


RESEARCH ARTICLE

Open Access



# Forecasting incidence of hand, foot and mouth disease using BP neural networks in Jiangsu province, China

Wendong Liu<sup>1\*†</sup>, Changjun Bao<sup>1†</sup>, Yuping Zhou<sup>1†</sup>, Hong Ji<sup>1</sup>, Ying Wu<sup>1</sup>, Yingying Shi<sup>1</sup>, Wenqi Shen<sup>1</sup>, Jing Bao<sup>2</sup>, Juan Li<sup>2</sup>, Jianli Hu<sup>1</sup> and Xiang Huo<sup>1</sup>

## Abstract

**Background:** Hand, foot and mouth disease (HFMD) is a rising public health problem and has attracted considerable attention worldwide. The purpose of this study was to develop an optimal model with meteorological factors to predict the epidemic of HFMD.

**Methods:** Two types of methods, back propagation neural networks (BP) and auto-regressive integrated moving average (ARIMA), were employed to develop forecasting models, based on the monthly HFMD incidences and meteorological factors during 2009–2016 in Jiangsu province, China. Root mean square error (RMSE) and mean absolute percentage error (MAPE) were employed to select model and evaluate the performance of the models.

**Results:** Four models were constructed. The multivariate BP model was constructed using the HFMD incidences lagged from 1 to 4 months, mean temperature, rainfall and their one order lagged terms as inputs. The other BP model was fitted just using the lagged HFMD incidences as inputs. The univariate ARIMA model was specified as ARIMA (1,0,1)(1,1,0)<sup>12</sup> (AIC = 1132.12, BIC = 1440.43). And the multivariate ARIMAX with one order lagged temperature as external predictor was fitted based on this ARIMA model (AIC = 1132.37, BIC = 1142.76). The multivariate BP model performed the best in both model fitting stage and prospective forecasting stage, with a MAPE no more than 20%. The performance of the multivariate ARIMAX model was similar to that of the univariate ARIMA model. Both performed much worse than the two BP models, with a high MAPE near to 40%.

**Conclusion:** The multivariate BP model effectively integrated the autocorrelation of the HFMD incidence series. Meanwhile, it also comprehensively combined the climatic variables and their hysteresis effects. The introduction of the climate terms significantly improved the prediction accuracy of the BP model. This model could be an ideal method to predict the epidemic level of HFMD, which is of great importance for the public health authorities.

**Keywords:** Hand, foot and mouth disease, Forecasting, BP neural networks

## Background

Hand, foot and mouth disease (HFMD) is a common infectious disease caused by various enteroviruses, especially enterovirus 71 (EV-71) and Coxsackie virus A16 (CV-A16) [1]. Recently, enteroviruses other than EV-71 and CV-A16 have been increasing in both mild and severe cases and Coxsackie virus A6 (CV-A6) has been

emerging as another predominant serotype in some regions [2]. HFMD mostly affects children under 5 years of age. It is a rising public health problem and has attracted considerable attention worldwide [3–5]. It is especially widespread in Asia-Pacific areas [1] and presents a general increasing incidence in recent decades [6–8]. China is one of the Southeast Asian countries with the most serious HFMD epidemics [9]. It is a Class C notifiable infectious disease in China. Since 2009, the annual incidence of HFMD has never been less than 100 per 100,000, and caused hundreds of deaths in each year.

\* Correspondence: [jscdclwd@sina.cn](mailto:jscdclwd@sina.cn)

<sup>†</sup>Wendong Liu, Changjun Bao and Yuping Zhou contributed equally to this work.

<sup>1</sup>Jiangsu Province Center for Diseases Control and Prevention, Nanjing, China  
Full list of author information is available at the end of the article



Effective prevention and control of HFMD has become a major challenge in the field of public health [1, 10].

Incidence forecasting of an infectious disease is essential for the public health authorities to better understand the epidemic characteristics and track its seasonal changes in advance. Accurate predicating is a vital basis to optimize decisions and configure resources for preventing and controlling infectious diseases. So, it is of great significance to establish a scientific, appropriate and reliable prediction model and improve the model performance to the best [11, 12]. Recently, some researchers are interested in forecasting the incidence of HFMD, using the liner time series models. For example, an ARIMA (1,0,1, 0,1,0)<sup>12</sup> model was constructed to forecast the HFMD incidence in Sichuan, China [13]. In another study, multivariable ARIMA models using search engine query data and climate factors as exogenous variables were developed to predict the HFMD epidemic in Guangdong, China [14]. However, the assumption of linearity in many time series events may not be satisfied in practice. The accuracy of the liner forecasting models therefore needs to be improved. Models based on artificial neural networks (ANN) can effectively extract nonlinear relationships in data. They have been widely used in infectious diseases predictions because of their characteristics of robustness, fault tolerance, and adaptive learning ability. As one of the common ANN, back propagation neural networks (BP model) is widely used in many areas, such as economic and engineering. It has also been introduced into forecasting infectious diseases [15, 16]. To date, however, there has been no literature report on using BP model to predict the epidemic of HFMD.

The purpose of this study was to develop an optimal BP model to predict the future trend of HFMD in Jiangsu province, China, with special emphasis on elucidating the effects of meteorological factors as predictors. Meanwhile, the performance of BP model was compared with ARIMA model. It was expected that the findings in this work would be useful for the prevention and control of HFMD.

## Methods

### Data sources

The monthly case numbers of HFMD in Jiangsu province during 2009–2016 were obtained from the National Notifiable Disease surveillance System (<http://www.cdpc.chinacdc.cn>). The demographic data were collected from the Jiangsu provincial statistics department. And the monthly meteorological data were gained from Jiangsu Meteorological Service Center. The meteorological variables used in this study included rainfall (RF), sunshine duration (SD), relative humidity (RH), atmospheric pressure (AP), minimum temperature (MIN\_T), mean temperature (MEAN\_T), maximum temperature (MAX\_T) and wind velocity (WV).

