

Regional distribution of non-human H7N9 avian influenza virus detections in China and construction of a predictive model

Zeying Huang¹✉, Haijun Li², Beixun Huang¹

¹Institute of Food and Nutrition Development,
 Ministry of Agriculture and Rural Affairs, 100081, Beijing, China

²University of Sanya, 572000, Sanya, Hainan province, China
 huangzeying@caas.cn

Received: January 4, 2021 Accepted: June 10, 2021

Abstract

Introduction: H7N9 avian influenza has broken out in Chinese poultry 10 times since 2013 and impacted the industry severely. Although the epidemic is currently under control, there is still a latent threat. **Material and Methods:** Epidemiological surveillance data for non-human H7N9 avian influenza from April 2013 to April 2020 were used to analyse the regional distribution and spatial correlations of positivity rates in different months and years and before and after comprehensive immunisation. In addition, positivity rate monitoring data were disaggregated into a low-frequency and a high-frequency trend sequence by wavelet packet decomposition (WPD). The particle swarm optimisation algorithm was adopted to optimise the least squares support-vector machine (LS-SVM) model parameters to predict the low-frequency trend sequence, and the autoregressive integrated moving average (ARIMA) model was used to predict the high-frequency one. Ultimately, an LS-SVM-ARIMA combined model based on WPD was constructed. **Results:** The virus positivity rate was the highest in late spring and early summer, and overall it fell significantly after comprehensive immunisation. Except for the year 2015 and the single month of December from 2013 to 2020, there was no significant spatiotemporal clustering in cumulative non-human H7N9 avian influenza virus detections. Compared with the ARIMA and LS-SVM models, the LS-SVM-ARIMA combined model based on WPD had the highest prediction accuracy. The mean absolute and root mean square errors were 2.4% and 2.0%, respectively. **Conclusion:** Low error measures prove the validity of this new prediction method and the combined model could be used for inference of future H7N9 avian influenza virus cases. Live poultry markets should be closed in late spring and early summer, and comprehensive H7N9 immunisation continued.

Keywords: epidemiological survey, aetiological virus detection monitoring, wavelet packet decomposition, LS-SVM-ARIMA model, PSO algorithm.

Introduction

H7N9 avian influenza is a disease caused by the novel H7N9 recombinant avian influenza virus, which is mainly characterised by pulmonary inflammation and fever (7). In early 2013, human infection with H7N9 avian influenza virus was discovered for the first time in China. Subsequently, birds infected with the virus were found in live poultry markets (25). At the beginning of 2017, a highly pathogenic variant of it was found in poultry flocks in some Chinese provinces, which then caused a number of outbreaks (32). In order to control H7N9 avian influenza in commercial flocks, China imposed treatment with a bivalent inactivated vaccine of recombinant avian influenza virus (H5+H7) in September

2017 (24). Both low-pathogenic and high-pathogenic strains of H7N9 avian influenza virus coexist and are still evolving at present (5). Therefore, the risk of its spread among various animals is not countered fully by vaccination and should not be ignored.

It is very important to construct predictive models for the early detection of H7N9 outbreaks (22), and many studies have established them. For example, Gilbert *et al.* (9) constructed a spatial econometric model to predict the risk of H7N9 avian influenza infection in live poultry markets in Asia, Young *et al.* (33) constructed a niche model to predict the prevalence of H5N1 and H9N2 avian influenza in poultry flocks, and Li *et al.* (16) developed a Bayesian inference system to predict the infection rate of H7N9 also in poultry flocks. Qiang and

Kou (23) used the wavelet packet decomposition (WPD) method to predict the transmission of avian influenza virus between humans and various poultry species, Walsh *et al.* (28) adopted the gradient-boosted tree to build a model for predicting the probability of isolating avian influenza virus from wild bird samples, and lastly, Yang *et al.* (31) predicted the spread of H9N2 avian influenza with a system dynamics model. Although these previous studies achieved good results, most of them adopted the single-model prediction method. However, avian influenza virus infection is a complex biochemical process, characterised by periodicity, complexity and nonlinearity, which underscore the need for a combined model based on modern prediction theory to improve prediction accuracy (23).

Continuous monitoring of the distribution and changes in the positivity rates of H7N9 avian influenza virus has been significant for eliminating infection sources (3). Positive virus detection indicates viral infection in poultry, and the higher the positivity rate (number of positive results/number of samples), the greater the avian influenza virus infection (26). Infected poultry have been the main source of avian influenza virus. After infection, the poultry shed the virus through saliva, nasal secretions and faeces, causing the spread of the epidemic (14). Also, it is possible that contacts of domestic poultry with migratory birds, which are natural reservoirs for avian influenza viruses, are a potential source of infection (12). Relevant studies have shown that the rate of avian influenza virus detection correlated with outbreaks to some extent, *i.e.* the higher the positive rate, the greater the possibility of an outbreak (10). Some studies have used positive samples of H7N9 to analyse the transmission regularity of the virus (1, 11) and to judge the effect of closing live poultry markets (29) and comprehensive H7N9 immunisation of at-risk animals (30).

