Forecasting Hand-Foot-and-Mouth Disease Cases Using a Wavelet-based SARIMA-NNAR Hybrid Model

Gongchao Yu
The Fifth Affiliated Hospital of Zhengzhou University https://orcid.org/0000-0003-0937-4685

Huifen Feng (huifen.feng@163.com)
The Fifth Affiliated Hospital of Zhengzhou University https://orcid.org/0000-0002-5790-1294

Shuang Feng
The Fifth Affiliated Hospital of Zhengzhou University

Jing Zhao
The Fifth Affiliated Hospital of Zhengzhou University

Jing Xu
The Fifth Affiliated Hospital of Zhengzhou University

Research article

Keywords: Hand-foot-and-mouth disease, enterovirus, HFMD, Artificial neural network

Posted Date: March 30th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-19355/v1

Version of Record: A version of this preprint was published at PLOS ONE on February 5th, 2021. See the published version at https://doi.org/10.1371/journal.pone.0246673.

Abstract

Background: Hand-foot-and-mouth disease(HFMD) is one of the most common diseases in children, which has high morbidity. Reliable forecasting is significant for prevention and control. Recently, hybrid models have been becoming popular and wavelet analysis has been widely used. Better prediction accuracy may be achieved with wavelet-based hybrid models. Thus, our aim is to forecast number of HFMD cases with wavelet-based hybrid models.

Materials and methods: We fitted a wavelet-based SARIMA(seasonal autoregressive integrated moving average)-NNAR(neural network nonlinear autoregressive) hybrid model with HFMD weekly cases from 2009 to 2016 in Zhengzhou, China. At the same time, single SARIMA model, simplex NNAR model and pure SARIMA-NNAR hybrid model were established as well for comparison and estimation.

Results: The wavelet-based SARIMA-NNAR hybrid model had an excellent performance whether in fitting or in forecasting compared to other models. Its fitted and forecasting time series were approximate to the actual observed time series.

Conclusions: This wavelet-based SARIMA-NNAR hybrid model that we fitted is suitable for forecasting number of HFMD cases. It will facilitate prevention and control of HFMD.

Introduction

Hand-foot-and-mouth disease is an acute infectious disease caused by enterovirus, which is mostly among young children[1]. The most cases are mild and self-limited with symptoms of fever and herpes(or rash)on the hands, feet and mouth[2]. However, few of children may have severe complications, such as meningitis, brainstem encephalitis, neurogenic pulmonary oedema, pulmonary haemorrhage, and circulatory failure[1-2]. There is no effective treatment of HFMD [3], so prevention and control are especially important. Although EV71 vaccine was introduced and the number of cases of EV71 and CA16 decreased but other enteroviruses gradually increased[1][4]. The incidence of HFMD still remained high[4-6]. Active early intervention is important. If accurate forecasting can be made, we are able to response in advance, which facilitates decreasing the incidence and reducing the disease burden. So, reliable forecasting is extremely important for the prevention and control of it.

Many scholars used all kinds of models to forecast the incidence of HFMD. Among these models, the traditional ARIMA model is utilized widely[7-9]. Linearity is the necessary condition of its application. However, time series in real-world are often uncertain and complex[10], especially the epidemic time series[11]. It may contain both linear and nonlinear structures[12-13]. The ability of SARIMA to fit non-stationary time series is limited[14-15]. A research[16], which compared the performance of SARIMA model and Back-Propagation neural networks, demonstrated that SARIMA is inferior to BP neural networks. Some practical studies have shown that its prediction effect is worse than that of hybrid models combined with neural network[13][17].

Artifical neural network(ANN), which is adaptive and nonlinear[18-19], are appropriate for excavating the nonlinear relationships in time series[18]. Due to their powerful nonlinear mapping ability, they are thought to be able to achieve any required accuracy[10]. Among them, nonlinear autoregressive neural network(NARNN), one of the dynamic neural networks, is suitable for time series forecasting[14-15] due to their dynamic property and high fault tolerance performance[20]. Some scholars also call it neural network nonlinear autoregressive(NNAR) model. However, some scholars think that the ANN model cannot extract linear patterns of data as well as nonlinear[21]. Just using ANN model alone may not be the best solution for real-world time series[15].

In recent years, combined models have been surging in order to avoid shortcomings of single models and improve prediction accuracy[22-23]. It is a common case that SARIMA models are combined with ANN models[14][24-25]. This combination is also applied in HFMD forecasting[26], in which SARIMA fits linear relationships and ANN fits nonlinear relationships. Such combination can make use of the unique strength of both models adequately and improve forecasting accuracy. However, some researches[14][21] argued that hybrid model does not necessarily outperform its constituents' performances.

The models discussed above did well in time series forecasting, but they are not absolutely perfect. We still need to explore better forecasting models. Wavelet analysis has been used as a data preprocessing method and combined with other forecasting models in environmental science[27], hydrology[18] and financial time series[28]. It does not require stationarity of time series, which is often the basic requirement of traditional methods[11]. So, it's suitable for non-stationary and noisy signal processing[29]. In some researches[18][30], wavelet analysis decomposes original series into approximation component and detail components. The approximation component, which is similar to original data but more smooth, was used to construct the SARIMA model. While the detail components, which are high-frequency and may contain noise, were utilized to establish ANN model. Then sum the forecst from SARIMA model and ANN model to get the final forecasting results. The results of the studies indicated that this wavelet-based combined model is superior to single models[30]. But, this kind of model is not used to forecast HFMD cases untill now.

