

Identifying the Dynamic Pattern and Influencing Factors of Influenza in Northwest China from 2013 to 2020, Based on Dynamic Regression Model and Wavelet Analysis

Ruonan Wang

Ningxia Medical University

Jiancai Du

Ningxia Centers for disease prevention and control

Jiangping Li

Ningxia Medical University

Yajuan Zhang

Ningxia Medical University

Jing Wen

Ningxia Medical University

Yu Zhao (✉ zhaoyu@nxmu.edu.cn)

Ningxia Medical University <https://orcid.org/0000-0002-3759-2893>

Research article

Keywords: Influenza, Environmental factors, Public awareness, Dynamical regression model, Wavelet analysis

Posted Date: October 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-92809/v1>

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

[Read Full License](#)

Identifying the dynamic pattern and influencing factors of influenza in Northwest China from 2013 to 2020, based on dynamic regression model and wavelet analysis

Ruonan Wang^{1,*}, Jiancai Du^{2,*}, Jiangping Li¹, Yajuan Zhang¹, Jing Wen¹, Yu Zhao^{1,†}

1. School of Public Health and Management, Ningxia Medical University, Ningxia Yinchuan 750004, China

2. Ningxia Centers for Diseases Prevention Control, Ningxia Yinchuan 750004, China.

Abstract. Background: Influenza remains a serious global public health problem and a substantial economic burden. The dynamic pattern of influenza differs considerably among geographic and climatological areas, however, the factors underlying these differences are still uncertain. The aim of this paper is to characterize the dynamic pattern of influenza and its potential influencing factors in Northwest China. **Methods:** Influenza case in Ningxia China from Nov. 2013 to Jun. 2020 was served as influenza proxy. Firstly, the baseline seasonal ARIMA model of influenza cases and seasonal pattern were analyzed. Then, the dynamic regression model was used to identifying the potential influencing factors of influenza. In addition, the wavelet analysis was further used to explore the coherence between influenza cases and these significant influencing factors. **Results:** The high risk periods of influenza in Ningxia presented a winter cycle outbreaks pattern and the fastigium came in January. The seasonal ARIMA(0, 0, 1)(1, 1, 0)₁₂ was the optimal baseline forecast model. The dynamic regression models and wavelet analysis indicated that PM_{2.5} and public awareness are significantly positively associated with influenza, as well as minimum temperature is negatively associated. **Conclusion:** Meteorological (minimum temperature), pollution (PM_{2.5}) and social (public awareness) factors may significantly associated with influenza in Northwest China. Decreasing PM_{2.5} concentration or increasing the public awareness prior to the fastigium of influenza may be the serviceable methods to reduce the disease risk of influenza, which have an important implication for policy-makers to choose an optimal time for influenza prevention campaign.

Keywords: Influenza; Environmental factors; Public awareness; Dynamical regression model; Wavelet analysis;

*These authors contributed equally to this work.

†Corresponding author, Email: zhaoyu@nxmu.edu.cn.

1. Introduction

Influenza is one of the most serious diseases worldwide, resulting in an estimated 1 billion cases, 3~5 million hospitalizations, and 290,000~650,000 respiratory deaths globally per year [1]. Usually, the virus is spread through the air from coughs or sneezes [1]. The disease is caused by influenza viruses, can be divided into A, B, C three types. The influenza A viruses antigenic variation often happens, thus the infectious, spread rapidly and easily happened a pandemic [2]. We are now face with a situation where the virus has spread widely, when many people in all age groups in many countries have some immunity to the new virus, and where seasonal influenza A-type and influenza B-type viruses are being reported successively in many provinces [3]. The variation of influenza virus antigens [4] (mainly is hemagglutinin and neuraminidase) is constantly mutating, and then it makes that influenza pandemic has so far been unable to control effectively [5]. Based on this overall picture, evidence is strong that the recent influenza pandemic patterns are transitioning towards seasonal patterns of influenza [6]. WHO recommended that [3] all countries should maintain monitoring for influenza, through routine respiratory disease surveillance and reporting as well as monitoring and investigating unusual disease patterns suggestive of potential changes in the severity. Therefore, identifying the dynamics pattern of influenza and its potential influencing factors are interesting issues to help for guiding the decision making by disease prevention and control agencies and governments.

Time-series methods have served as an powerful tools to explore the dynamics of numerous epidemics [7], however, seasonality as a temporal pattern of systematic periodic oscillation within a predetermined cycle that can be characterized by peak timing, amplitude, and duration [8]. The dynamic pattern of influenza continues to attract many researchers' attentions, however there was likely to be no definitive explanation that covers all regions, climates and populations, at all times. The influenza in each area was most likely the result of a number of factors contributing to different degrees to the observed incidence and timing of influenza infections. Given this, there were some retrospective studies on the seasonality and trend analysis of influenza data to describe the trends of influenza incidence [9, 10]. At the same time, the mathematical models have become a powerful tools to investigate the time-series of infectious diseases recently, thus, various models (such as generalized regression [11], stochastic model [12], compartmental model [13, 14] and autoregressive moving average models [7, 15]) have been used to forecast influenza in order to analyze the trends and predict the root cause of the influenza incidence epidemic. In addition, the wavelet analysis recently has become an powerful tool to investigate the potential inter-connectedness between influenza and climate factors according to the method proposed by Grinsted, Moore and Jevrejeva [16, 17]. These methods provide the approaches to investigated the dynamic pattern of influenza, which is important to analyze and know the epidemic situation of influenza.

The dynamics pattern of influenza might be related to many factors, including influenza virus mutation, people susceptibility, public awareness and climate changes, etc. In many environmental factors, the environmental pollution and climate are two important factors that can affect the influenza. The dynamic pattern of influenza differ considerably among geographic and climatological

areas. For example, in the Northwest of China, coal fire-power industries and heating systems, as well as vehicle emissions, all conduce to air pollution (airborne fine particulate matter $PM_{2.5}$, PM_{10} , and SO_2 , etc.), which has affected the transmission of respiratory system infectious diseases [19, 20]. Numerous epidemiological studies have consistently demonstrated that exposure to ambient $PM_{2.5}$ is associated with increased respiratory health outcomes [21, 22]. Lowen and Steel [23] found that humidity and temperature play important roles in shaping the influenza seasonality. Experimental researches in guinea pigs indicated that the influenza is stable at low relative humidity (RH) and relatively unstable at intermediate RH. Shaman, Goldstein and Lipsitch [24] pointed out that lower temperature and absolute humidity (AH) increase influenza virus survival and transmission in temperate regions. Some studies respectively found that the influenza negative associated with temperature and AH in Beijing, North of China [25]. In addition, some studies illustrated that public awareness and preventive behaviors may significantly associated with infectious diseases [12, 18, 26, 27]. As well known, health promotion efforts may increase the level of preventive awareness during disease outbreak, but most of the researches were based on the methods of questionnaire, the effect of public awareness on the influenza varies as time was still uncertain. Thus, identifying the potential interaction relationship between influenza and environmental factors (environmental pollution, climate and public awareness) is meaningful to provide scientific references for influenza and consequently carry out prevention strategies.

Ningxia Hui Ethnic Autonomous Region is located in the northwest inland plateau of China, with a typical continental semi-humid semi-arid climate, where have approximately 650 million populations [28]. Existing surveillance result has shown that influenza was one kind of major category C infectious diseases (accounted for about 2.85% of infectious diseases) in Ningxia, Northwest China [29]. However, there are few related studies focus on the influenza in Northwest China. In order to precisely predict of dynamic pattern and effectively identify the potential influencing factors of influenza, in this study, we analyzed the dynamic pattern of influenza by using ARIMA or ARIMAX models, and explored the various environmental factors, including environmental pollution ($PM_{2.5}$, PM_{10} , SO_2 , NO_2 , CO and O_3), climate (minimum temperature and maximum temperature) and public awareness (Baidu index), as possible drivers of influenza in Ningxia, Northwest China, from 2013 to 2020. These analyses aim to provide evidence for the prevention and control strategies of influenza in future.

2. Materials and methods

2.1 Data collection

The data information was collected by three parts, the surveillance data for influenza, meteorological and public awareness data in Ningxia. The monthly influenza case in Ningxia Hui Autonomous Region from Nov. 2013 to Jun. 2020 (80 months) was obtained from the Notifiable Disease Surveillance System (NNDSS) (see Figure 1 (a)). The meteorological factors including $PM_{2.5}$ (particulate matter $< 2.5\mu m$ in diameter, unit: $\mu g/m^3$), PM_{10} (particulate matter $<$

$10\mu\text{m}$ in diameter, unit: $\mu\text{g}/\text{m}^3$), SO_2 (sulfur dioxide, unit: $\mu\text{g}/\text{m}^3$), CO (unit: mg/m^3), NO_2 (unit: $\mu\text{g}/\text{m}^3$), O_3 (unit: $\mu\text{g}/\text{m}^3$), minimum temperature (unit: $^\circ\text{C}$) and maximum temperature (unit: $^\circ\text{C}$) were gathered from the National meteorological information Center (as shown in Figure 1 (c)-(j)) [30]. In addition, we used the monthly Baidu Index data from China’s largest search engine as a proxy of public awareness [18,31,32]. We extracted the monthly Baidu Index data according to the Chinese keyword ”influenza” and position located in Ningxia, China (see Figure 1 (b) for more details). The total data set was divided into two sets: the data from Nov. 2013 to Dec. 2018 was viewed as training set and the data from Jun. 2019 to Jun. 2020 was viewed as the validation set.