### BP neural networks

ANN is a family of intelligent methods that mimic the biological neural networks. BP model is one of the most common ANN, developed by Rumelhart and McClelland in 1986 [17]. Since the distinguish performance, BP model has been popularly used in many practical fields including public health [16, 18]. Typical BP model has a three-layer network construction, consisting of an input layer, a hidden layer and an output layer. Each layer consists of a number of neuron nodes. The upper layer and lower layer nodes are connected by the connection weights.

BP model is trained with a back-propagation algorithm, in which the external input information at the input nodes is propagated forward to calculate the outputs. Then the error between the predicted values and the target outputs is propagated backward to modify the connection weights and thresholds. BP model training includes three steps: (1) the forward feeding of the input training pattern, (2) the calculation and back-propagation of the associated error, and (3) the adjustment of the weights and thresholds. Given  $n$  nodes in the input layer,  $m$  nodes in the hidden layer and one node in the output layer, the outputs of each node in the hidden layer and the output layer are calculated according to the following formulas:

$$net_j = f\left(\sum_{i=0}^n \omega_{ij}x_i + b_j\right) \quad (i = 0, 1, \dots, n; j = 1, 2, \dots, m)$$

$$\hat{y} = f\left(\sum_{j=1}^m \omega_j net_j + b\right) \quad (j = 1, 2, \dots, m)$$

In the formulas,  $net_j$  is the output of the  $j^{\text{th}}$  node in the hidden layer,  $\omega_{ij}$  denotes the connection weight between input node  $i$  and hidden node  $j$ ,  $x_i$  the  $i^{\text{th}}$  input,  $b_j$  the threshold of hidden node  $j$ ,  $\hat{y}$  the output of the last layer (i.e., the predicting value),  $\omega_j$  the connection weight between hidden node  $j$  and the output node,  $b$  the threshold of output node,  $f$  the activation function of a node which is usually a sigmoid function as follow:

$$f(x) = \frac{1}{1 + \exp(-x)}$$

### ARIMA model

Auto-regressive integrated moving average (ARIMA), also called Box-Jenkins model, is a traditional method to study the time series data [19]. ARIMA model deals with non-stationary time series with a differencing process based on ARMA model. As an extension of ARIMA model, seasonal ARIMA has both non-seasonal and seasonal components [19, 20]. It is denoted as ARIMA(p,d,q)(P,D,Q)<sup>s</sup> in which  $p$ ,  $d$ ,  $q$  indicate orders of non-seasonal auto-regression (AR), differencing and moving average (MA) terms;  $P$ ,  $D$ ,  $Q$  are orders of

seasonal AR, differencing and MA, respectively; the superscript *s* indicates seasonal period (*s* = 12 in this study). The process of fitting ARIMA model involves four stages. First, the original time series is transformed by logarithmic algorithm, difference or seasonal difference to achieve stationarity. Second, autocorrelation function (ACF) and partial autocorrelation function (PACF) of the stationary time series are calculated and plotted to identify the initial *p*, *q*, *P*, *Q* parameters. Alternative ARIMA models are established with different model parameters. Third, Akaike information criterion (AIC) and Bayesian information criterion (BIC) are conducted to access the goodness-of-fit of the ARIMA models, the one with the minimum AIC and BIC values is considered as the optimal model. Fourth, Box-Ljung test for the residual series of the optimal model is conducted to determine if the residual series is white noise sequence (*p* > 0.05). Finally, prospective prediction is conducted using the optimal model.

Since the incidence time series of HFMD commonly shows significant cyclical and seasonal patterns [21, 22], ARIMA model was considered to erect the benchmark model. Once the univariate ARIMA model was selected, the multivariate ARIMA model including climate factors as external regressors [23] was further developed. In this study, ARIMA model that incorporates climate factors was referred as ARIMAX.

**Model evaluation**

Four models were fitted in this study, BP model with meteorological variables, BP model without meteorological variables, ARIMA model and ARIMAX model. The data between 2009 and 2014 were used as training set to fit models, and data between 2015 and 2016 were used as testing set to evaluate the forecasting accuracy of different models. Root mean square error (RMSE) and mean absolute percentage error (MAPE) were selected as the measures to evaluate the performance of the models, which were calculated as the following formulas:

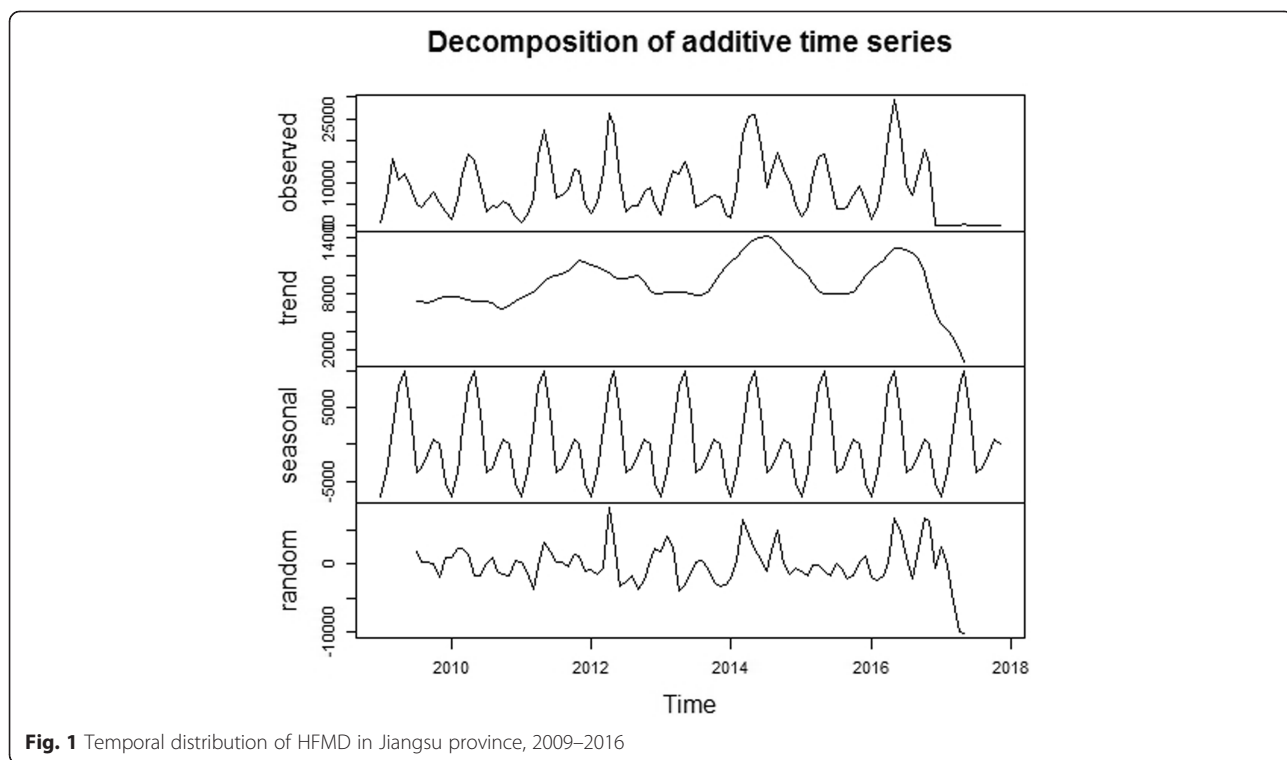
$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{y_t}$$

where *n* means number of real data or predicted values, *y<sub>t</sub>* means real data and *ŷ<sub>t</sub>* means predicted value.

**Statistical software**

All statistical analyses were completed using R software version 3.5.0. Particularly, ARIMA models were performed with R package “forecast” version 8.5. Meanwhile, BP models were constructed with R package “nnet” version 7.3–12.



**Fig. 1** Temporal distribution of HFMD in Jiangsu province, 2009–2016

**Table 1** Spearman correlation coefficients between HFMD and meteorological factors in Jiangsu province, 2009–2016

	HFMD	RF	SD	RH	AP	MIN_T	MEAN_T	MAX_T
RF	0.334*							
SD	0.129	-0.142						
RH	0.167	0.752*	-0.404*					
AP	-0.400*	-0.723*	-0.330*	-0.471*				
MIN_T	0.409*	0.741*	0.300*	0.608*	-0.943*			
MEAN_T	0.410*	0.719*	0.334*	0.576*	-0.947*	0.998*		
MAX_T	0.409*	0.686*	0.387*	0.526*	-0.949*	0.990*	0.996*	
VV	0.045	0.151	0.099	-0.241*	-0.175	-0.033	-0.035	-0.029

\*:  $p < 0.05$

**Results**

**General description**

Totally 917,285 cases were detected during 2009–2016 in Jiangsu province, China, reaching an average annual incidence rate of 145.39 per 100,000. As shown in Fig. 1, the incidence presented no long-term trend in the 8 years. However, there was a distinct seasonality, and two incidence peaks were observed in each year, the higher occurred between April and June, the lower occurred between November and December.

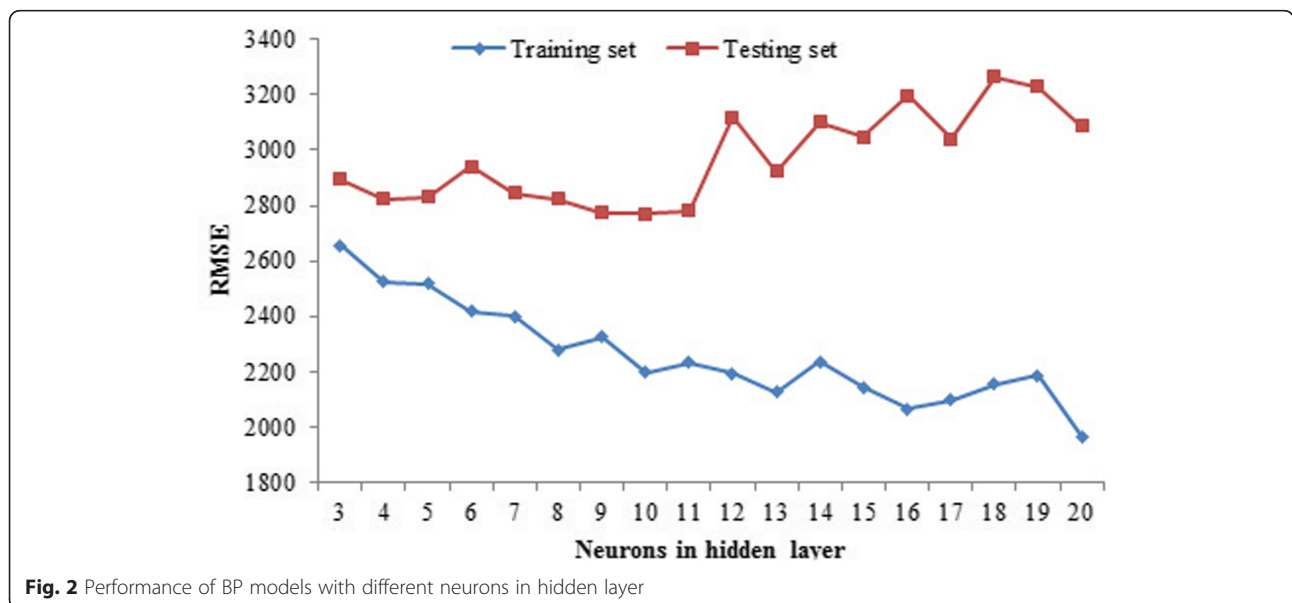
Univariate spearman correlation analysis indicated that all the meteorological factors were significantly associated with the incidence of HFMD, except sunshine duration, relative humidity and wind velocity. Notably, strong correlations were detected among mean temperature, maximum temperature, minimum temperature and atmospheric pressure, with correlation coefficients  $> 0.9$ . See Table 1. To avoid multicollinearity, just monthly mean temperature and rainfall were considered in the following models. Further cross-correlation analysis indicated that both mean temperature and rainfall significantly related