In view of the above situation, we proposed to fit a wavelet-based SARIMA-NNAR hybrid model to forecast number of HFMD cases. We are expected that this model is suitable for forecasting HFMD cases and facilitate the prevention and control of it.

Materials And Methods

Data collection and processing

The weekly number of HFMD cases was obtained from Zhengzhou Center for Disease Control and Prevention, China. There is no missing data. We divided the data into training set and validation set. In the training set , the weekly number of HFMD cases during 2009 to 2015 were used to fit models. While in the validation set, the weekly number of HFMD cases in 2016 were used to estimate the performance of

models. We plotted the time series and used “stl” function in “stats” package of “R” software to decompose the time series for the purpose of exploring its trend, seasonality and error.

Establishing SARIMA model

Autoregressive integrated moving average(ARIMA) model is one of the mostly used models to forecast number of cases of infectious diseases. If a seasonal component is included, we call it seasonal autoregressive integrated moving average(SARIMA) model. Generally, it is denoted as SARIMA(p, d, q)(P, D, Q)s, where p is the order of autoregressive(AR) model, d the number of difference, q the order of moving average(MA) model, P the order of AR seasonal model, D the number of seasonal difference, Q the order of MA seasonal model, and s the length of the seasonal period. Due to the obvious seasonality of the HFMD cases and one-year period, a SARIMA model was constructed and s was set to 52 weeks.

Firstly, stationarity is necessary to fit a SARIMA model. Augmented Dickey-Fuller (ADF) unit root test is frequently used to test the stationarity. Differencing and seasonal differencing are often used to transform it into a stationary series. Second, select the order of model according to the autocorrelation function(ACF) and partial autocorrelation function(PACF). Then, select an optimal model based on the Akaike information criterion (AIC) and Bayesian information criterion (BIC) and estimate the model parameters. Finally, examine the residuals with ACF, PACF and Box-Ljung test. The residuals are supposed to be white noise and have no autocorrelation.

We used the “auto.arima” function in “forecast” package of “R” software to fit models with different values of p, d, q, P, D and Q. Then, the best model will be selected by minimizing AICc. AICc is the corrective AIC. Given the same value of d and D, the minimum AICc corresponds to the best model.

Building NNAR model

Artificial neural networks are based on mathematical models of the brain. The basic structure includes input layer, hidden layers and output layer. In this research, we use neural network nonlinear autoregressive(NNAR) model[31], which is a feed-forward neural network with a single hidden layer and uses lagged values of the time series as inputs. It is denoted as $NNAR(p,P,k)_m$ model, where p is the number of inputs lags, P seasonal lags, k the number of nodes in the hidden layer and m the length of seasonal period. The “nnetar” function in “forecast” package of “R” software can automatically find optimal parameter p, P, and k. For seasonal time series, the default values are P=1 and p is chosen from the optimal linear model fitted to the seasonally adjusted data. If k is not specified, it is set to k= (p+P+1)/2(rounded to the nearest integer)[31].

Constructing SARIMA-NNAR combined model

In the first place, a SARIMA model was fitted. Then, its residual series were inputted to NNAR model. The nonlinear relationships that the residuals may contain can be mined adequately by neural networks. The

final combined forecasting values of the time series were the sum of predictions from SARIMA model and adjusted residuals from NNAR model.

Formulating wavelet based SARIMA-NNAR hybrid model

We chose discrete wavelet transfromatiom(DWT),which is often used in time series decomposition[30]. There are different wavelet, such as Daubechies, Coiflets and Symlets. Via consulting references[18][30], we chose a Daubechies wavelet, which is denoted as “db2” in MATLAB and one or two decomposition levels. The “db2” wavelet was used to decompose the original data into approximation component and detail components in different levels(one or two). The approximation component is low-frequency and similar to original data but more smooth. The detail components are high-frequency which usually contain noise. Afterwards, a SARIMA model was fitted to the approximation component while a NNAR model was fitted to the detail components. Final results were computed via summing the results from SARIMA model and NNAR model. Wavelet decomposition and reconstruction were performed with MATLAB software(Version R2014a).

Evaluate the performance of four models

The models fitted with training set were used to forecast forward 52 weeks. Every week’s number of cases were forecasted on the basis of value before. Three indices were computed to measure the accuracy of fitness and forecasting for the four models:the root mean square error(RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The formulas for calculation are as follows.

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

$$\text{MAE:}\frac{1}{n}\sum_{t=1}^{n}\left|y_t-\hat{y}_t\right|$$

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

Here, y_t means the observed time series at time t and $\hat{y}_t$ means the fitted or forecast time series.

Results

General information

There are totally 128682 cases reported in Zhengzhou, China from 2009 to 2016. The peaks of cases often occurred between May and July. The time series plot of weekly HFMD cases is showed in Fig.1. We can see obvious seasonality from the results of “stl” decomposition. There is one-year period in HFMD prevalence.

The best-performing SARIMA model

The ADF test implied that the time series is nonstationary(P=0.4451). After one difference and one seasonal difference, the time series became stationary(ADF test P=0.04478)(Fig.2). Using the “auto.arima” funtion with d=1 and D=1, the optimal model was selected as SARIMA $(1,1,3)(0,1,1)_{52}$ with the lowest AICc of 3668.13. The residuals plot, the corresponding ACF plot and a histogram are displayed in Fig.3. The Ljung-Box test of the residuals, which P value is 0.6359, demonstrated that the residuals have no autocorrelation.