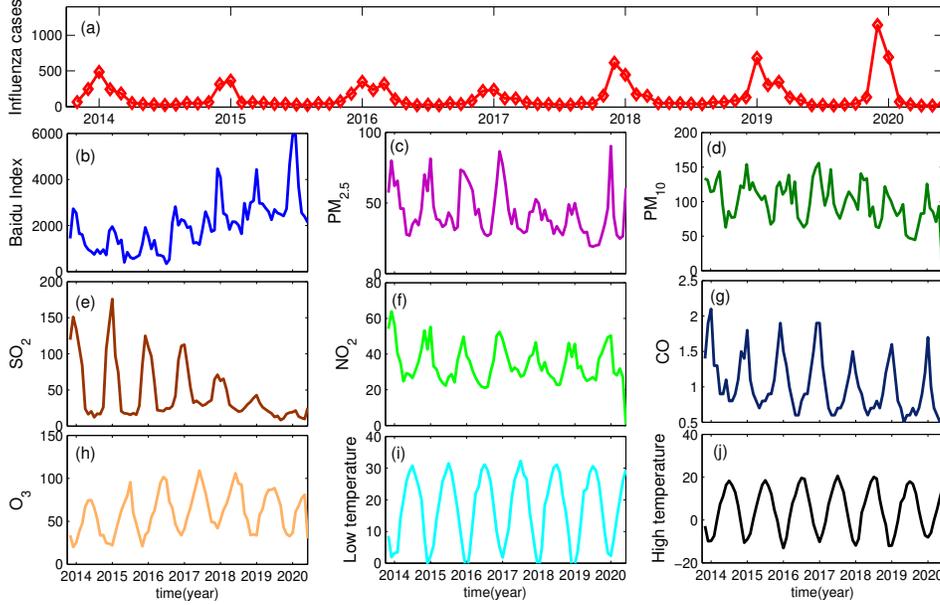


Figure 1: The time series of the monthly (a) influenza cases, (b) Baidu index, (c) $\text{PM}_{2.5}$ (d) PM_{10} , (e) SO_2 ,(f) NO_2 ,(g) CO , (h) O_3 , (i) minimum temperature and (j) maximum temperature in Ningxia, China from Nov. 2013 to Jun. 2020.

2.2 Seasonal ARIMA model

The time series analysis commonly identified as the time domain approach and the frequency domain approach. The time domain approach mainly reveals the evolution of time series from the perspective of sequence autocorrelation. The frequency domain approach mainly resort to Fourier analysis to explore the time series from the frequency viewpoint [33].

The autoregressive integrated moving average (ARIMA) model is one kind of time domain approach to explain all of the interesting dynamics of a time series. Since 1970, Box and Jenkins proposed the landmark Box-Jenkins method for parameter estimation and forecasting of ARIMA models [34]. Considering the ARIMA model contains the seasonal components, the product sea-

sonal ARIMA model (ARIMA(p,d,q)×(P,D,Q)_τ) was given as :

$$\phi(B)\Phi(B^\tau)(1-B)^d(1-B^\tau)^D X_t = \theta(B)\Theta(B^\tau)\epsilon_t, \quad (1)$$

where

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \\ \Phi(B^\tau) &= 1 - \Phi_1 B^\tau - \Phi_2 B^{2\tau} - \dots - \Phi_P B^{P\tau}, \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q, \\ \Theta(B^\tau) &= 1 - \Theta_1 B^\tau - \Theta_2 B^{2\tau} - \dots - \Theta_Q B^{Q\tau}, \end{aligned}$$

where B denotes the backshift operator, ϵ_t denotes the estimated residuals at time t with 0 mean and constant variance, τ is the seasonal cycle length. p, P, d, D, q, Q are non-negative parameters, which were needed to be determined when fitting the seasonal ARIMA model. p and P are the orders of autoregressive and seasonal autoregressive respectively, d, D are the orders of difference and seasonal difference respectively, q, Q are the orders of the moving average and seasonal moving average respectively [35].

Generally, the following methodological steps should be carried out to establish the ARIMA models.

1. Observed the character of the time series, which should be a non-white noise sequence, which usually examined by Ljung-Box (Q) test. Moreover, the data needs stationary to be a time series. The difference or seasonal difference is required to process the non-stationary original sequence, and the stationarity of the sequence can be tested by Augmented Dickey-Fuler (ADF) test.
2. Model (1) are usually identified predominantly with the autocorrelation function (ACF) as well as with the partial autocorrelation function (PACF), which can provide a possible range of the parameters p, P, q, Q . The maximum likelihood estimation method is carried out to estimate the parameters of the model (2). In addition, Ljung-Box test checked whether the residuals of Model (1) is a white noise, which can indicate Model (1) extracted the primary information of the time series.
3. According to the above step, there may exist several models meet the conditions, Akaike information criterion (AIC) and Bayesian information criterion (BIC) are used to identify the optimal model.

2.3 Dynamic regression model

To further identify the influencing factors of influenza in Northwest China, dynamic regression model (also known as transfer function model, abbreviated as ARIMAX) plays an important role in describing and analyzing the relationships among several time series [34]. The dynamic regression model not only can understand the dynamic relationships over time between the influencing factors and seasonal influenza, but also can improve accuracy of forecasts for influenza series by utilizing

the additional information available from the related environmental influencing factors series in the forecasts.

The dynamic regression model can be written as:

$$\begin{cases} y_t = \mu + \sum_{i=1}^k \frac{\Theta_i(B)B^{l_i}}{\Phi_i(B)}x_{it} + \epsilon_t, \\ \epsilon_t = \frac{\Theta(B)}{\Phi(B)}a_t, \end{cases} \quad (2)$$

where y_t denotes the response variable, x_t denotes the input variable, μ is the intercept of regression, $\Phi_i(B)$ is the polynomial of i -th autoregression coefficient of input variable, $\Theta_i(B)$ is the polynomial of i -th moving average coefficient of input variable, l_i is the delay order of i -th input variable, ϵ_t is the residual sequence of regression, $\Phi(B)$ is the autoregressive coefficient polynomial of residual sequence, $\Theta(B)$ is the moving average coefficient polynomial of residual sequence, and a_t is a white noise sequence with mean 0.

To establish the optimal dynamic regression model, we include the meteorological or socio-behaviors factors (Baidu Index as a proxy variable of public awareness) into the model to check whether this factor can improve accuracy of forecasts for the baseline seasonal ARIMA model. The following sequential procedure for fitting the dynamic regression model to data. Firstly, use the cross-correlation functions (CCF) between the influenza cases and meteorological or socio-behaviors factors to suggest a potential influencing factor and its corresponding order of delay. Then, to reduce the effect of autocorrelation of sequences on the processing results, the pre-whitening process is preformed to filter the input and the response variables, which can be used to establish an optimal ARIMA model for the meteorological or socio-behaviors factors. The CCF can calculate the cross-correlation coefficient of the filtered input and response variables, and the cross-correlation diagram can determine the lag relationship between the input and response variables. If the meteorological or socio-behaviors factors significantly ($P < 0.05$) correlated with the influenza cases, then the meteorological or socio-behaviors factors may include into the multivariate ARIMA model. The best selection criteria for the dynamic regression model are AIC and mean absolute percentage error (MAPE). The values of AIC or MAPE are smaller, the predictive performance more better. MAPE is calculated by the following formula:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|X_t - \hat{X}_t|}{X_t} \times 100\%, \quad (4)$$

where X_t is the real incidence at time t , \hat{X}_t is the estimated incidence at time t , and n is the number of predictions.

2.4 Wavelet analysis

In addition, an important approach of analyzing time series data in the frequency domain, as well as the time domain, is the investigation and exploitation of the properties of coherency. Coherency is a frequency measure based on the correlation between two series at a given frequency. Due to the epidemiological time-series are typically noisy, complex and strongly non-stationary [36], recently, wavelet analysis has been proposed to characterize the non-stationary time-series and also

to estimate dependencies among the signals. Following the general approach proposed by Cazelles et al. [36], wavelet analysis performs a time-frequency decomposition of the signal, which provide the estimation of its spectral characteristics as a function of time and frequency [37]. Meanwhile, the potential causality links between two time-series can be explored by the wavelet coherence (phase relationships), which revealed areas with high common power in time-frequency space [16]. Wavelet coherence (WTC) finds regions in time frequency space where the two time series co-vary.

WTC shows the significant coherence against red noise in time-frequency space , which can describe the significant covariance at specific periods (frequencies) and phase shift between two time series. The phase difference between the two series is indicated by arrows. Arrows pointing to the right mean that the variables are in phase. Arrows pointing to the left mean that the variables are out of phase. The down arrows show that climate factor is leading. The up arrows mean that influenza virus is leading. In phase indicate that variables will be having cyclical effect on each other and out of phase or anti-phase shows that variable will be having ant-cyclical effect on each other. To explore the period synchronization phenomenon, and characterize the potential association between the time-series of the potential influencing factors and influenza cases at the monthly interval from Nov. 2013 to Jun. 2020, we used the wavelet analysis approach.

All these time series models were established in R software (version 4.0.2). The packages "TSA", "forecast", "tseries", "fUnitRoots", "zoo" are used. In addition, Wavelet analysis was established by MATLAB software proposed by Grinsted, Moore and Jevrejeva [16] (<https://noc.ac.uk/business/marine-data-products/cross-wavelet-wavelet-coherence-toolbox-matlab>). A P -value of less than 0.05 was used for statistical significance.

3. Results

3.1 Seasonal ARIMA model

A total of 10,745 influenza cases were reported during this period. (see Figure. 1, a time series plot was used to evaluate the evolutionary trend over a period of years and the division of training set and testing set). Since the influenza case data presents a significant periodic oscillation, we shall first choose seasonal index to assess the risk period and seasonal pattern. Seasonal index can be used as the estimate of the seasonal component, which was a well known and widely used techniques for dealing with seasonal patterns of data [38, 39]. To assess the risks of each month in a year, we calculated the seasonal index of influenza for each month, as shown in Figure. 2, January has the high risk of influenza in each year, which is three times more higher than that of the other months. Thus, the infected risk of influenza in Ningxia usually presented a winter cycle outbreaks pattern, and the fastigium of influenza in December and January during one year.

