correlated with the incidence of infectious diarrhea. All of them exhibited a certain degree of importance, which suggested that the RF model comprehensively combined the climatic variables and their hysteresis effects. In particular, the models partly underestimated the diarrhea incidence in 2017. It's largely due to the sharp increase of diarrhea incidence in 2017, which indicated that the potential influencing factors might have changed over 52 weeks, such as increase in the number of outbreaks, or changes in pathogen spectrum[22-23]. In addition to meteorological factors, some other variables should be considered to better optimize the prediction accuracy of the RF model.

Atmospheric pressure, precipitation and relative humidity were all correlated with the incidence of infectious diarrhea in Jiangsu Province with 0-2 weeks' lag, zero-week lag and 0-3 weeks' lag, respectively. However, Tao et al. [20] showed that atmospheric pressure and relative humidity were related to 0-1 week's lag diarrhea morbidity in Lanzhou city (northwest China), respectively. Relative humidity was related to four-week lag diarrhea incidence in Beijing city (north China)[21]. This difference may because of the regional differences in pathogen composition and climate conditions. Furthermore, the meteorological factors significantly contributed to the forecasting ability of the RF model, with atmospheric pressure as the main contributor. A high atmospheric pressure may affect survival of diarrhea causing microorganisms in the environment, although effects tend to be species specific[24]. The precipitation had moderate importance in the RF model, especially the three-week lag effect. This implied that precipitation in the previous three weeks may influence the morbidity and thus can be used to predict it. Relative humidity was identified as the least important factor. Relative humidity in Jiangsu Province exhibited a narrow variation at the weekly level, and did not fit better with the incidence rate. These findings may help future studies on specific relationship between climate and infectious diarrhea.

Notably, the prediction performance is likely to vary in different climatic regions. The generalizability of the RF model for the diarrhea incidence in Jiangsu Province to other regions might not be straightforward. However, the use of the RF model incorporating meteorological factors in the detection and prediction of diarrhea may provide an opportunity for reallocating healthcare resources more efficiently in other regions. In addition, considering the autocorrelation and obvious seasonality of diarrhea, the 0-4 weeks' lag morbidity and time variable were incorporated in the RF model, which were more important than meteorological factors in improving the prediction accuracy of the RF model. These strategies should be used as a reference when fitting similar RF models.

This study had a few limitations. First, some mild cases might use home therapies, and some cases with atypical symptoms may be misdiagnosed, therefore, the reported data may underestimate morbidity. Second, in this study, only meteorological variables were considered to improve the prediction ability. Other factors associated with infectious diarrhea may also be used as good predictors, this should be studied further. Third, our study was based on provincial-wide data, and did not consider geographical disparities and pathogen heterogeneity within the province. Further studies accounting for geographical disparities, such as rural or urban, may be useful. Fourth, similar to other machine learning methods such as artificial neural networks, the RF model cannot explain the specific non-liner relationship between meteorological factors and disease.

Conclusions

The RF model with meteorological factors had satisfactory prediction accuracy. It could be used to predict the epidemic level, and has potentials for practical applications. The autocorrelation and seasonal variation in the data of the dependent variables are crucial for the predication model. Meanwhile, the effects of meteorological factors and the cumulative effects over a period of time were combined to improve the model. Future studies should explore the RF model with meteorological factors as well as other variables to develop a useful tool for predicting other major infectious diseases.

Abbreviations

Random forest (RF); Autoregressive integrated moving average (ARIMA); Mean absolute percentage (MAPE); classification and regression tree (CART); Auto-regression (AR); Moving average (MA); Akaike information criterion (AIC); Bayesian information criterion (BIC); Root mean square error (RMSE)

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee at Jiangsu Provincial Center for Disease Control and Prevention, China. As the incidence was statistical summary data, it was not necessary to obtain informed consent.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used in this study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Jiangsu Provincial Major Science & Technology Demonstration Project (No. BE2017749); Jiangsu Provincial natural science foundation (No. BK20151595); Jiangsu Provincial Medical Youth Talent (No. QNRC2016542); Key Medical Discipline of Epidemiology (No. ZDXK A2016008) and National Major S&T Projects (No.2018ZX10714-002)

Authors' contributions

XYF, WDL and JA designed the study. ML, YW, YYS, WQS prepared and processed the initial data, XYF and WDL analyzed the data and drafted the manuscript. CJB revised the manuscript. All authors read and approved the final manuscript

Acknowledgements

We gratefully acknowledge Jiangsu Meteorological Service Center for supplying the daily meteorological data.

References

[1]GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015.Lancet. 2016;388(10053):1545-1602.

[2]GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1459-1544.