The best-performing NNAR model

Due to the seasonality, value of 1 was set to P. Along with automatic selection as $(6,1,4)_{52}$ by ”nnetar” function, we also tried different p values from one to ten(Table 1). Taking into consideration three indices of performance both in training set and validation set, we finally selected NNAR$(8,1,5)_{52}$ as the optimal model. The residuals was displayed in Fig.4. The Ljung-Box test, which P value is 0.5917, demonstrated that the residuals have no autocorrelation.

The best-performing SARIMA-NNAR model

The SARIMA model was fitted as explained before. The optimal NNAR model which used the residual series generated from the SARIMA model was selected as NNAR$(1,1,2)_{52}$. The residuals are showed in Fig.5. The Ljung-Box test, which P value is 0.5199, demonstrated that the residuals have no autocorrelation.

The best-performing wavelet-based hybrid SARIMA-NNAR model

The “db2” wavelet decomposed the original data into approximate component(cA) and detail component(cD) in one level or two levels(Table 2). The performance of wavelet-based hybrid model using different decomposition levels were exhibited in Table 3. From the perspective of training set, two level of decomposition did better, while as for validation set, one level of decomposition exceeded. In view that the purpose of model is forecasting, we regarded one level as better decomposion level.

Compare accuracy among four kinds of models

The performance of four kinds of models in training set and validation set were showed in Table 4. When it comes to training set, the RMSE, MAE and MAPE of single NNAR model are lowest, followed by wavelet-based hybrid model. As for validation set, the indices of wavelet-based hybrid model are lowest in most situation except for MAE and MAPE of single NNAR model.

The fitted and forecasting time series plot of four models were displayed in Fig.6. In terms of training set, all models fitted well, while the fitted time series by NNAR model was especially approximate to original data. As far as validation set is concerned, the peak of wave forecasted by single SARIMA model and regular SARIMA-NNAR model were lower than that of actual observed data. That forecasted by NNAR

model may deviate from real time series slightly. It seems that the wavelet-based hybrid model may performed better in validation set compared to other models.

Discussion

In this paper, a wavelet-based SARIMA-NNAR hybrid model was fitted to forecast the number of HFMD cases with data from 2009 to 2016 in Zhengzhou, China. In order to estimate its performance, we compared it with single SARIMA model, simplex NNAR model and pure SARIMA-NNAR hybrid model. We found that the wavelet-based hybrid model had an excellent performance whether in fitting or in forecasting compared to other models. Its fitted and forecasting time series were approximate to the actual observed time series. This wavelet-based SARIMA-NNAR hybrid model is suitable for forecasting number of HFMD cases. It will facilitate prevention and control of HFMD.

As a new and powerful data preprocessing method, wavelet analysis is capable of decomposing original data into approximation component and detail components in different levels. Through allowing different components to be forecasted by different models and summing the results as the eventual consequence, this wavelet-based SARIMA-NNAR hybrid model facilitates accuracy improvement of forecasting.

However, there still are some flaws in this study. Firstly, the original data were collected from Zhengzhou Center for Disease Prevention and Control, which may have the possibility of false reporting and omissive reporting. The quality of data may influence the construction process and performance of model to some degree. Secondly, there are different kinds of wavelet and different levels of wavelet decomposition. Moreover, different detail componets can be modeled respectively. We just chose ”db2” wavelet and one or two levels of decomposition. With two level of decomposition, we got two detail components but just synthesize them to one total detail component via wavelet reconstruction. We didn’t try different wavelet, more level of decomposition and different methods to deal with detail components. More trials maybe yield better results. Thirdly, this hybrid model should be updated in time to maintain it’s accuracy with new data. At last, the influencing factors of HFMD are complex and a variety of factors should be taken into account to predict, such as climate[31] and transmission dynamic[32].

Conclusions

In this research, a wavelet-based SARIMA-NNAR hybrid model was fitted to forecast number of HFMD cases with weekly data from 2009 to 2016 in Zhengzhou, China. The wavelet-based hybrid model had an excellent performance whether in fitting or in forecasting compared to other models. This wavelet-based SARIMA-NNAR hybrid model is suitable for forecasting number of HFMD cases. It will facilitate prevention and control of HFMD.

Declarations

Acknowledgements

We thank the Zhengzhou Center for Disease Control and Prevention.

Authors' contributions

GY and HF conceived and designed the study. GY, SF, JZ and JX collected and organized the data. GY were in charge of statistical analysis and wrote the manuscript. HF revised and approved the final version of the manuscript. All authors read and approved the final submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (grant number 81473030) and the Henan Province Special Project of Key Research, Development and Promotion (grant number 192102310376).

Availability of data and materials

All data generated or analyzed during this study are included in this published article. The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The weekly number of HFMD cases were obtained from Zhengzhou Center for Disease Control and Prevention, China. There is not personal information.So, it is not nessary to get ethics approval and consent.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. **Qiu J, *et al***. The clinical and epidemiological study of children with Hand, Foot, and Mouth Disease in Hunan, China from 2013 to 2017. *Scientific Reports* 2019; **9**: 11662.
2. **Zhuang ZC, *et al***. Epidemiological research on Hand, Foot, and Mouth Disease in mainland China. *Viruses* 2015; **7**: 6400-6411.
3. **Li X, *et al*.** Improving the efficacy of conventional therapy by adding Andrographolide Sulfonate in the treatment of severe Hand, Foot, and Mouth Disease: a randomized controlled trial. *Evidence-Based Complementary and Alternative Medicine*. Published online: 15 January 2015. doi:10.1155/2013/316250.