Since the Ljung-Box (Q) test of influenza case time series obtain that $P = 3.196 \times 10^{-8} < 0.05$, indicating that the influenza case is not a white noise sequence. In addition, the ADF test of influenza case time series shows $P = 0.01 < 0.05$, which means that the sequence of influenza cases is a stationary sequence.

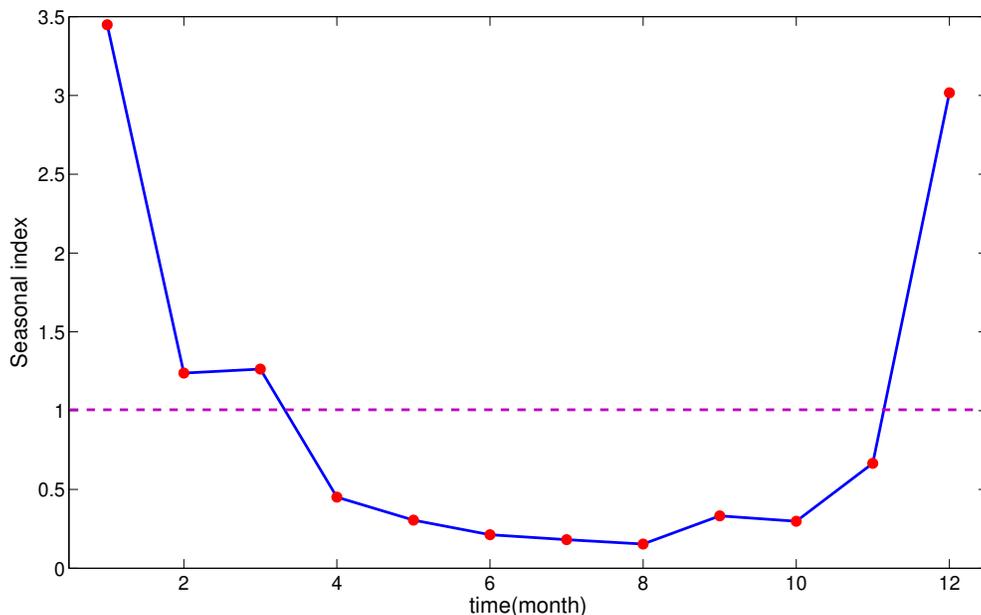


Figure 2: The seasonal index of influenza cases from Nov. 2013 to Jun. 2020.

Now, we can determine the ranges of seasonal ARIMA model parameters (p, q, P, Q) based on the autocorrelation function (ACF) and partial autocorrelation function (PACF). The ACF and PACF plot of were given as shown in Figure 3. ACF figure indicates that $q = 0$ or 1, and PACF figure shows that $p = 0, 1$ or 2. Since the cycle of influenza case is in years, the period is $S = 12$.

By utilizing of the AIC information criteria, the seasonal ARIMA(0, 0, 1)(1, 1, 0)₁₂ has the minimum AIC=659.22, AICc=659.68, BIC=665.3, thus, this model is the optimal. The parameter estimation and the white noise test of residual sequences of model (6) were shown in Table. 1, which indicates that the seasonal ARIMA(0, 0, 1)(1, 1, 0)₁₂ make better use of the information of data.

Then, the seasonal ARIMA(0, 0, 1)(1, 1, 0)₁₂ model of influenza has the following formula:

$$(1 - B^{12})X_t = \frac{(1-0.4343B)}{(1+0.6646B^{12})}\epsilon_t, \quad \epsilon_t \sim N(0, 6214). \quad (6)$$

Table 1. Parameter estimation results of ARIMA(0, 0, 1)(1, 1, 0)₁₂ model.

Parameter	Coefficient	Standard error	t-statistic	P-value
MA(1)	0.4343	0.114	3.805	$P < 0.001$
SAR(1)	-0.6646	0.1022	-6.503	$P < 0.001$

Now, the seasonal ARIMA(0, 0, 1)(1, 1, 0)₁₂ model is carried out to fit the influenza cases from Jul. 2019 to Jun. 2020. As shown in Figure 4, the fitted values and confidence intervals (95% and 50%) are given from Nov. 2013 to Dec. 2018, the predicted values and confidence intervals (95% and 50%) are employed from Jan. 2019 to Jun. 2020. The MAPE equals 40.5%. The fitting

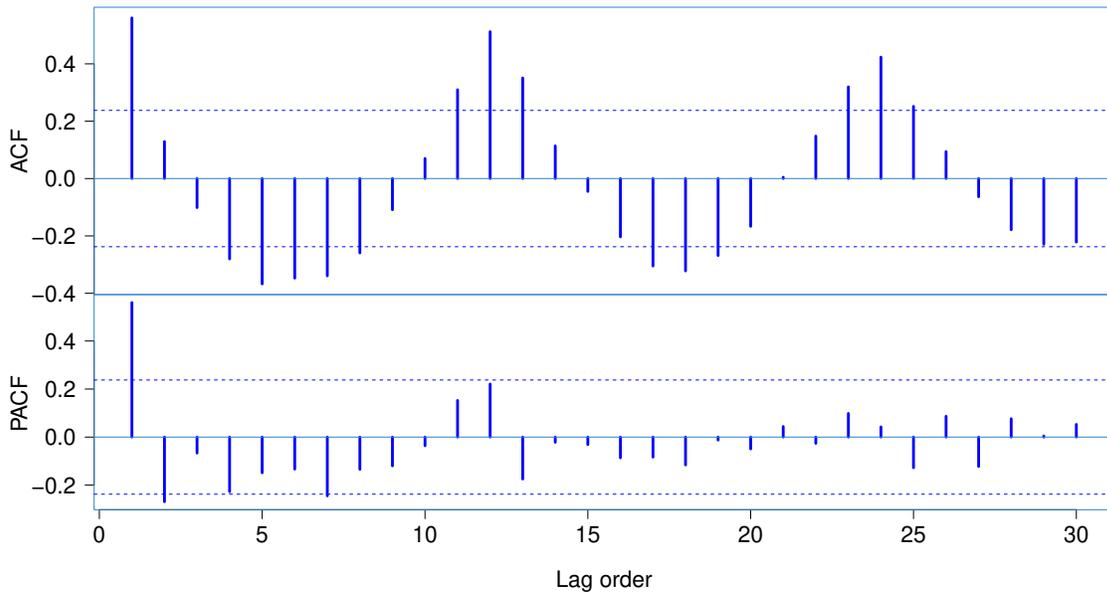


Figure 3: ACF and PACF of influenza cases from Nov. 2013 to Dec. 2018 of the original sequence.

and forecasting results were displayed in Figure 4, which showed that real data scattered in the confidence interval.

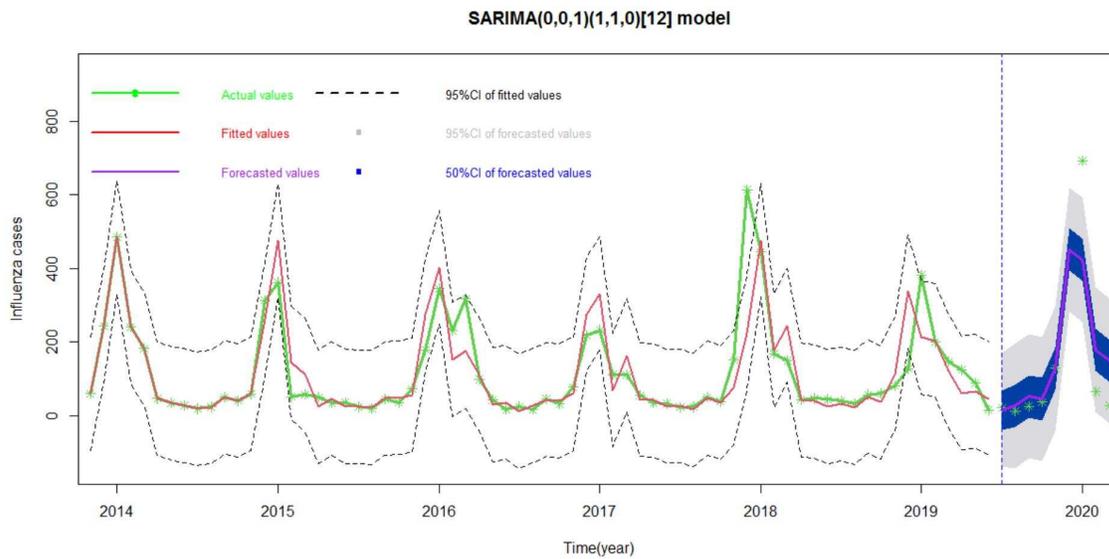


Figure 4: The fitting and predicting results of influenza cases by seasonal ARIMA(0,0,1)(1,1,0)₁₂ model. Here the green line is real data of influenza, red line is fitted values of model (6), purple line is the forecast values of model (6), black dotted line is the 95% confidence interval of fitted values, gray area and blue area are the 50% confidence belt of forecast values, respectively.

Table 2. The white noise test of residual sequence of model (6).

Model	Lag	χ^2	DF	P-value
	6	2.0935	5	0.8361
ARIMA(0, 0, 1)(1, 1, 0) ₁₂	12	12.538	11	0.3246
	18	14.952	17	0.5989

3.2 Dynamical regression model

To explore the optimal multivariate dynamic regression model, we considered the meteorological factors and socia-behavior factor into the baseline ARIMA(0, 0, 1)(1, 1, 0)₁₂ model to explore the factors can improve accuracy of forecasts. As shown in Table. 3, the optimal ARIMA models of the potential influencing factors were fitted, which determined the transformation of influenza series in pre-whitening. The cross-correlation function (CCF) between influenza cases and these environmental factors series were plotted in Figure 5 indicated the significant association between influenza cases and meteorological factors or socia-behavior factor, which can be used to suggest the potential lag orders.