[3]Zhang P, Zhang J. Surveillance on other infectious diarrheal diseases in China from 2014 to 2015.Chin J Epidemiol. 2017;38(4):424-430.(in Chinese)

[4]Yang E, Park HW, Choi YH, Kim J, Munkhdalai L, Musa I, et al. A Simulation-Based Study on the Comparison of Statistical and Time Series Forecasting Methods for Early Detection of Infectious Disease Outbreaks. Int J Environ Res Public Health. 2018;15(5):966.

[5]Zhang Y, Bi P, Hiller JE, Sun Y, Ryan P. Climate variations and bacillary dysentery in northern and southern cities of China. J Infect. 2007;55(2):194–200.

[6]Gao L, Zhang Y, Ding G, Liu Q, Zhou M, Li X, et al. Meteorological variables and bacillary dysentery cases in Changsha City, China. Am J Trop Med Hyg. 2014;90(4):697-704.

[7]Yan L, Wang H, Zhang X, Li MY, He J. Impact of meteorological factors on the incidence of bacillary dysentery in Beijing, China: A time series analysis (1970-2012). PLoS One. 2017;12(8):e0182937.

[8]Chou WC, Wu JL, Wang YC, Huang H, Sung FC, Chuang CY. Modeling the impact of climate variability on diarrhea-associated diseases in Taiwan (1996-2007). Sci Total Environ. 2010;409(1):43-51.

[9]Phung D, Huang C, Rutherford S, Chu C, Wang X, Nguyen M, et al. Association between climate factors and diarrhea in a Mekong Delta area. Int J Biometeorol. 2015;59(9):1321-31.

[10]Li Z, Wang L, Sun W, Hou X, Yang H, Sun, L, et al. Identifying high-risk areas of bacillary dysentery and associated meteorological factors in Wuhan, China. Sci Rep. 2013;3(1):3239.

[11]Breiman L. Machine Learning: Kluwer Academic Publishers. California, USA;2011.p.5-32.

[12]Keyel AC, Elison Timm O, Backenson PB, Prussing C, Quinones S, McDonough KA, et al. Seasonal temperatures and hydrological conditions improve the prediction of West Nile virus infection rates in Culex mosquitoes and human case counts in New York and Connecticut. PLoS One. 2019;14(6):e0217854.

[13]Machado G, Mendoza MR, Corbellini LG. What variables are important in predicting bovine viral diarrhea virus? A random forest approach. Vet Res. 2015;24(1):46-85.

[14]Kane MJ, Price N, Scotch M, Rabinowitz P. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinformatics. 2014;15(1):276.

[15]Wang L, Wang Y, Jin S, Wu Z, Chin DP, Koplan JP et al. Emergence and control of infectious diseases in China. Lancet. 2008;372(9649):1598-605.

[16] Box G, Jenkins G, Reinsel G. Time series analysis: forecasting and control. Hoboken. New Jersey: John Wiley & Sons;2008.

[17]Xu Q, Li R, Liu Y, Luo C, Xu A, Xue F, et al. Forecasting the Incidence of Mumps in Zibo City Based on a SARIMA Model. Int J Environ Res Public Health. 2017;14(18):925.

[18]Tian CW, Wang H, Luo XM. Time-series modelling and forecasting of hand, foot and mouth disease cases in China from 2008 to 2018. Epidemiol Infect. 2019;147(1):28.

[19] Zhang Y, Bi P, Hiller JE. Meteorological variables and malaria in a Chinese temperate city: A twenty-year time-series data analysis. Environ Int. 2010;36(5):439-45.

[20]Tao Y, Liu ZM, Mi SQ, Song J, Qiang L. Effects of meteorological factors on other types of infectious diarrhea. J Lanzhou Univ: Nat Sci. 2015;51(5):646-651.(in Chinese)

[21]Wang J, Xu MM, Mo YZ, Pan XC. Correlation between meteorological factors and infectious diarrhea in a district of Beijing. J Environ Health. 2013;30(11):991-995.(in Chinese)

[22]Fu JG, Shi C, Xu C, Lin Q, Zhang J, Yi QH, et al. Outbreaks of acute gastroenteritis associated with a re-emerging GII.P16-GII.2 norovirus in the spring of 2017 in Jiangsu, China. PLoS One. 2017;12(12):e0186090.

[23]Ma T, Zhang M, Hong L, Wang X, Dai WJ, Wu ZW, et al. Outbreak investigation of acute gastroenteritis associated with G□.P7-G□.6 norovirus in a primary school of Nanjing in 2017. Mod Pre Med. 2018;45(22):4188-4191.(in Chinese)

[24]Tang JW. The effect of environmental parameters on the survival of airborne infectious agents. J R Soc Interface. 2009;6(6):737-746.