4. **Wang J, *et al***. Epidemiological characteristics of hand, foot, and mouth disease in Shandong, China, 2009–2016. *Scientific Reports*. Published online: 21 August 2017. doi:10.1038/s41598-017-09196-z.
5. **Li J, *et al***. The surveillance of the epidemiological and serotype characteristics of hand, foot, mouth disease in Neijiang city, China, 2010-2017: A retrospective study. *PLoS ONE.* Published online: 6 June 2019. doi:10.1371/journal.pone.0217474.
6. **Qi L, *et al***. Epidemiological characteristics and spatial-temporal distribution of Hand, Foot, and Mouth Disease in Chongqing, China, 2009-2016. *International Journal of Environmental Research and Public Health* 2018; **15**: 270.
7. **Liu L, *et al***. Predicting the incidence of hand, foot and mouth disease in Sichuan province, China using the ARIMA model. *Epidemiology and Infection* 2015; **144**: 144-151.
8. **Liu S, *et al***. Predicting the outbreak of hand, foot, and mouth disease in Nanjing, China: a time-series model based on weather variability. *International Journal of Biometeorology* 2018; **62**:565-574.
9. **Tian CW, Wang H, Luo XM**. Time-series modelling and forecasting of hand, foot and mouth disease cases in China from 2008 to 2018. *Epidemiology and Infection* 2019; **147**: e82.
10. **Wang KW, *et al***. Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. *Epidemiology and Infection* 2017; **145**: 1118-1129.
11. **Cazelles B, *et al***. Time-dependent spectral analysis of epidemiological time-series with wavelets. *Journal of the Royal Society, Interface* 2007; **4**: 625-636.
12. **Yolcu U, Egrioglu E, Aladag CH**. A new linear & nonlinear artificial neural network model for time series forecasting. *Decision Support Systems* 2013; **54**: 1340-1347.
13. **Purwanto, Eswaran C, Logeswaran R**. An enhanced hybrid method for time series using linear and neural network models. *Applied Intelligence* 2012; **37**: 511-519.
14. **Zhou L, *et al***. Time series model for forecasting the number of new admission inpatients. *BMC Medical Informatics and Decision Making* 2018; **18**: 39.
15. **Zhou L, *et al***. A hybrid model for predicting the prevalence of schistosomiasis in humans of Qianjiang City, China. *PLoS ONE*. Published online: 13 August 2014. doi:10.1371/journal.pone.0104875.
16. **Liu W, *et al***. Forecasting incidence of hand, foot and mouth disease using BP neural networks in Jiangsu province, China. *BMC Infectious Diseases* 2019; **19**: 828.
17. **Zhang GP**. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 2003; **50**: 159-175.
18. **Shafaei M, *et al***. A wavelet-SARIMA-ANN hybrid model for precipitation forecasting. *Journal of Water and Land Development* 2016; **28**: 27-36.
19. **Grossi E, Buscema M**. Introduction to artificial neural networks. *European Journal of Gastroenterology & Hepatology* 2007; **19**: 1046-1054.
20. **Kaastra I, Boyd M**. Design a neural network for forecasting financial and economic series. *Neurocomputing* 1996; **10**: 215-236.

21. **Taskaya-Temizel T, Casey M**. A comparative study of autoregressive neural network hybrids. *Neural Networks* 2005; **18**: 781-789.
22. **Zhang GP**. A neural network ensemble method with jitterred training data for time series forecasting. *Information Sciences* 2007; **177**: 5329-5346.
23. **Zou JJ, *et al*.** Application of a combined model with seasonal autoregressive integrated moving average and support vector regression in forecasting hand-foot-mouth disease incidence in Wuhan, China. *Medicine* 2019; **98**: e14195.
24. **Wang H, *et al*.** Time-series analysis of tuberculosis from 2005 to 2017 in China. *Epidemiology and Infection* 2018; **146**: 935-939.
25. **Wang Y, Shen Z, Jiang Y**. Comparison of autoregressive integrated moving average model and generalised regression neural network model for prediction of haemorrhagic fever with renal syndrome in China: a time-series study. *BMJ Open* 2019; **9**: e025773.
26. **Yu L, *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**: e98241.
27. **Wang F, *et al*.** Chlorophyll a simulation in a lake ecosystem using a model with wavelet analysis and artificial neural network. *Environmental Management* 2013; **51**: 1044-1054.
28. **Jin J, Kim J**. Forecasting natural gas prices using wavelets, time series, and artificial neural networks. *PLoS ONE* 2015; **10**: e0142064.
29. **Cazelles B, *et al*.** Wavelet analysis of ecological time series. *Oecologia* 2008; **156**: 287-304.
30. **Wang Y, *et al*.** Time series modeling of pertussis incidence in China from 2004 to 2018 with a novel wavelet based SARIMA-NAR hybrid model. *PLoS ONE* 2018; *13*: e0208404.
31. **Feng H, *et al*.** Time series analysis of Hand-Foot-Mouth Disease hospitalization in Zhengzhou: establishment of forecasting models using climate variables as predictors. *PLoS ONE* 2014; **9**: e87916.
32. **Takahashi S, *et al*.** Hand, Foot, and Mouth Disease in China: modeling epidemic dynamics of enterovirus serotypes and implications for vaccination. *PLoS Medicine* 2016; **13**: e1001958.