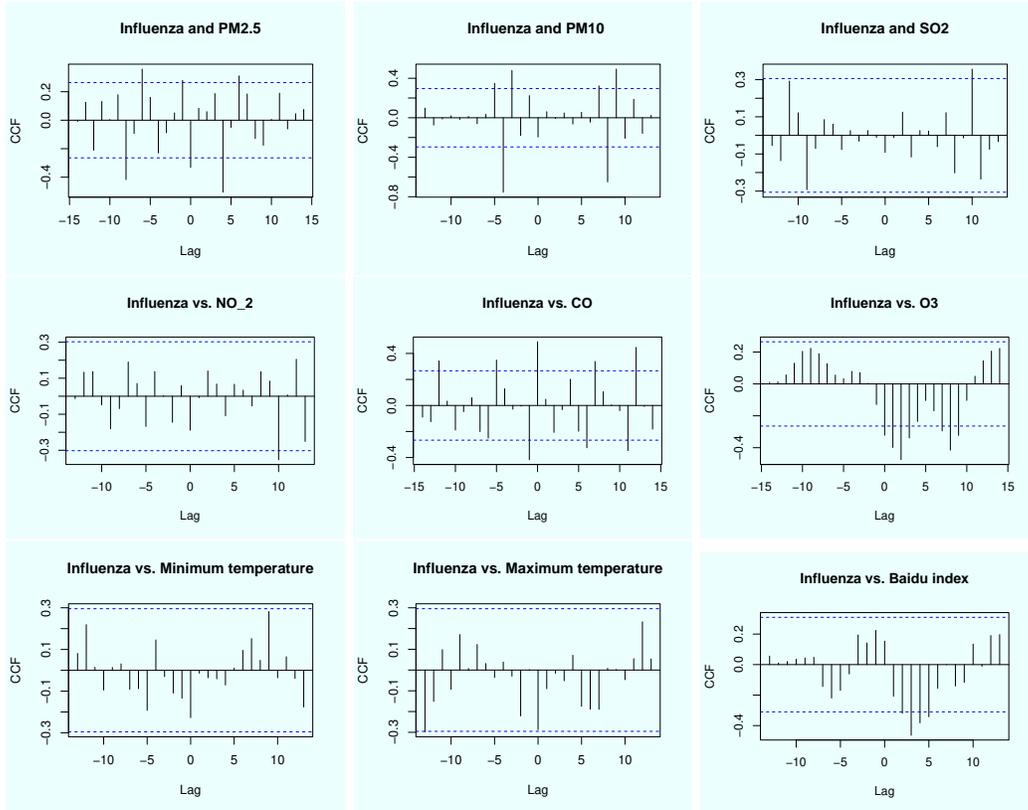


Figure 5: The Cross-correlations between the pre-whitened influenza cases and environmental factors (PM_{2.5}, PM₁₀, SO₂, NO₂, CO, O₃, Minimum temperature, Maximum temperature and Baidu index) from 2013 to 2019.

According to the AIC criterion, as listed in Table. 4, the optimal lag order predictors was determined in pre-whitening. There are three ARIMAX models have lower AIC and passed the parameters test, that is, ARIMA(0, 0, 1)(1, 1, 0)₁₂+PM_{2.5}, ARIMA(0, 0, 1)(1, 1, 0)₁₂+Minimum temperature, and ARIMA(0, 0, 1)(1, 1, 0)₁₂+Baidu index. These results demonstrated that PM_{2.5}, Minimum temperature and Baidu index may be the potential significant influencing factors of influenza in Northwest China.

3.3 Wavelet analysis

According to the ARIMAX models, we found that PM_{2.5}, minimum temperature, Baidu index might be close related with influenza cases, thus, further identifying the co-varies and correlation direction between the above environmental factors and influenza is meaningful for early warning and vaccine strategies of the influenza prevention.

Table 3. Comparison of ARIMAX models with different predictors.

Predictor	Optimization models	AIC	AICc	BIC
PM _{2.5}	ARIMA(1, 0, 2)(0, 1, 1) ₁₂	425.09	426.29	435.22
PM ₁₀	ARIMA(0, 0, 1)(1, 1, 1) ₁₂	497.69	498.48	505.79
SO ₂	ARIMA(3, 0, 0)(1, 1, 0) ₁₂	464.05	465.76	476.2
NO ₂	ARIMA(2, 0, 0)(1, 1, 0) ₁₂	349.87	351.07	360
CO	ARIMA(2, 0, 0)(0, 1, 1) ₁₂	-40.81	-39.61	-30.68
O ₃	ARIMA(0, 1, 1)(0, 1, 0) ₁₂	389.57	389.8	393.59
Minimum temperature	ARIMA(0, 0, 0)(1, 1, 0) ₁₂	244.24	244.47	248.3
Maximum temperature	ARIMA(0, 0, 0)(1, 1, 0) ₁₂	215.11	215.34	219.16
Baidu Index	ARIMA(3, 1, 0)(1, 1, 0) ₁₂	869.83	871.06	879.87

Table 4. Parameters estimation of ARIMAX models with different predictors.

Predictor	Lag	Coefficient	Standard error	P-value	AIC	MAPE(%)
PM _{2.5}	4	4.6948	2.8421	0.05	654.73	36.92
Minimum temperature	9	-13.0307	4.8619	0.0047	650.38	39.32
Baidu Index	3	0.2897	0.1939	0.07	645.36	34.9

As shown in Figure 6, wavelet transform and wavelet transform coherence (WTC) provided the information with respect to whether two time series are correlated or co-varies at a particular time and frequency. In Figure 6 (a), the wavelet power spectrum of influenza demonstrated that from Jul. 2017 to Sep. 2018 of period 12, it has the maxima of the undulations of the wavelet power spectrum, which implies that this area may be the most relevant region. Here we notice that the common features we found by eye from the individual wavelet transforms stand out as being significant at the 5% level. In Figure 6 (b), the results of WTC shows that the influenza is statistically significantly associated with PM_{2.5} at the period of approximately 40 months band (2015-2018) ($P < 0.05$), with arrows pointing to the right, which indicating that the influenza

and $PM_{2.5}$ are positively correlated during 2015 to 2018. At the same time, we also found a significantly negatively correlation between minimum temperature and influenza at the period of approximately 40 months band (2015-2018) ($P < 0.05$) and the arrows pointed to right (see Figure 6 (c)). For Baidu index, we also observed one significant high power spectrum in the 30-60 month band (2016-2018) with in-phase, which means that the series are positively correlated during the period (see Figure 6 (d) for details).

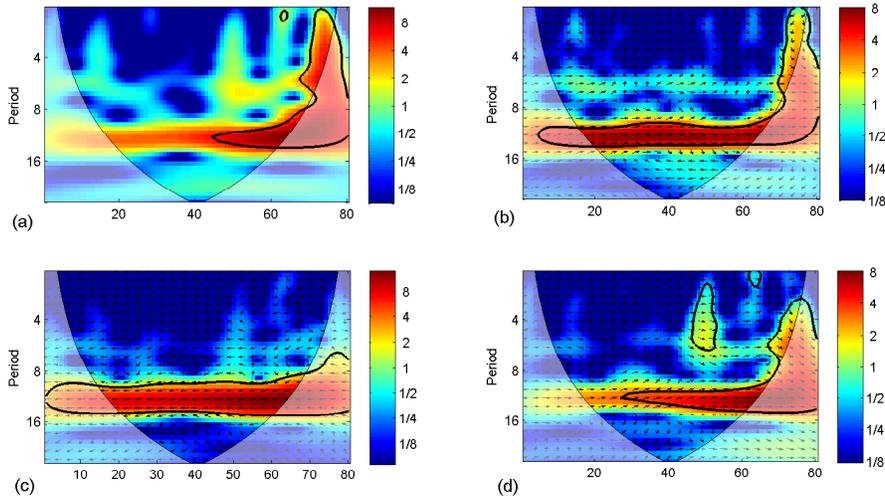


Figure 6: (a) Wavelet power spectrum analysis of monthly influenza prevalence from Nov. 2013 to Jun. 2020 in Ningxia China. (b) WTC between influenza and $PM_{2.5}$, (c) WTC between influenza and minimum temperature, (d) WTC between influenza and Baidu index. The 5% significance level against red noise is shown as a thick contour. The cone of influence, which indicates the region affected by edge effects, is shown with dark red shade. The color code for power ranges from blue (low power) to red (high power). X-axis is the time since Nov. 2013, and y-axis denotes frequency periods.

5. Conclusion

One important feature of influenza was the role of seasonal drivers that tend to limit the spread of the virus to particular periods of time of the year [9]. As a result of these seasonal drivers, influenza virus infections can generate recurrent epidemics year after year or even multiple waves of pandemic influenza [40]. Regardless of whether a place was in a pandemic situation or not, influenza viruses pose a risk of disease to many individuals and therefore, individuals should take prudent steps to reduce their risk of infection [6]. Therefore, identifying the seasonal pattern may contribute to the prediction of high risk period in the future and have implications for the outbreaks preparedness.

In this study, we reported monthly influenza cases data in Ningxia, Northwest China from Dec. 2013 and Jun. 2020, and investigated the dynamic patterns of influenza over the eight-

year period. The seasonal ARIMA model and seasonal index of influenza cases were analyzed, the results indicated that the high risk periods of influenza presented a winter cycle outbreaks pattern and the fastigium came in January. The dynamic regression model was used to identifying the potential influencing factors of influenza. Moreover, the wavelet analysis was further used to explore the coherence or co-varies between influenza cases and these significant influencing factors. These results found that $PM_{2.5}$ and public awareness are significantly positively associated with influenza, as well as minimum temperature is negatively associated with influenza. The seasonal dynamic pattern of influenza may be ascribed to the meteorological, pollution and social factors.