with the incidence of HFMD at lag 1, with correlation coefficients of 0.235 ( $p = 0.0216$ ) and 0.251 ( $p = 0.0146$ ) respectively. These lagged terms were also considered in the following modelling.

**Model fitting**

**Multivariate BP model**

According to the results of autocorrelation analysis, the HFMD time series presented significant autocorrelation at lag 1–4. Given this, its four lagged terms were considered as predictors in the BP model. Ultimately, eight variables, including the monthly case numbers lagged from 1 to 4 months ( $\times 1-\times 4$ ), the monthly mean temperature and rainfall ( $\times 5, \times 6$ ), and the one order lagged mean temperature and rainfall ( $\times 7, \times 8$ ), were taken as inputs of the BP model. The current monthly case number was taken as output of the model. To determine the number of neurons in the hidden layer, 18 BP models with different neurons in the hidden layer were built, and RMSE was employed to evaluate their performance. As shown in Fig. 2, the model performed better on the training set



**Fig. 2** Performance of BP models with different neurons in hidden layer

**Table 2** Selection of the univariate ARIMA model

Model	AIC	BIC	Ljung-Box test	
			Q <sup>a</sup>	p
ARIMA(1,0,1)(1,1,0) <sup>12</sup>	1132.12	1140.43	10.233	0.5272
ARIMA(1,0,0)(1,1,0) <sup>12</sup>	1135.73	1141.96	17.418	0.1432
ARIMA(2,0,1)(1,1,0) <sup>12</sup>	1133.77	1144.16	10.187	0.4422
ARIMA(2,0,0)(1,1,0) <sup>12</sup>	1134.98	1143.29	13.207	0.2944
ARIMA(3,0,1)(1,1,0) <sup>12</sup>	1134.63	1147.09	9.118	0.4455
ARIMA(3,0,0)(1,1,0) <sup>12</sup>	1132.76	1143.15	8.6807	0.5811
ARIMA(4,0,1)(1,1,0) <sup>12</sup>	1136.61	1151.15	8.4172	0.4136
ARIMA(4,0,0)(1,1,0) <sup>12</sup>	1134.64	1147.11	8.6288	0.4917

<sup>a</sup>Q denotes the statistics of Ljung-Box test

with more neurons in the hidden layer, which means the more neurons in the hidden layer, the better goodness-of-fit. The prospective prediction accuracy reached the best on the testing set when the number of neurons in the hidden layer was 11. Accordingly, the best BP model structure was determined as 8–11-1, which means there were 8 nodes in the input layer, 11 nodes in the middle layer and one node in the output layer.

#### Univariate BP model

To assess the effects of meteorological factors on the forecasting of HFMD, a BP model without climate variables was constructed, just using the cases numbers lagged from 1 to 4 months ( $\times 1-\times 4$ ) as the model inputs. The model fitting process was the same as the former BP model. And the best model structure was determined as 4–4-1.

#### Univariate ARIMA model

The original time series of HFMD achieved stationary after one order seasonal differencing. As shown in Table 2, eight alternative univariate ARIMA models were constructed. The results of Ljung-Box test suggest that all the residual series of these models were white noise sequences. Based on the AIC and BIC, the best-fitting model was determined to be ARIMA(1,0,1)(1,1,0)<sup>12</sup>, with a minimum AIC = 1132.12 and a minimum BIC = 1140.43.

#### Multivariate ARIMAX model

Mean temperature ( $\times 5$ ), rainfall ( $\times 6$ ) and their one order lagged terms ( $\times 7$  and  $\times 8$ , respectively) were added into the optimal univariate ARIMA(1,0,1)(1,1,0)<sup>12</sup> model as exogenous variables, individually or in combination. Accordingly, 15 multivariate ARIMAX models were fitted. As shown in Table 3, these models were all statistically significant (Ljung-Box test  $p > 0.05$ ). ARIMA(1,0,1)(1,1,0)<sup>12</sup> with one order lagged

**Table 3** Selection of the multivariate ARIMAX model

Model	AIC	BIC	Ljung-Box test	
			Q <sup>a</sup>	p
ARIMA+ $\times 5$	1134.05	1144.43	10.679	0.4005
ARIMA+ $\times 6$	1134.08	1144.47	10.274	0.4347
ARIMA+ $\times 7$	1132.37	1142.76	10.982	0.3759
ARIMA+ $\times 8$	1134.07	1144.45	10.272	0.4348
ARIMA+ $\times 5 + \times 6$	1135.97	1148.43	10.903	0.2985
ARIMA+ $\times 5 + \times 7$	1134.01	1146.47	10.406	0.3357
ARIMA+ $\times 5 + \times 8$	1135.97	1148.43	10.85	0.3023
ARIMA+ $\times 6 + \times 7$	1134.21	1146.68	11.149	0.2811
ARIMA+ $\times 6 + \times 8$	1136.06	1148.52	10.28	0.3456
ARIMA+ $\times 7 + \times 8$	1134.00	1146.46	11.144	0.2815
ARIMA+ $\times 5 + \times 6 + \times 7$	1135.93	1150.48	10.502	0.2468
ARIMA+ $\times 5 + \times 7 + \times 8$	1135.62	1150.16	10.325	0.2587
ARIMA+ $\times 5 + \times 6 + \times 8$	1137.94	1152.49	10.926	0.2201
ARIMA+ $\times 6 + \times 7 + \times 8$	1135.99	1150.53	11.157	0.2066
ARIMA+ $\times 5 + \times 6 + \times 7 + \times 8$	1137.61	1154.23	10.294	0.1861

<sup>a</sup>Q denotes the statistics of Ljung-Box test

temperature as external predictor was the optimal ARIMAX model, with a minimum AIC = 1132.37 and a minimum BIC = 1142.76.