Tables

Table 1 Accuracy of ten candidate NNAR models in the training set and validation set

NNAR model	Training set			Validation set		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
(1,1,2)52	74.17	42.83	21.33	471.67	295.87	57.88
(2,1,2)52	64.66	38.75	22.91	515.23	355.63	68.88
(3,1,2)52	62.15	35.95	23.07	455.17	315.41	62.93
(4,1,3)52	53.84	32.13	21.46	394.90	274.97	64.58
(5,1,4)52	46.57	28.85	21.68	349.08	243.48	57.56
(6,1,4)52	45.94	28.36	20.65	359.88	262.90	62.48
(7,1,4)52	45.71	28.30	21.91	332.62	231.62	54.41
(8,1,5)52	37.90	23.77	19.53	309.44	211.96	57.48
(9,1,6)52	34.15	21.77	19.28	323.15	215.22	52.52
(10,1,6)52	33.58	20.56	19.67	344.24	230.52	55.00

Table 2 Models for approximate component and detail component and P value of Ljung-Box test for residuals

Level of decomposition	SARIMA for cA	P value(cA)	NNAR for cD	P value(cD)
One	(4,1,0)(1,1,0)52	0.1476	(16,1,9)52	0.0546
Two	(4,1,1)(1,1,1)52	0.1198	(9,1,6)52	0.5681

cA: approximate component; cD: detail component. P value(cA): P value of Box.test of residuals from SARIMA for cA. P value(cD): P value of Box.test of residuals from NNAR for cD.

Table 3 Accuracy of the optimal models with one or two wavelet decomposition level.

Level of decomposition	Training set			Validation set		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
One	71.29	37.45	21.81	296.18	227.25	61.32
Two	56.67	28.46	19.12	314.34	240.74	61.45

Table 4 Accuracy of the training set and 52 weeks forecasting in validation set.

	Training set			52 weeks		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
SARIMA	77.38	46.27	32.96	307.02	238.45	64.01
NNAR	37.90	23.77	19.53	309.44	211.96	57.48
SARIMA+NNAR	78.60	51.03	36.81	304.33	236.84	65.58
Wavelet hybrid	71.29	37.45	21.81	296.18	227.25	61.32

SARIMA+NNAR: regular SARIMA-NNAR hybrid model; Wavelet hybrid: wavelet-based SARIMA-NNAR hybrid model.