Many studies have illustrated some evidences for the association between particulate air pollution and human illness (especially cardiovascular and respiratory diseases) [28,41,42,43]. In China, the main research attention focused on the impact of $PM_{2.5}$ on the respiratory diseases for the developed provinces in southern China. Lei et al. [11] pointed out that children and the elderly people are more likely to be affected by air pollution due to their relatively weak immune systems, or/and less exposure to air pollutants. Many experiments have show that the exposure to air pollutant may result in lung function decline [44], and particulate matter of air pollutants may cause neurogenic inflammation of sensory nerve endings in the trachea [45]. The associations between the $PM_{2.5}$ concentration and influenza cases have been explored widely. However, the different regions meteorological/environmental conditions were diverse and the influenza dominant strain subtype in each year varied markedly [4]. The analysis of Ningxia can be viewed as a representation of an arid and semi-arid climate characteristics in Northwest China, since it was difficult to degrade and diffuse the environmental pollutant. Thus, decreasing $PM_{2.5}$ concentration prior to the fastigium of influenza may be the serviceable methods to reduce the disease risk of influenza.

Up to now, there were some results with respect to the association between temperature and influenza, especially in warm regions [46]. The temperature difference is an important feature of climate that can have important impacts on the influenza virus transmission [47] and may cause the seasonal influenza outbreaks [48]. Su et al. [17] observed a negative association between average temperature and three influenza virus subtype by using of Spearman's correlation and WTC. Chong et al. [49] verified that the temperature may be connected with the influenza virus epidemic, not only influenza virus A, but also influenza virus B. In this study, we also observed that minimum temperature is negatively associated with influenza cases, which was consistent with the finding of related studies of Lowen and Steel [17]. The mechanism may be that the influenza transmission was most efficient at low temperature (e.g. 5 °C), the virus transmission and survival capacity may decrease as temperature increase from 5 °C ro 20 °C and completely blocked at 30 °C in guinea pigs experiments [23]. Some studies also found that influenza virus, which is more stable at low temperature, may decrease the activities of proteases, reduce the mucus and ciliary movement and inhibit defense and immunity toward infection [50,51,52]. Thus, the minimum temperature may shape the seasonal dynamic pattern of influenza by affecting the host susceptibility and increasing the transmission of influenza virus in winter.

Public awareness may be conducive to the prevention of influenza infection have been reported by several groups [26, 27]. Specifically, by using a structured questionnaire, Khowaja et al. [26] found that three types of preventive awareness during the influenza epidemic were emerged, and were related to preventive behaviors against influenza infection. Balkhy et al. [27] identified awareness, attitudes, and practices related to influenza A (H1N1) among the Saudi public. Public awareness play an important role in helping curb epidemic, and health messages delivered through various media, such as media, internet, etc. Zhao et al. [18] pointed out that google index was a effective proxy indicator of public awareness, which had an important influence on infectious disease transmission. In this study, we used the Baidu index as a proxy of public awareness, and found that public awareness are significantly positively associated with influenza case in Northwest China. One of the most important experiences was raising public awareness, which effectively prevents the spread of influenza, and enables residents to proactively prevent the disease. These results provided the scientific support for further measure to prevent and control influenza virus, including vaccination time, the time of taking health education to the high-risk populations.

However, this study also has several limitations: 1) the time-series model has some inaccurate forecasts for some periods may due to the changes of dominant strains each years, however, our time-series data was not include subtype-specific of influenza virus (e.g., H1N1, H3N2, Victoria, Yamagata subtypes). 2) The influenza surveillance data used in our study were limited to sentinel hospitals, limited to one single province and also limited to a relatively short period of time. Thus, our data may underestimate the influenza cases, but we thought that the trends of influenza infected number in consistent seasonal pattern was still hold. 3) In addition, the influenza-like illness should be include in the future, which can provide a more stronger evidence for different region characters. These issues leave to the further consideration.

Despite all these limitations, our works concerned with prediction of the seasonal dynamic pattern of influenza and association between influenza cases and potential meteorological (minimum temperature), pollution (PM_{2.5}) and social (public awareness) factors. These analysis will help policy makers and health institution in order to explore potential benefits of vaccination strategy (vaccine recommendations of flu high-risk season) with a view to control and/or eradication of transmission of influenza appropriate for specific periods of time.

Ethics approval and consent to participate No involved.

Consent for publication All authors consent for publication.

Funding Research is supported by the Natural Science Foundation of Ningxia (No: 2020AAC03186) and the National Natural Science Foundation of China (12061058).

Conflict of interest The authors declare that they have no conflict of interest.

Authors' contributions Contributions of authors: R.N. Wang and J.C. Du contributed to study design, analyzed the data and drafted the manuscript. J. Weng collected and managed the data. Y. J. Zhang contributed the drawing figures. J. P. Li, Y. Zhao provided expert review of the manuscript, and reviewed and approved the manuscript.

Acknowledgements The authors would like to thank Prof. Kai Wang for his valuable suggestions in preparing the manuscript.

Availability of data The data used to support the findings of this study is available from the online website or corresponding author upon request.

References

- [1] Epidemic and pandemic' flu. *Lancet*. 2000, 355(9203): 509.
- [2] H. Nair, B. W. Abdullah. M. Katz, et al. Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis. *Lancet*, 2011, 378: 1917-1920.
- [3] Centres for Disease Control and Prevention. Flu. <https://www.cdc.gov/flu/>.
- [4] J. Park, Y. Ryu. Transmissibility and severity of influenza virus by subtype. *Infection, Genetics and Evolution*. 2018, 65: 288-292.
- [5] A. J. Caton, F. L. Raymond, G. G. Brownlee, J. W. Yewdell, W. Gerhard. Antigenic variation in influenza virus. *Biochemical Society Transactions*. 1983, 11(4): 435-441.
- [6] World Health Organization. What is post-pandemic? http://www.who.int/csr/disease/swineflu/frequently_asked_questions/post_pandemic/en/.
- [7] A. Azeez, D. Obaromi, A. Odeyemi, J. Ndege, R. Muntabayi. Seasonality and trend forecasting of tuberculosis prevalence data in Eastern cape, South Africa, using a hybrid model. *Int. J. Environ. Res. Public Health*. 2016,13, 757.
- [8] M. Moorthy, D. Castronovo, A. Abraham, S. Bhattacharyya, S. Gradus, J. Gorski, Y. N. Naumov, N. H. Fefferman, E. N. Naumova. Deviations in influenza seasonality: odd coincidence or obscure consequence? *Clinical Microbiology and Infection*. 2012, 18(10): 955-962.
- [9] A. Mummert, H. Weiss, L. P. Long, M. Jos, J. M. Amig, X. F. Wan. A perspective on multiple waves of influenza pandemics. *Plos One*. 2013. <http://dx.doi.org/10.1371/journal.pone.0060343>.
- [10] Y. J. Liang, L. Q. Fang, H. Pan, K. Z. Zhang, H. D. Kan, J. R. Brook, Q. H. Sun. PM_{2.5} in Beijing-temporal pattern and its association with influenza. *Environmantal Health*. 2014, 13: 102.

- [11] H. Lei, Z. Lian, C. Jin, et al. Acute effects of air pollution on influenza-like illness in Nanjing, China: A population-based study. *Chemosphere*. 2016, 147:180-187.
- [12] Y. Zhao, J. P. Li, X. Ma. Stochastic periodic solution of a susceptible-infective epidemic model in a polluted environment under environmental fluctuation. *Computational and Mathematical Methods in Medicine*. 2018, <https://doi.org/10.1155/2018/7360685>.
- [13] Y. Zhao, M. T. Li, S. L. Yuan. Analysis of transmission and control of tuberculosis in mainland China, 2005-2016, based on the age-structure mathematical model. *Int. J. Environ. Res. Public Health*. 2017, 14(10), 1192.
- [14] Y. Zhao, L. P. Zhang, S. L. Yuan. The effect of media coverage on threshold dynamics for a stochastic SIS epidemic model. *Physica A*. 2018, 512(15): 248-260.
- [15] W. R. Yan, Y. Xu, X. B. Yang, Y. K. Zhou. A hybrid model for short-term bacillary dysentery prediction in Yichang city, China. *Jpn. J. Infect. Dis.*, 2010, 63: 267-270.
- [16] A. Grinsted, J. C. Moore, S. Jevrejeva. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*. 2004, 11: 561-566.
- [17] W. Su , T. Liu, X. Geng, et al. Seasonal pattern of influenza and the association with meteorological factors based on wavelet analysis in Jinan City, Eastern China, 2013-2016. *PeerJ*. 2020, 8: e8626.
- [18] S. Zhao, S. S. H. Musa, J. Qin, D. H. He. Associations between public awareness, local precipitation, and cholera in Yemen in 2017. *The American Journal of Tropical Medicine and Hygiene*, 2019, 101(3): 521-524.
- [19] D. W. Dockery, Health Effects of Particulate Air Pollution. *Annals of Epidemiology*. 2009, 19(4): 257-263.
- [20] B. Brunekreef, S. T. Holgate. Air pollution and health. *Lancet*, 2002, 360(9341): 1233-1242.
- [21] H. L. Lin, W. J. Ma, H. Qiu, M. G. Vaughn, E. J. Nelson, Z. M. Qian, L. W. Tian. Is standard deviation of daily PM_{2.5} concentration associated with respiratory mortality? *Environmental Pollution*. 2016, 216: 208-214.
- [22] K. A. Johannson, J. R. Balmes, H. R. Collard. Air pollution exposure: a novel environmental risk factor for interstitial lung disease? *Chest*. 2015, 147(4): 1161-1167.
- [23] A. C. Lowen, J. Steel. Roles of humidity and temperature in shaping influenza seasonality. *Journal of Virology*. 2014, 88(14): 7692-7695.
- [24] J. Shaman, E. Goldstein, M. Lipsitch. Absolute humidity and pandemic versus epidemic influenza. *American Journal of Epidemiology*. 2011, 173(2): 127-135.