#### Prediction performance comparison

The comparison of the models was summarized in Table 4, and the predicting outputs were displayed in Fig. 3. The multivariate BP model performed the best in both model fitting stage and prospective forecasting stage. The predicted values matched the real HFMD incidences very well, with a MAPE no more than 20%, which suggested that this model would be able to accurately estimate the prevalence and seasonal fluctuation of HFMD. The BP model without climate variables had much higher RMSE and MAPE than the multivariate BP model on both training set and testing set. Its forecasting performed well in 2015 but presented a high error in 2016. The performance of the multivariate ARIMAX model was similar to that of the univariate ARIMA model. Both performed much worse than the two BP models. They could not accurately predict the real data in the study area, with a high MAPE near to 40%.

#### Discussion

Accurately identifying the epidemic trend in advance is of critical importance for infectious diseases prevention and control. As HFMD is a common infectious disease throughout the world, modeling its epidemic has been concerned and actively studied in recent years. Some researchers have put forward different prediction methods

**Table 4** Comparison of the four models

model	RMSE		MAPE	
	Training set	Testing set	Training set	Testing set
Multivariate BP	2125.68	2653.73	16.59	18.57
Univariate BP	3055.32	2981.11	20.46	28.55
ARIMA(1,0,1)(1,1,0) <sup>12</sup> + x <sup>a</sup>	3313.60	4476.06	26.11	36.43
ARIMA(1,0,1)(1,1,0) <sup>12</sup>	3377.48	4476.39	28.02	36.67

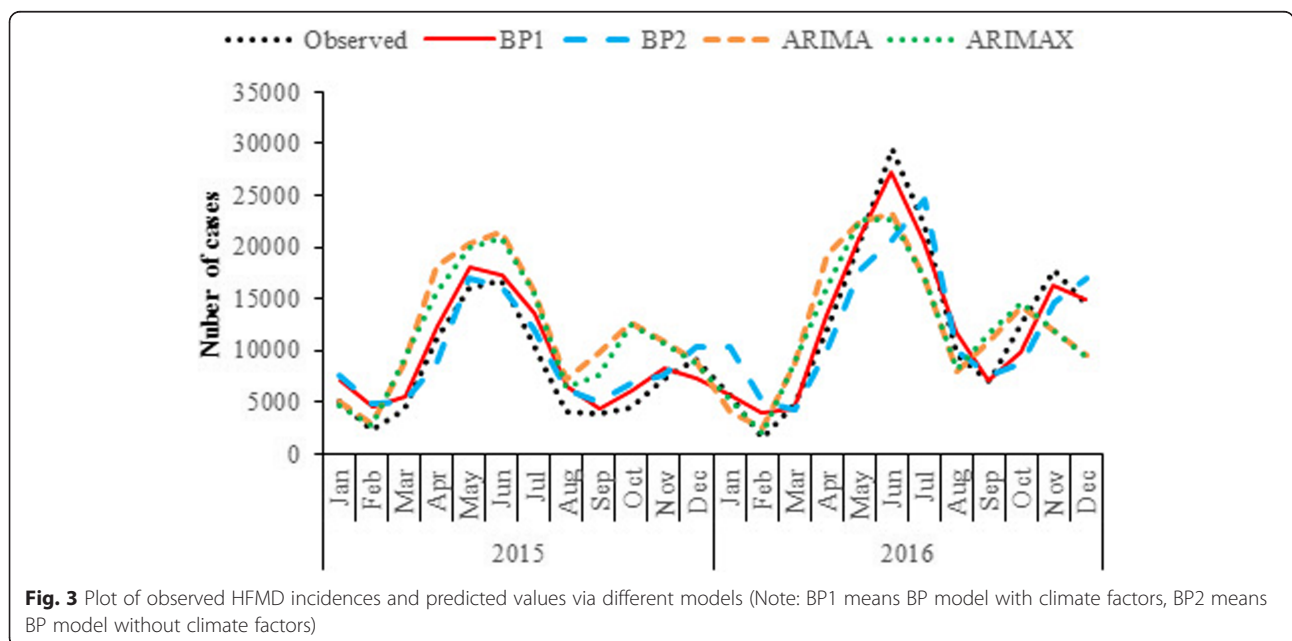
<sup>a</sup>x7 means one order lagged mean temperature

for HFMD. For examples, Yu et al. [24], developed a new hybrid model with ARIMA and nonlinear auto-regressive neural network. Zhong et al. [25], employed XGBoost, one of the machine learning methods, to forecast HFMD with multiple environmental factors.

In this study, an optimized multivariate BP model with meteorological factors was constructed. This model presented a satisfactory accuracy in forecasting the HFMD incidences in Jiangsu province, China. It reached a MAPE less than 20% in the prospective forecasting stage and accurately estimated the seasonal fluctuation of HFMD in the next 24 months. The predictive performance is much better than that in many similar studies. It may serve as a reliable tool for the public health authorities in the practice of HFMD prevention and control. Notably, BP model has a risk of over-fitting, which is a critical issue that usually leads to poor generalization [26]. In this work, it was observed that the accuracy of prospective prediction getting worse when the neurons in the hidden layer was more than 11, which suggested that too many neurons in hidden layer maybe causes severe over-fitting. Hence, how to determine an optimized model structure is an important issue. Unfortunately, it

is still controversial. In this study, this work was completed based on MAPE, the BP model with the minimum MAPE on testing set was selected as the best optimal model.