Figures

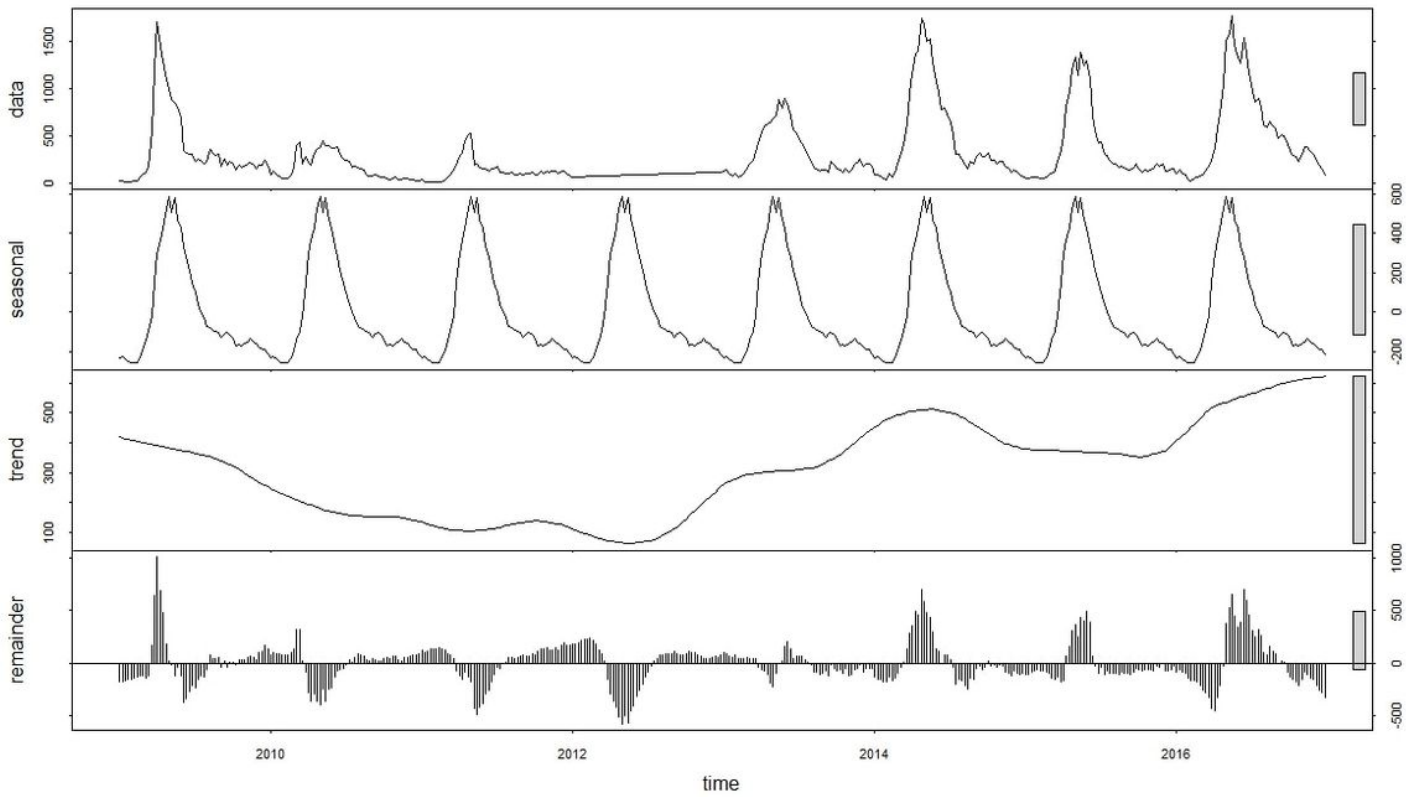

Figure 1

Time series plot of weekly HFMD cases from 2009 to 2016 in Zhengzhou, China.

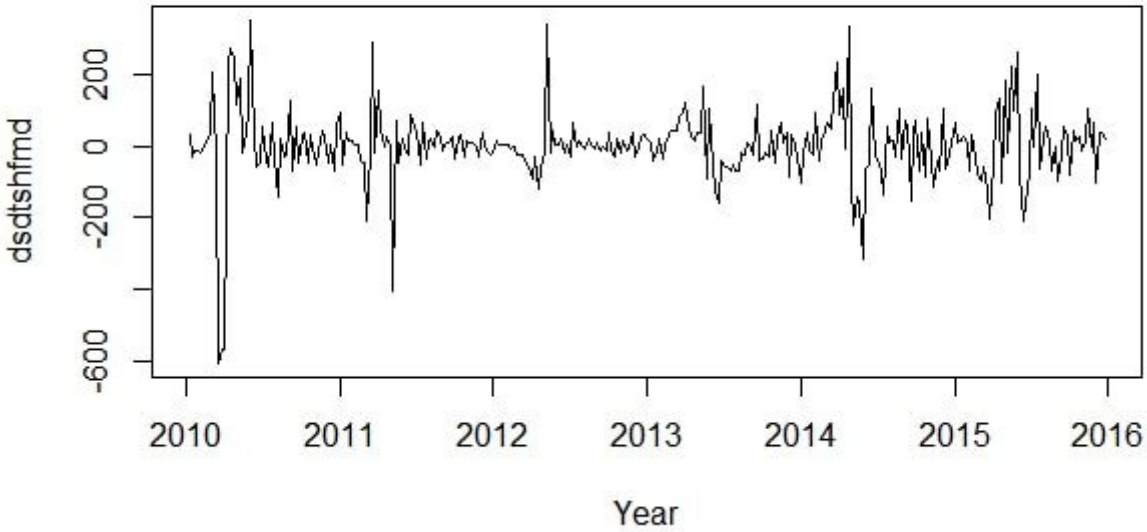

Figure 2

Regular differenced and seasonal differenced time series plot.

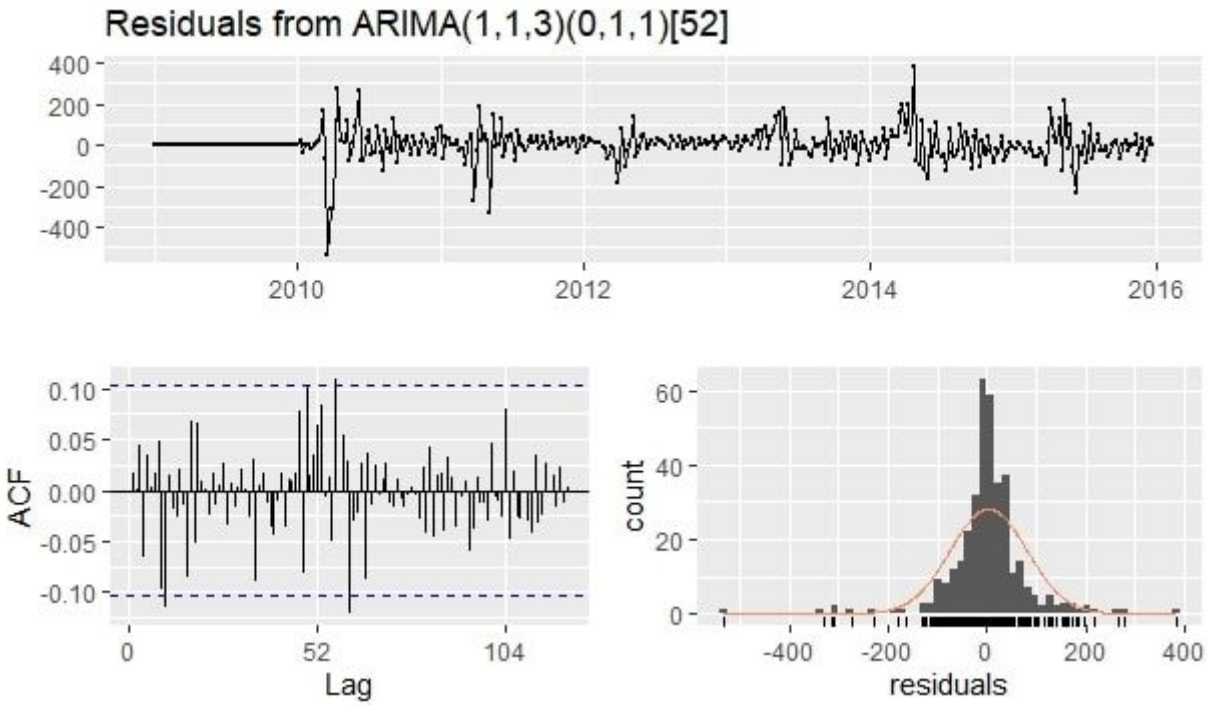

Figure 3

The residuals plot, the corresponding ACF plot and a histogram from ARIMA(1,1,3)(0,1,1)52.