- [25] J. Y. Sun, X. Y. Sun, Z. Xiao. Correlation between meteorological parameters and influenza activity in Xicheng District of Beijing. *Occupation and Health*. 2018, 34(20): 2811-2814.
- [26] Z. A. Khowaja, M. I. Soomro, A. K. Pirzada, et al. Awareness of the Pandemic H1N1 Influenza global outbreak 2009 among medical students in Karachi, Pakistan. *Journal of Infection in Developing Countries*. 2011, 5(3): 151-155.
- [27] H. H. Balkhy, M. A. Abolfotouh, R. H. AlHathloul, et al. Awareness, attitudes, and practices related to the swine influenza pandemic among the Saudi public. *BMC Infectious Diseases*. 2010, 10: 42.
- [28] H. L. Zhang, Q. A. Wang, Y. J. Zhang, Y. Yang, Y. Zhao, J. R. Sang, Y. L. Zhang, Y. E. Zhang, F. Xie, S. S. Li, Y. H. Zhang, Y. M. Guo. Modeling the impacts of ambient temperatures on cardiovascular mortality in Yinchuan: evidence from a northwestern city of China. *Environmental Science and Pollution Research*. 2018, 25(6): 6036-6043.
- [29] Ningxia Centers for Diseases Prevention Control. <http://www.nxcdc.org/>.
- [30] National meteorological information center. Available from: <http://data.cma.cn/>.
- [31] Baidu Index. Available from: <http://index.baidu.com/v2/index.html>.
- [32] Y. Zhao, R. N. Wang, J. P. Li, Y. H. Zhang, H. F. Yang. Analysis of the transmissibility change of 2019-Novel coronavirus pneumonia and its potential factors in China from 2019 to 2020. *BioMed Research International*, 2020, 3842470.
- [33] H. S. Robert, S. S. David. *Time series analysis and its applications with R examples*. 2017, Springer International Publishing AG, USA.
- [34] G. E. P. Box, G. M. Jenkins. *Time series analysis, forecasting, and control*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2015.
- [35] R. A. Yaffee. *Introduction to time series analysis and forecasting with applications of SAS and SPSS*. Academic Press. Inc. London, 2000.
- [36] B. Cazelles, M. Chavez, G. C. de Magny, J. Guégan, S. Hales. Time-dependent spectral analysis of epidemiological time-series with wavelets. *J. R. Soc. Interface*. 2007, 4: 625-636.
- [37] C. Torrence, G. P. Compo. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* 1998, 79: 61-78.
- [38] T. I. Peter. A Seasonal Index for Business. *Decision Sciences*. 1997, 28(2): 335-355.
- [39] D. J. Cowden. Moving seasonal indexes. *Journal of the American Statistical Association*, 1942, 37: 523-524.
- [40] R. Omori, A. Sasaki. Timing of the emergence of new successful viral strains in seasonal influenza. *J. Theor. Biol.* 2013, 329: 32-38.

- [41] K. J. Hu, Y. M. Guo, D. Y. Hu, R. G. Du, X. C. Yang, J. M. Zhong, F. R. Fei, F. Chen, G. B. Chen, Q. Zhao, J. Yang, Y. Q. Zhang, Q. Chen, T. T. Ye, S. S. Li, J. G. Qi. Mortality burden attributable to PM in Zhejiang Province, China. *Environmental International*. 2018, 121: 515-522.
- [42] F. Dominici, R. D. Peng, M. L. Bell, L. Pham, A. McDermott, S. L. Zeger, J. M. Samet. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA*. 2006, 295(10): 1127-1134.
- [43] C. Feng, L. Jian, W. W. Sun, Y. Zhang, Q. Wang. Impact of ambient fine particulate matter (PM_{2.5}) exposure on the risk of influenza-like-illness: a time-series analysis in Beijing, China. *Environmental Health*. 2016, 15(1): 1-12.
- [44] E. Samoli, P. T. Nastos, A. G. Paliatsos, K. Katsouyanni, K. N. Priftis, E. Samoli, A. G. Paliatsos, K. Katsouyanni. Acute effects of air pollution on pediatric asthma exacerbation: evidence of association and effect modification. *Environ. Res.* 2011, 111: 418-424.
- [45] B. Veronesi, M. Oortgiesen. Neurogenic Inflammation and particulate matter (PM) air pollutants. *Neurotoxicology*. 2001, 22: 795-810.
- [46] J. Tamerius, C. Uejio, J. Koss, J. Shaman. Seasonal characteristics of influenza vary regionally across US. *PLOS ONE*. 2019, 14(3): e0212511.
- [47] A. C. Lowen, S. Mubareka, J. Steel, P. Palese. Influenza virus transmission is dependent on relative humidity and temperature. *Plos Pathogens*. 2007, 3(10): 1470-1476.
- [48] Q. T. Pham, C. Marc, N. D. Tran, T. Vu Dinh, T. Y. Nguyen, T. H. Nguyen, J. W. Daniel, F. B. Maciej, H. Peter. Seasonality of absolute humidity explains seasonality of influenza-like illness in Vietnam. *Epidemics*. 2015, 13: 65-73.
- [49] K. C. Chong, T. C. Lee, S. Bialasiewicz, et al. Association between meteorological variations and activities of influenza A and B across different climate zones: a multi-region modelling analysis across the globe. *Journal of Infection*. 2019, 80(1): 84-98.
- [50] C. M. Wong, L. Ynag, T. Q. Thach, P. Y. Chau, K. P. Chan, G. N. Thomas, T. H. Lam, T. W. Wong, A. J. Hedley, J. S. Peiris. Modification by influenza on health effects of air pollution in Hongkong. *Environ. Health Perspect.* 2009, 117(2): 248-253.
- [51] X. T. Yang, D. Liu, J. He, et al. Classification and regression tree model study on correlation between influenza epidemic and meteorological factors indifferent areas of Gansu, 2010 to 2016. *Disease Surveillance*. 2019, 34(5): 440-445.
- [52] J. Shaman, M. Kohn, Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America*. 2009, 106(9): 3243-3248.

Figures

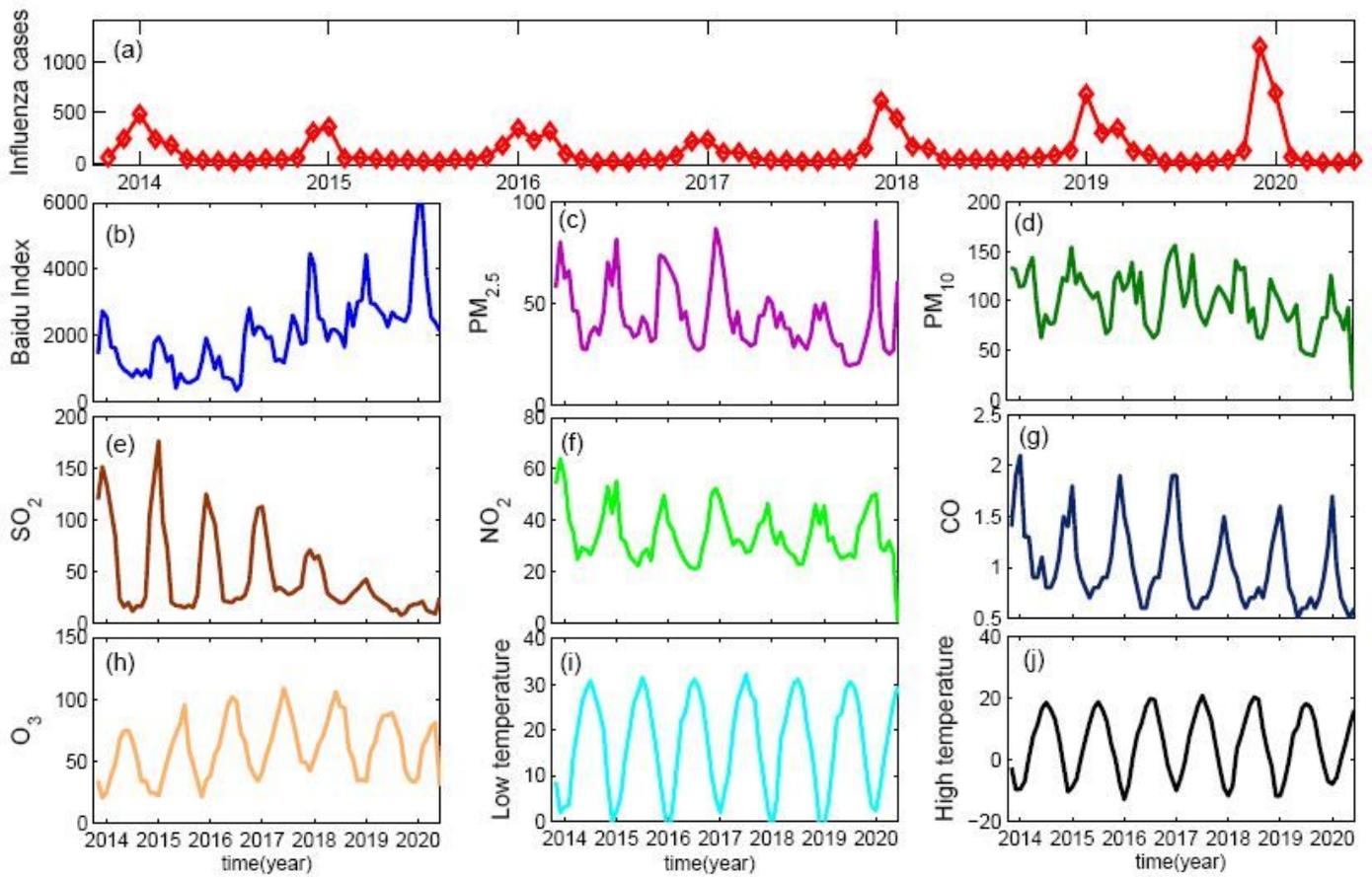


Figure 1

The time series of the monthly (a) influenza cases, (b) Baidu index, (c) PM_{2.5} (d) PM₁₀, (e) SO₂, (f) NO₂, (g) CO, (h) O₃, (i) minimum temperature and (j) maximum temperature in Ningxia, China from Nov. 2013 to Jun. 2020.