A substantial studies have proposed that infectious diseases are climate sensitive [27–29]. Climatic factors may influence the survival and spread of infectious pathogens in the environment, the host susceptibility and exposure probability [30–32]. The effects of meteorological factors, such as temperature, rainfall and relative humidity, on the epidemics of HFMD have attracted considerable concern recently [22, 33]. Song et al. [34], developed a seasonal ARIMA model with lagged precipitation as predictor to forecast the incidence of HFMD. Unfortunately, the model did not present a satisfactory performance. Similarly, the ARIMAX model we developed using lagged temperature as predictor did not achieve a good enough accuracy for practical application. And the introduced climate variable did not improve the performance of the ARIMA model. It may be due to two reasons. Firstly, ARIMA model is essentially a linear method. However, meteorological factors were proved to be nonlinearly associated with the epidemic of HFMD, so



**Fig. 3** Plot of observed HFMD incidences and predicted values via different models (Note: BP1 means BP model with climate factors, BP2 means BP model without climate factors)

ARIMA model is inappropriate to fit the relationships between predictors and HFMD incidence. Besides, data size is not sufficient for the model to fully extract the underline pattern contained in data. Consequently, the model could not achieve a satisfied predictions. Zhao et al. [35], also constructed an ARIMA model with temperature as predictor based on the data in Huainan City, China. It presented a well goodness-of-fit. However, its extrapolated predictive capability was not explained. Consequently, its practical application value is doubtful. In this study, BP neuron network was employed to forecast the HFMD incidence with monthly average temperature, rainfall and their lagged terms as predictors. This model performed much better than the BP model without climatic variables, which suggested that climate factors can improve the prediction effect. Meanwhile, we also found that both of the two BP models performed much better than the ARIMA models. This may indicate that BP model is more suitable than ARIMA model to predict the HFMD incidence in the study region.

It is worth mentioning that the multivariate BP model developed in this study achieved accurate estimations of the HFMD incidences in the next 24 months. Thus, it could be used to predict the medium to long term epidemic level of HFMD, which is of great important for the public health authorities. As shown in Fig. 3, given the whole test set, the BP model without climate variables performed relatively poor. Interestingly, the first few predicted values matched the real incidences very well. It suggested that this model may have the potential to be used for short-term forecasting, which is necessary to be further verified in practice.

Some limitations need to be mentioned. First, the epidemic of HFMD is affected by many factors, including natural and social environmental factors, etiological factors, and so on. In this study, just meteorological variables were considered to improve the predication ability. Other factors associated with HFMD may also be used as good predictors, which deserves progressive studies. Second, because some mild cases might use home therapies, and some cases with atypical symptoms may be misdiagnosed, so the data reported may underestimate the HFMD incidence, which may affect the precision of the predictions. Third, the optimal BP model was constructed based on the data in Jiangsu province, China, generalizability of our findings to other regions with different epidemic characteristics of HFMD and climate situations might not be straightforward. But the use of the BP model incorporating climate factors in the detection and prediction of HFMD may provide an opportunity for re-allocating healthcare resources more efficiently in other regions or countries. Besides, similar to many other neural network models, BP

model can not explain the specific association between risk factors and disease.

## Conclusion

In this study, four models were constructed to forecast the incidence of HFMD in Jiangsu province, China. The BP models performed much better than the ARIMA models. The introduction of mean temperature, rainfall and their one order lagged terms significantly improved the prediction accuracy of the BP model. On the contrary, neither the univariate ARIMA model nor the multivariate ARIMAX model achieved satisfactory prediction accuracy. The climate factors did not optimize the performance of the ARIMA model. In general, the multivariate BP model comprehensively combined the autocorrelation of the independent, the climatic variables and their hysteresis effects. It is an ideal method to predict the HFMD epidemic, which has a good prospect of practical application.

## Abbreviations

ACF: Auto-correlation function; AIC: Akaike information criterion; ANN: Artificial neural networks; AP: Atmospheric pressure; ARIMA: Auto-regressive integrated moving average; ARIMAX: ARIMA model that incorporates climate factors; BIC: Bayesian information criterion; BP: Back propagation neural networks; CV-A16: Coxsackie virus A16; CV-A6: Coxsackie virus A6; EV-71: Enterovirus 71; HFMD: Hand, foot and mouth disease; MAPE: Mean absolute percentage error; MAX\_T: Maximum temperature; MEAN\_T: Mean temperature; MIN\_T: Minimum temperature; PACF: Partial auto-correlation function; RF: Rainfall; RH: Relative humidity; RMSE: Root mean square error; SD: Sunshine duration; WV: Wind velocity

## Acknowledgements

Not applicable.

## Authors' contributions

Conceived and designed the experiments: WL, CB, YZ. Performed the experiments: WL, JH, XH. Analyzed the data: WL, HJ, YW, YS, WS. Contributed reagents/materials/analysis tools: JB, JL, YS. Wrote the paper: WL, BC, YZ. All authors read and approved the final manuscript.

## Funding

This study was supported by Research project of Jiangsu Provincial natural science foundation (No. BK20151595), Jiangsu Provincial Medical Youth Talent (No. QNRC2016542) and Jiangsu Provincial Major Science & Technology Demostation Project (No. BE2017749). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Availability of data and materials

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

## Ethics approval and consent to participate

This study was approved by the Ethics Committee at Jiangsu Provincial Center for Disease Control and Prevention, China (No. JSCDC20190301). As the HFMD incidence was statistical summary data, it was not necessary to obtain informed consent.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Jiangsu Province Center for Diseases Control and Prevention, Nanjing, China. <sup>2</sup>Jiangsu Meteorological Service Center, Nanjing, China.