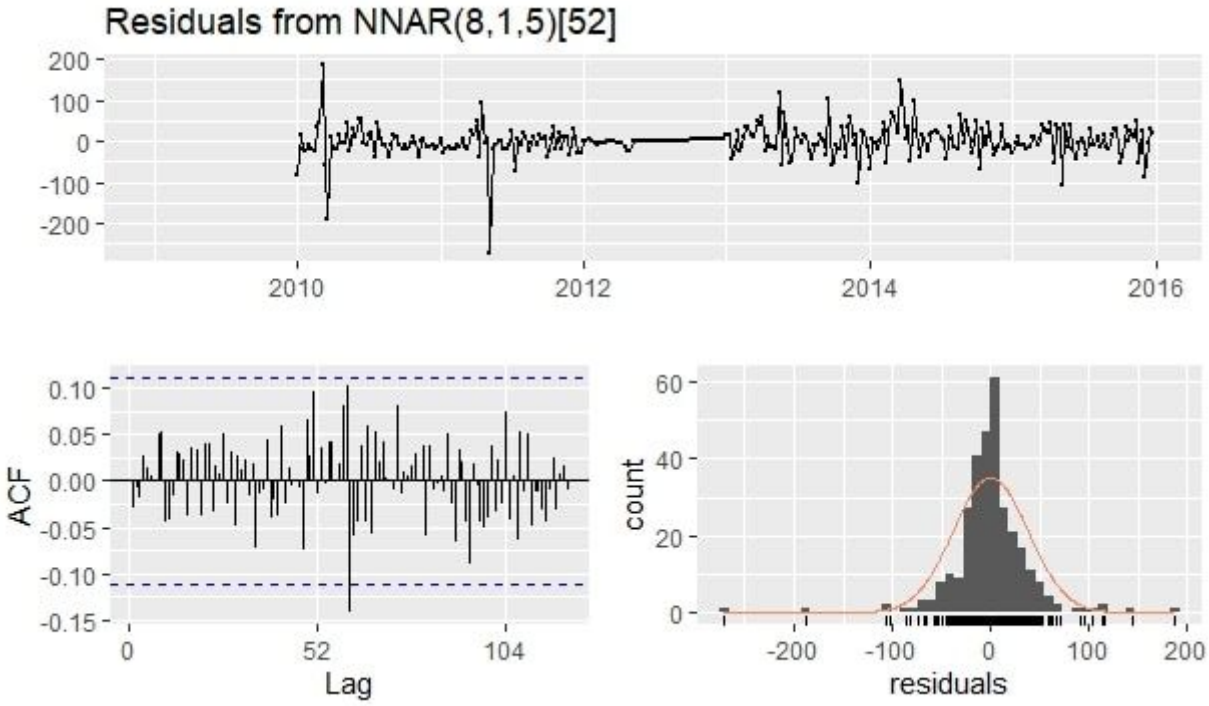

Figure 4

The residuals plot, the corresponding ACF plot and a histogram from NNAR(8,1,5)52.