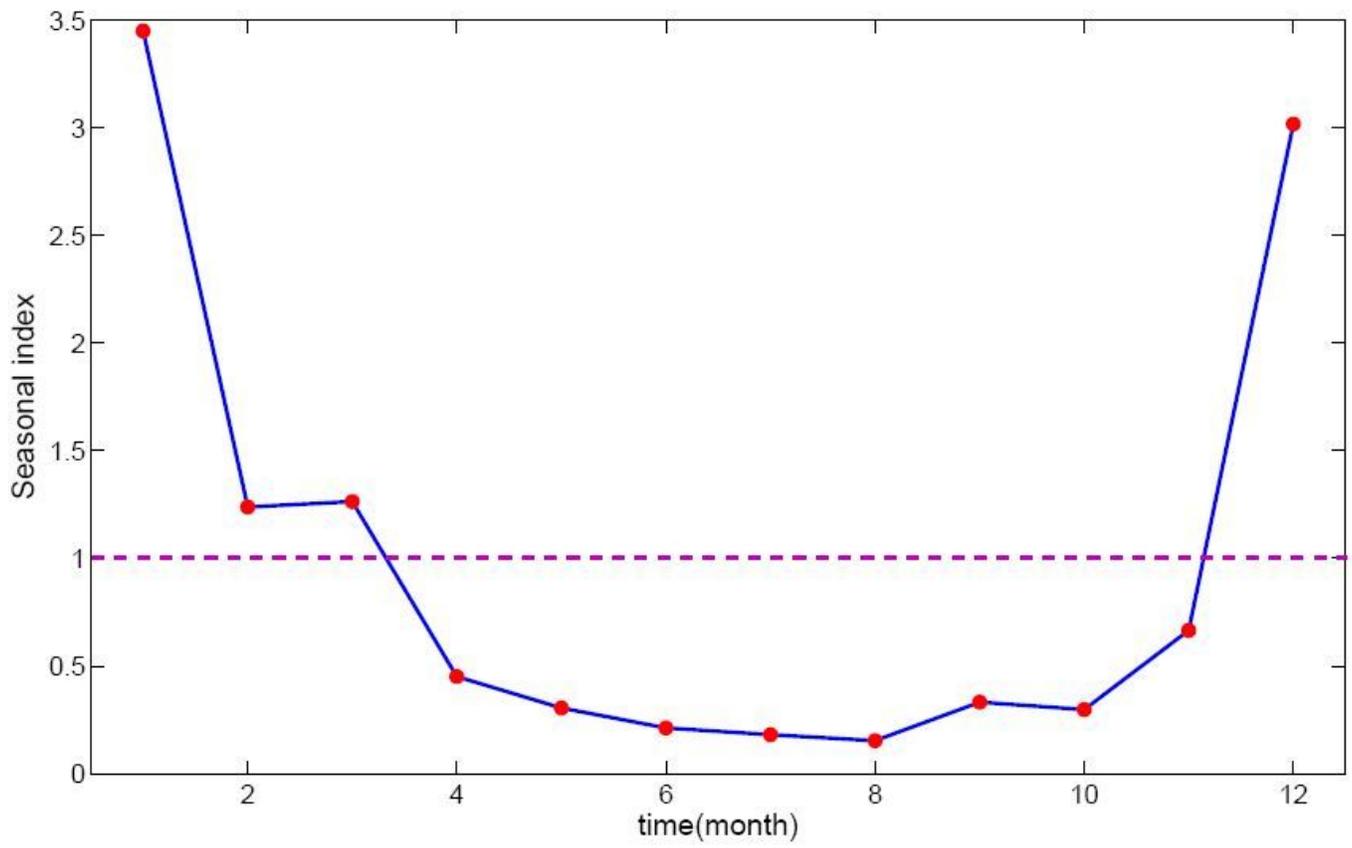


Figure 2

The seasonal index of influenza cases from Nov. 2013 to Jun. 2020.

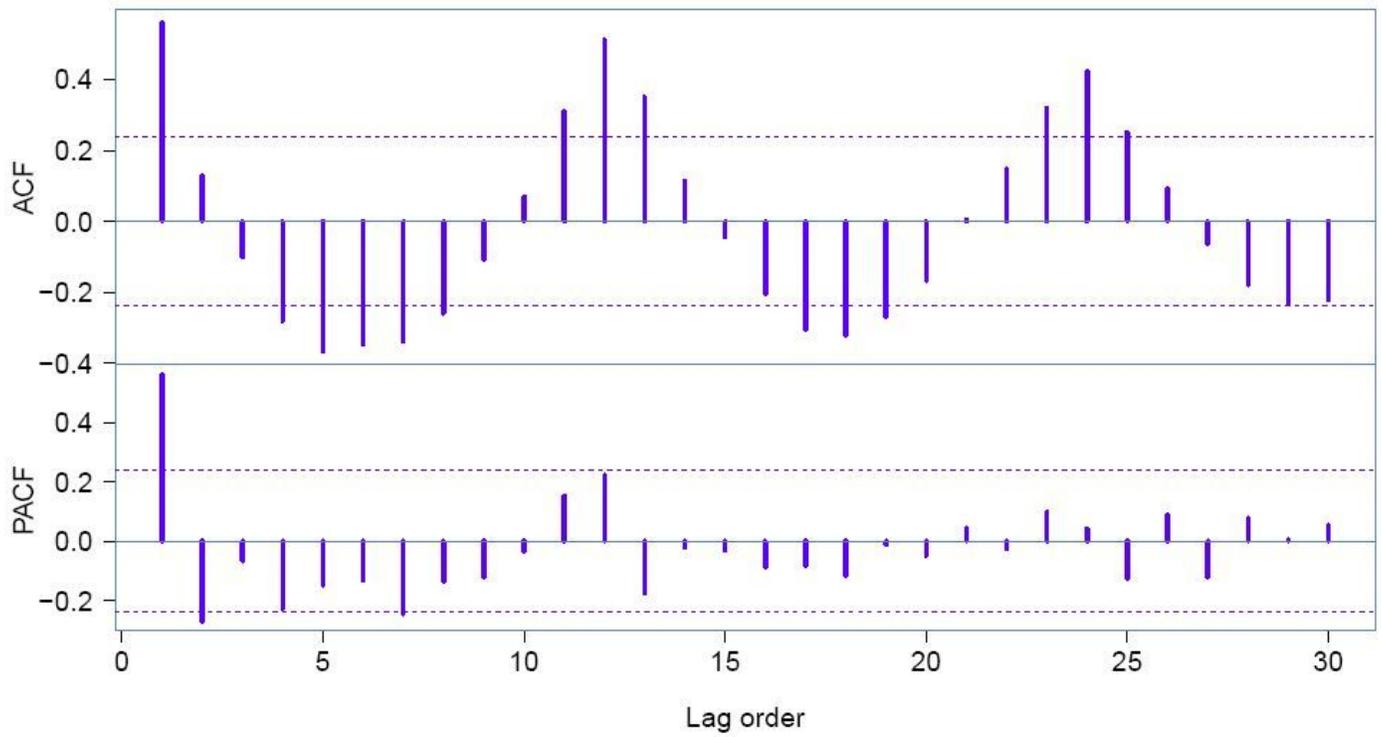


Figure 3

ACF and PACF of influenza cases from Nov. 2013 to Dec. 2018 of the original sequence.