Received: 23 January 2019 Accepted: 10 September 2019

Published online: 07 October 2019

### References

- Xing W, Liao Q, Viboud C, Zhang J, Sun J, Wu JT, et al. Hand, foot, and mouth disease in China, 2008–12: an epidemiological study. *Lancet Infect Dis*. 2014;14(4):308–18. [https://doi.org/10.1016/s1473-3099\(13\)70342-6](https://doi.org/10.1016/s1473-3099(13)70342-6).
- Li Y, Chang ZR, Wu P, Liao QH, Liu FF, Zheng YM, et al. Emerging enteroviruses causing hand, foot and mouth disease, China, 2010–2016. *Emerg Infect Dis*. 2018;24(10):1902–6. <https://doi.org/10.3201/eid2410.171953> PubMed PMID: WOS:000444801900015.
- Ho MT, Chen ER, Hsu KH, Twu SJ, Chen KT, Tsai SF, et al. An epidemic of enterovirus 71 infection in Taiwan. *New Engl J Med*. 1999;341(13):929–35. <https://doi.org/10.1056/Nejm199909233411301> PubMed PMID: WOS:000082658100001.
- Chan KP, Goh KT, Chong CY, Teo ES, Lau GKK, Ling AE. Epidemic hand, foot and mouth disease caused by human enterovirus 71, Singapore. *Emerg Infect Dis*. 2003;9(1):78–85 PubMed PMID: WOS:000180503300012.
- Zhang Y, Zhu Z, Yang WZ, Ren J, Tan XJ, Wang Y, et al. An emerging recombinant human enterovirus 71 responsible for the 2008 outbreak of hand foot and mouth disease in Fuyang city of China. *Virology*. 2010;7:94. <https://doi.org/10.1186/1743-422X-7-94> PubMed PMID: WOS:000279550400001.
- Fujimoto T, Chikahira M, Yoshida S, Ebira H, Hasegawa A, Totsuka A, et al. Outbreak of central nervous system disease associated with hand, foot, and mouth disease in Japan during the summer of 2000: detection and molecular epidemiology of enterovirus 71. *Microbiol Immunol*. 2002;46(9):621–7 PubMed PMID: 12437029.
- Ang LW, Koh BK, Chan KP, Chua LT, James L, Goh KT. Epidemiology and control of hand, foot and mouth disease in Singapore, 2001–2007. *Ann Acad Med Singap*. 2009;38(2):106–12 PubMed PMID: 19271036.
- Koh WM, Badaruddin H, La H, Chen MI, Cook AR. Severity and burden of hand, foot and mouth disease in Asia: a modelling study. *BMJ Glob Health*. 2018;3(1):e000442. <https://doi.org/10.1136/bmjgh-2017-000442> PubMed PMID: 29564154; PubMed Central PMCID: PMC5859810.
- Huang J, Liao Q, Ooi MH, Cowling BJ, Chang Z, Wu P, et al. Epidemiology of recurrent hand, foot and mouth disease, China, 2008–2015. *Emerg Infect Dis*. 2018;24(3). <https://doi.org/10.3201/eid2403.171303> PubMed PMID: 29460747; PubMed Central PMCID: PMC5823341.
- Yang B, Liu F, Liao Q, Wu P, Chang Z, Huang J, et al. Epidemiology of hand, foot and mouth disease in China, 2008 to 2015 prior to the introduction of EV-A71 vaccine. *Euro Surveill*. 2017;22:50. <https://doi.org/10.2807/1560-7917.ES.2017.22.50.16-00824> PubMed PMID: 29258646; PubMed Central PMCID: PMC5743100.
- Zhang X, Liu Y, Yang M, Zhang T, Young AA, Li X. Comparative study of four time series methods in forecasting typhoid fever incidence in China. *PLoS One*. 2013;8(5):e63116. <https://doi.org/10.1371/journal.pone.0063116> PubMed PMID: 23650546; PubMed Central PMCID: PMC3641111.
- Wu W, Guo J, An S, Guan P, Ren Y, Xia L, et al. Comparison of two hybrid models for forecasting the incidence of hemorrhagic fever with renal syndrome in Jiangsu Province, China. *PLoS One*. 2015;10(8):e0135492. <https://doi.org/10.1371/journal.pone.0135492> PubMed PMID: 26270814; PubMed Central PMCID: PMC4536138.
- 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. *Epidemiol Infect*. 2016;144(1):144–51. <https://doi.org/10.1017/S0950268815001144> PubMed PMID: 26027606.
- Du Z, Xu L, Zhang W, Zhang D, Yu S, Hao Y. Predicting the hand, foot, and mouth disease incidence using search engine query data and climate variables: an ecological study in Guangdong, China. *BMJ Open*. 2017;7(10):e016263. <https://doi.org/10.1136/bmjopen-2017-016263> PubMed PMID: 28988169; PubMed Central PMCID: PMC5640051.
- Ren H, Li J, Yuan ZA, Hu JY, Yu Y, Lu YH. The development of a combined mathematical model to forecast the incidence of hepatitis E in Shanghai, China. *BMC Infect Dis*. 2013;13:421. <https://doi.org/10.1186/1471-2334-13-421> PubMed PMID: 24010871; PubMed Central PMCID: PMC3847129.
- Wang Y, Li J, Gu J, Zhou Z, Wang Z. Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China). *Appl Soft Comput*. 2015;35:280–90. <https://doi.org/10.1016/j.asoc.2015.05.047>.
- Du K, Swamy M. *Neural networks and statistical learning*. London: Springer; 2014.
- Gan R, Chen X, Yan Y, Huang D. Application of a hybrid method combining grey model and back propagation artificial neural networks to forecast hepatitis B in China. *Comput Math Methods Med*. 2015;2015:328273. <https://doi.org/10.1155/2015/328273> PubMed PMID: 25815044; PubMed Central PMCID: PMC4357037.