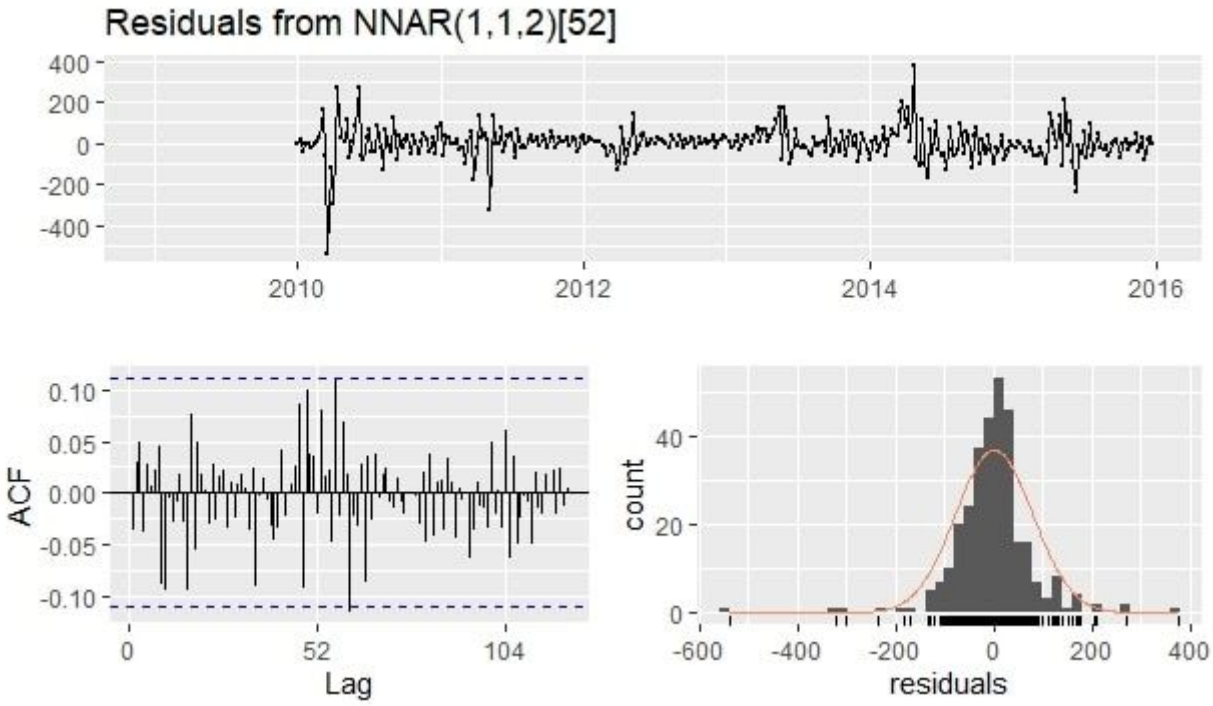

Figure 5

The residuals plot, the corresponding ACF plot and a histogram from NNAR(1,1,2)52.

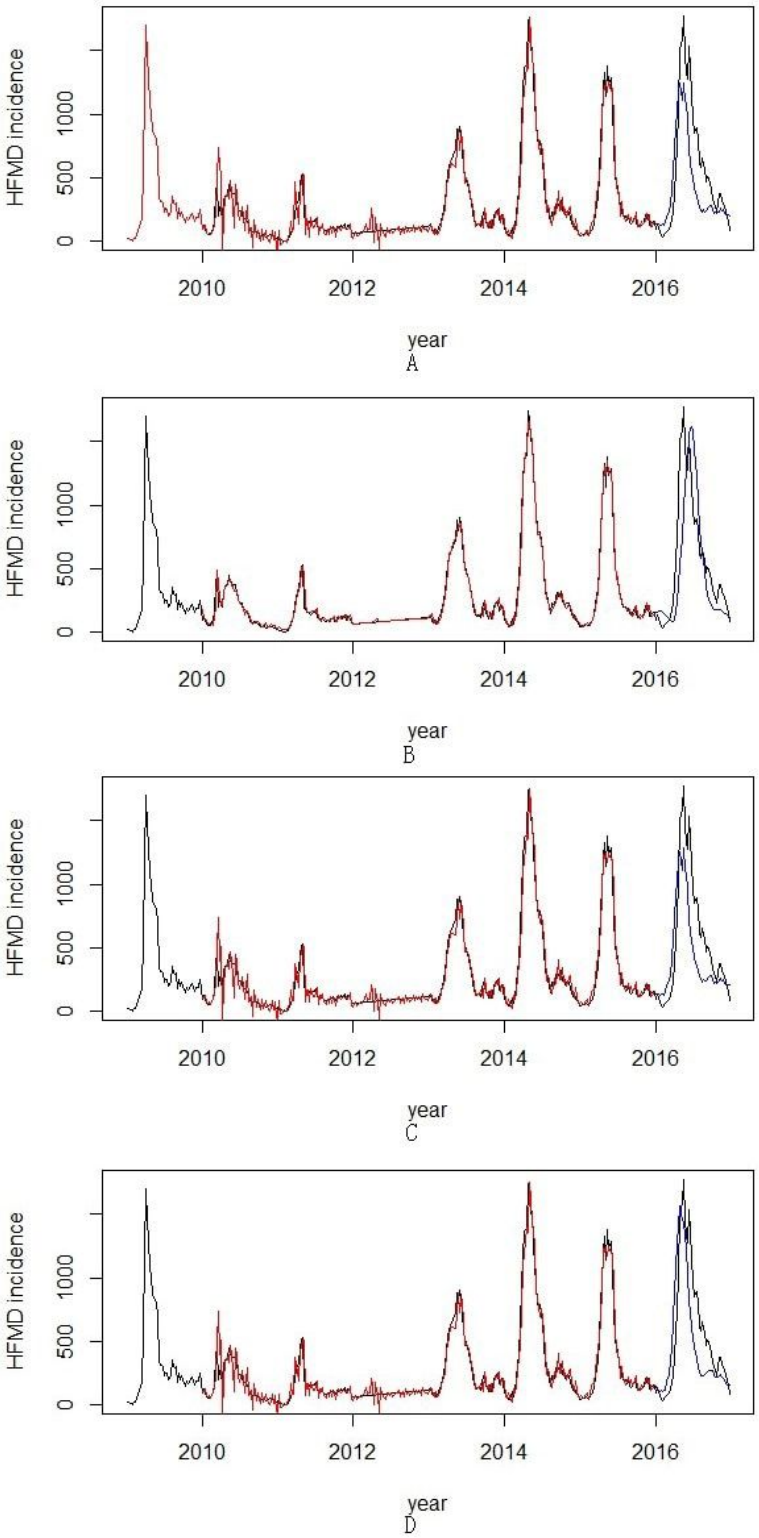

Figure 6

The fitted and 52 weeks forecasting time series plot of four models.(A: SARIMA model; B: NNAR model; C: SARIMA-NNAR hybrid model; D: wavelet-based SARIMA-NNAR model. Black line: original time series; red line: fitted time series in training set; blue line: forecasting time series in validation set)