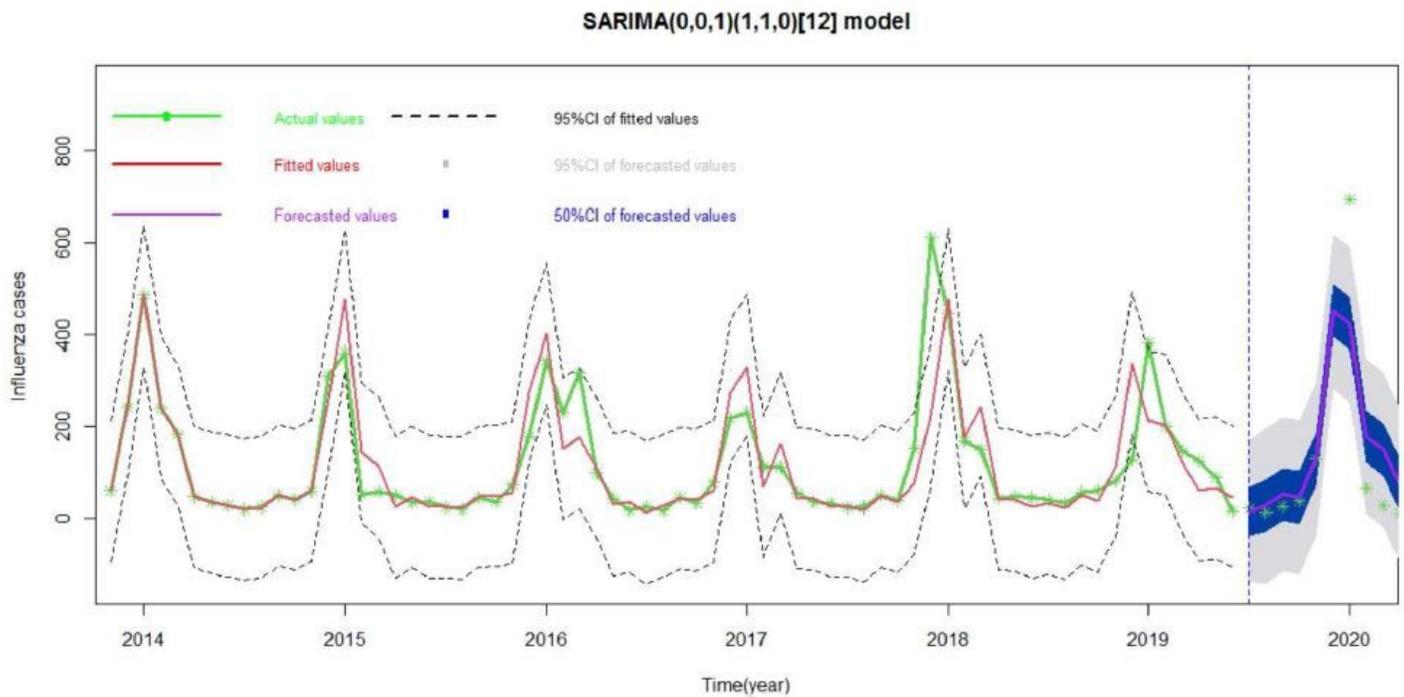


Figure 4

The fitting and predicting results of influenza cases by seasonal ARIMA(0, 0, 1)(1, 1, 0)₁₂ model. Here the green line is real data of influenza, red line is fitted values of model (6), purple line is the forecast values of model (6), black dotted line is the 95% confidence interval of fitted values, gray area and blue area are the 50% confidence belt of forecast values, respectively.

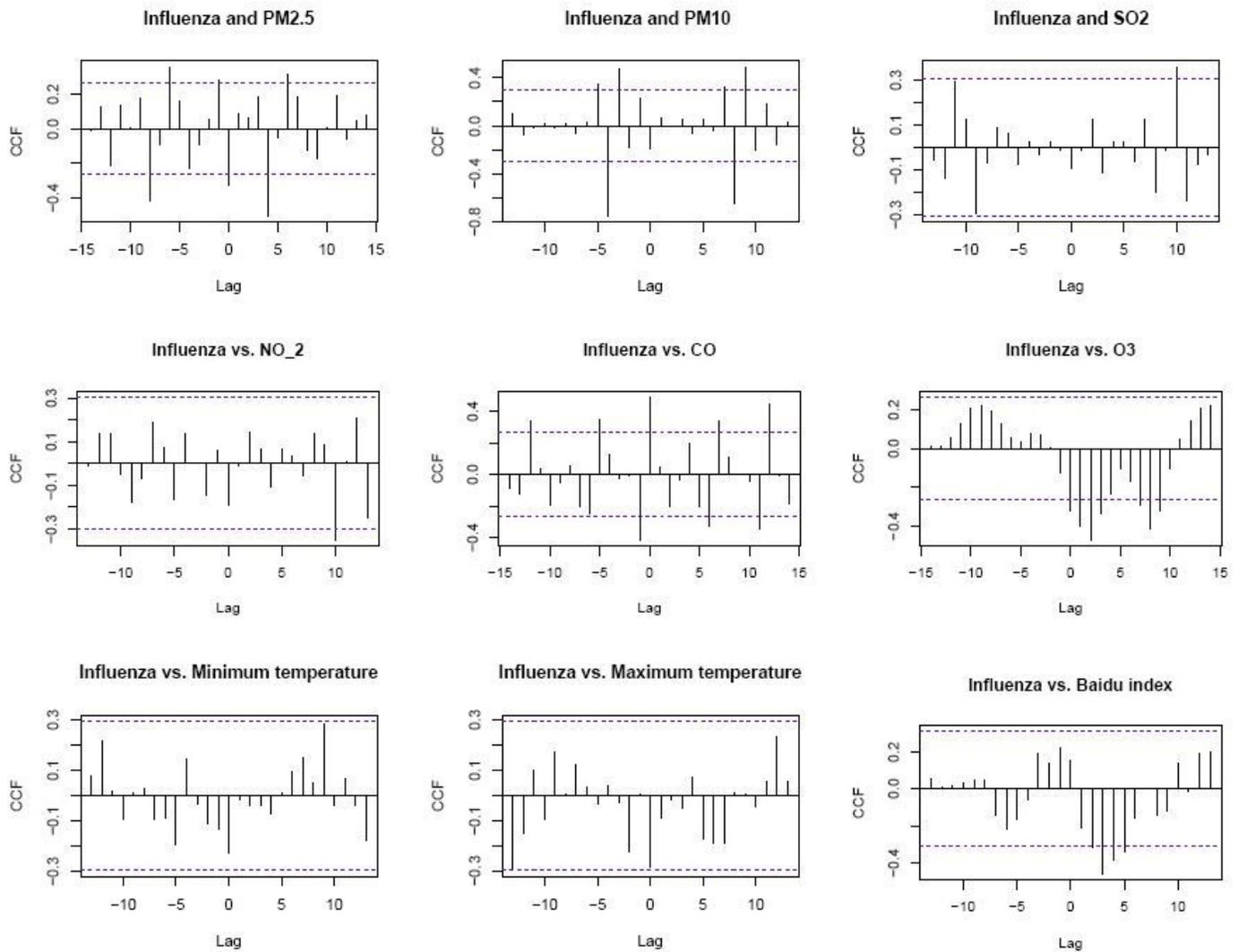


Figure 5

The Cross-correlations between the pre-whitened influenza cases and environmental factors (PM2.5, PM10, SO2, NO2, CO, O3, Minimum temperature, Maximum temperature and Baidu index) from 2013 to 2019.

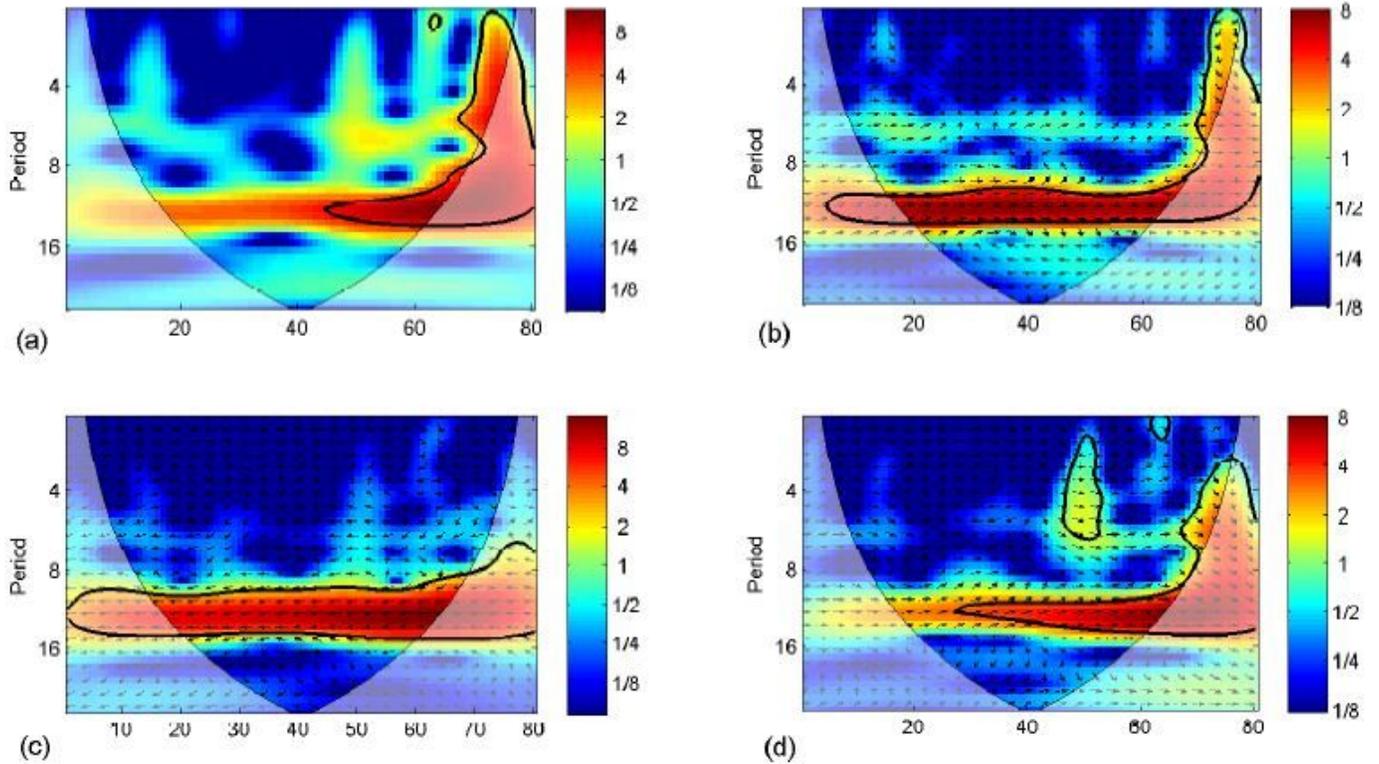


Figure 6

(a) Wavelet power spectrum analysis of monthly influenza prevalence from Nov. 2013 to Jun. 2020 in Ningxia China. (b) WTC between influenza and PM2.5, (c) WTC between influenza and minimum temperature, (d) WTC between influenza and Baidu index. The 5% significance level against red noise is shown as a thick contour. The cone of influence, which indicates the region affected by edge effects, is shown with dark red shade. The color code for power ranges from blue (low power) to red (high power). X-axis is the time since Nov. 2013, and y-axis denotes frequency periods.