Figures

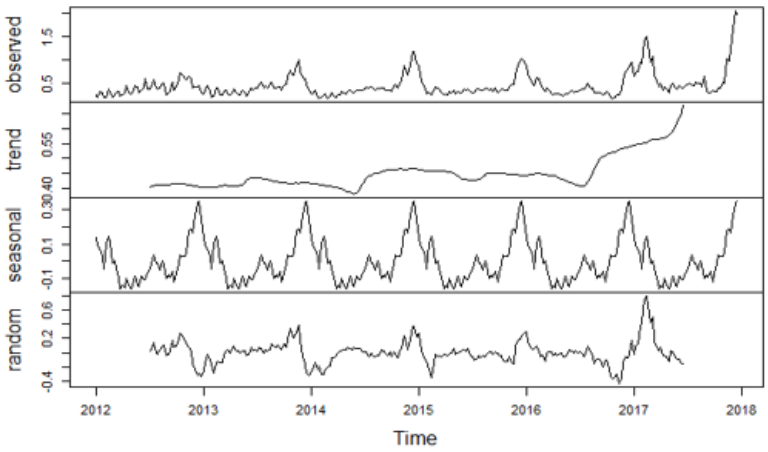

Figure 1

Weekly observed cases of infectious diarrhea in Jiangsu Province, 2012–2017 Note: From top to bottom, the lines represent actually observed cases, trend components, seasonal components, and random components, respectively.

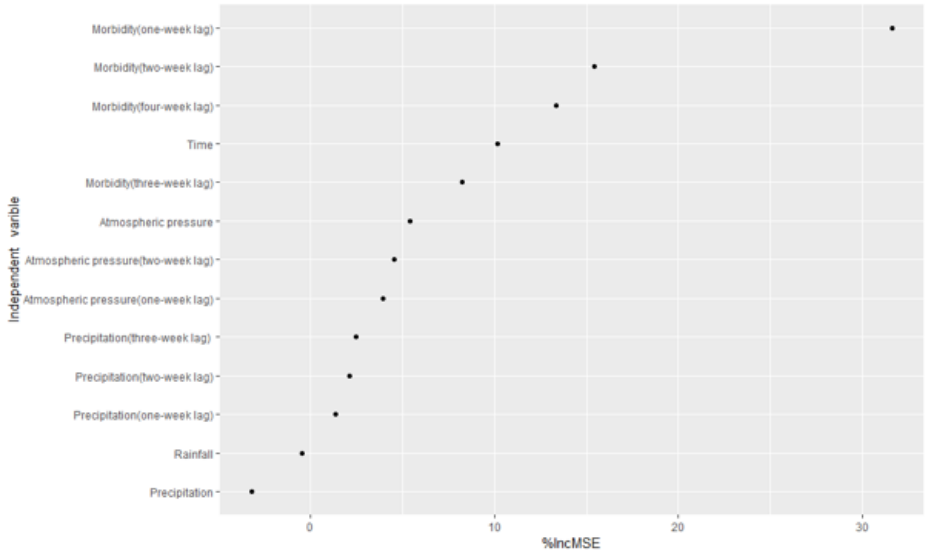

Figure 2

Variable importance in random forest regression model for infectious diarrhea

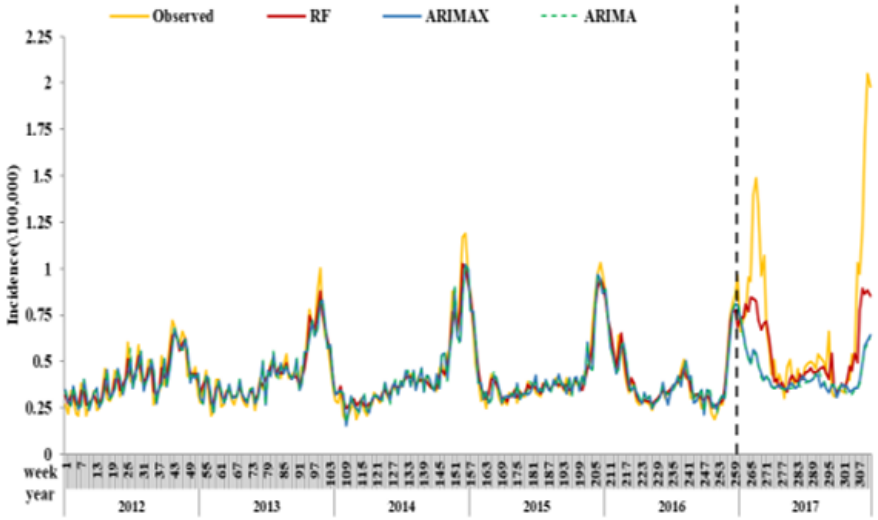

Figure 3

Observed infectious diarrhea incidences and values predicted by different models Note: Left side of the vertical line is model fitting stage, and the right is forecasting stage.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Calculations.jpg