- Box G, Jenkins G, Reinsel G. *Time series analysis: forecasting and control*. Hoboken: Wiley; 2008.
- Lara-Ramirez EE, Rodriguez-Perez MA, Perez-Rodriguez MA, Adeleke MA, Orozco-Algarra ME, Arrendondo-Jimenez JJ, et al. Time series analysis of onchocerciasis data from Mexico: a trend towards elimination. *PLoS Negl Trop Dis*. 2013;7(2):e2033. <https://doi.org/10.1371/journal.pntd.0002033> PubMed PMID: 23459370; PubMed Central PMCID: PMC3573083.
- Liu W, Ji H, Shan J, Bao J, Sun Y, Li J, et al. Spatiotemporal dynamics of hand-foot-mouth disease and its relationship with meteorological factors in Jiangsu Province, China. *PLoS One*. 2015;10(6):e0131311. <https://doi.org/10.1371/journal.pone.0131311> PubMed PMID: 26121573; PubMed Central PMCID: PMC4488144.
- Tian L, Liang F, Xu M, Jia L, Pan X, Clements ACA. Spatio-temporal analysis of the relationship between meteorological factors and hand-foot-mouth disease in Beijing, China. *BMC Infect Dis*. 2018;18(1):158. <https://doi.org/10.1186/s12879-018-3071-3> PubMed PMID: 29614964; PubMed Central PMCID: PMC5883540.
- Soebiyanto RP, Adimi F, Kiang RK. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. *PLoS One*. 2010;5(3):e9450. <https://doi.org/10.1371/journal.pone.0009450> PubMed PMID: 20209164; PubMed Central PMCID: PMC2830480.
- Yu L, Zhou L, Tan L, Jiang H, Wang Y, Wei S, et al. Application of a new hybrid model with seasonal auto-regressive integrated moving average (ARIMA) and nonlinear auto-regressive neural network (NARNN) in forecasting incidence cases of HFMD in Shenzhen, China. *PLoS One*. 2014;9(6):e98241. <https://doi.org/10.1371/journal.pone.0098241> PubMed PMID: 24893000; PubMed Central PMCID: PMC4043537.
- Zhong R, Wu Y, Cai Y, Wang R, Zheng J, Lin D, et al. Forecasting hand, foot, and mouth disease in Shenzhen based on daily level clinical data and multiple environmental factors. *Biosci Trends*. 2018;12(5):450–5. <https://doi.org/10.5582/bst.2018.01126> PubMed PMID: 30473551.
- Schittenkopf C, Deco G, Brauer W. Two strategies to avoid overfitting in feedforward networks. *Neural Netw*. 1997;10(3):12.
- Xiang J, Hansen A, Liu Q, Liu X, Tong MX, Sun Y, et al. Association between dengue fever incidence and meteorological factors in Guangzhou, China, 2005–2014. *Environ Res*. 2017;153:17–26. <https://doi.org/10.1016/j.envres.2016.11.009> PubMed PMID: 27883970.
- Azage M, Kumie A, Worku A, A CB, Anagnostou E. Effect of climatic variability on childhood diarrhea and its high risk periods in northwestern parts of Ethiopia. *PLoS One*. 2017;12(10):e0186933. <https://doi.org/10.1371/journal.pone.0186933> PubMed PMID: 29073259; PubMed Central PMCID: PMC5658103.
- Semenza JC, Suk JE, Estevez V, Ebi KL, Lindgren E. Mapping climate change vulnerabilities to infectious diseases in Europe. *Environ Health Perspect*. 2012;120(3):385–92. <https://doi.org/10.1289/ehp.1103805> PubMed PMID: 22113877; PubMed Central PMCID: PMC3295348.
- Lafferty, KD. The ecology of climate change and infectious diseases. *Ecology*. 2009;90(4):888–900. <https://doi.org/10.1890/08-0079.1>.
- Semenza JC, Menne B. Climate change and infectious diseases in Europe. *Lancet Infect Dis*. 2009;9(6):365–75. [https://doi.org/10.1016/S1473-3099\(09\)71004-5](https://doi.org/10.1016/S1473-3099(09)71004-5) PubMed PMID: 19467476.
- The Lancet Infectious D. Climate change: the role of the infectious disease community. *Lancet Infect Dis*. 2017;17(12):1219. [https://doi.org/10.1016/s1473-3099\(17\)30645-x](https://doi.org/10.1016/s1473-3099(17)30645-x).
- Qi H, Chen Y, Xu D, Su H, Zhan L, Xu Z, et al. Impact of meteorological factors on the incidence of childhood hand, foot, and mouth disease (HFMD) analyzed by DLNMs-based time series approach. *Infect Dis Poverty*. 2018;7(1):7. <https://doi.org/10.1186/s40249-018-0388-5> PubMed PMID: 29391070; PubMed Central PMCID: PMC5796399.



34. Song Y, Wang F, Wang B, Tao S, Zhang H, Liu S, et al. Time series analyses of hand, foot and mouth disease integrating weather variables. *PLoS One*. 2015;10(3):e0117296. <https://doi.org/10.1371/journal.pone.0117296> PubMed PMID: 25729897; PubMed Central PMCID: PMC4346267.
35. Zhao D, Wang L, Cheng J, Xu J, Xu Z, Xie M, et al. Impact of weather factors on hand, foot and mouth disease, and its role in short-term incidence trend forecast in Huainan City, Anhui Province. *Int J Biometeorol*. 2017;61(3):453–61. <https://doi.org/10.1007/s00484-016-1225-9> PubMed PMID: 27557791.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