China has been a major exporter of poultry products, the trade in which may spread domestic avian influenza virus to importing countries (34). Furthermore, there is a wide range and distribution of bird migrations within and in proximity to China, which possibly causes the spread of domestic avian influenza to neighbouring countries (6). The outbreak risk of H7N9 avian influenza related to China has obliged the Chinese government to employ effective preparedness measures, including early detection and timely sharing of surveillance data (22). The Ministry of Agriculture and Rural Affairs of the People's Republic of China has organised provincial veterinary departments to conduct special surveillance efforts against H7N9 avian influenza. Since April 2013, fixed virus monitoring points have been established on each province's poultry farms and commercial livestock farms and in pig slaughterhouses and wild bird habitats. The collected samples are sent to the National Avian Influenza Reference Laboratory for testing. When a virus detection in animal or environmental samples occurs, a retrospective investigation is carried out in accordance with guidelines issued by the Ministry of Agriculture and Rural Affairs to identify the source of the virus and

possible contaminated sites and implement culling or treatment of pathogen-positive poultry flocks (18). As the official monitoring data are relatively authoritative, some studies have tried to use the annual avian influenza monitoring data of the United States (28) or Southeast Asia (17) to predict the prevalence of the virus. This study regards China and also employs national monitoring data for the virus positivity rates to analyse the distribution characteristics of positive samples and construct a predictive model conforming to the trend of changes in the positive rate. We believe that this study has novelty and important practical significance for improving the prediction of major animal epidemics and curbing the spread of the virus. It could also provide an important insight into how the global known and future zoonotic influenza threats may be reduced.

Material and Methods

Monitoring efforts for H7N9 avian influenza by the Ministry of Agriculture and Rural Affairs of the People's Republic of China involve collecting chicken, duck, goose, wild bird and pig samples from certain sites for serological and pathogenic analysis. It should be noted that since September 2017, China has implemented comprehensive H7N9 immunisation for all domestic poultry operations, and the original positivity rate of serological samples has been replaced by a qualified rate. Consequently, the statistical calibre of pathogenic monitoring data has not changed, so the data could be used as an observation indicator of positivity rates. The rates in samples from surveillance for this pathogen from April 2013 to April 2020 were collected from information published by the Ministry of Agriculture and Rural Affairs (see Supplementary Materials for details of raw data). In addition, the affected species, number of H7N9 avian influenza outbreaks, number of cases, number of fatalities and number of culled animals reported by the provinces were obtained from the animal epidemic monthly report in the ministry's Veterinary Bulletin (19). In order to have a clear understanding of the distribution pattern of non-human H7N9 virus detections in China and to conduct spatial autocorrelation analysis, the monthly positive rates greater than 0 were all set to 1, indicating that H7N9 virus had been detected. A value of 0 means that no virus detections were reported.

Moran's *I*. "Spatial autocorrelation" usually describes the correlation among the values of a given variable with respect to a dependence on the relative locations between spatial units (8). If "neighbours" tend to have similar values, it is generally accepted as a positive spatial autocorrelation and it is a negative spatial autocorrelation if they tend to have dissimilar values. One of the most prominent measures of spatial autocorrelation is Moran's *I* index (MI), which is defined as the ratio between the local and global coherence (20).

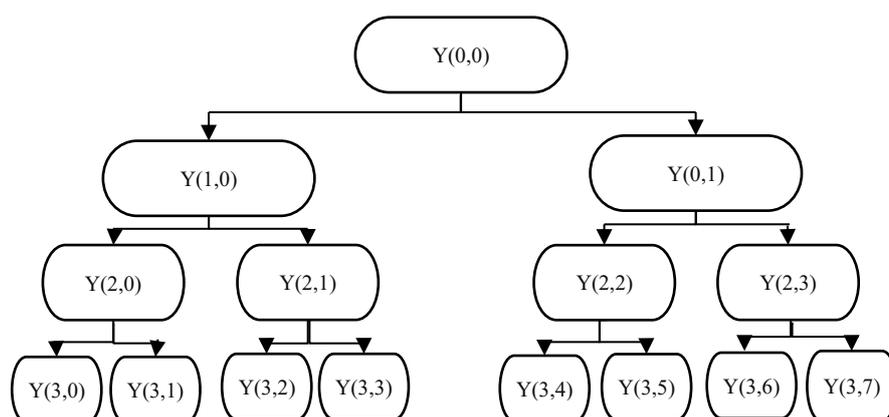


Fig. 1. Hierarchical Structure of WPD

Least squares support-vector machines model.

Our study adopted ArcGIS 10.2 software (Environmental System Research Institute, Redlands, CA, USA) and MI to analyse the spatial correlation between the number of months of H7N9 avian influenza virus detection in different provinces. The MI values ranged from -1 to 1 . Standardised statistics of Z and P values are generally used to test the significance of the spatial correlation.

Wavelet packet decomposition. WPD is a method for decomposing the high-frequency signal obtained by wavelet transform layer by layer. Theoretically, the effective information in a signal can be extracted by WPD into three layers (15). As shown in Fig. 1, the signal $Y = (0,0)$ represents the original data of positive rates of avian influenza virus detections. If the original data are non-stationary, they could be mapped into 2^3 subspaces by three-level WPD, from low to high frequency from left to right. $Y = (0, 1)$ and $Y = (1, 0)$ respectively represent the low-frequency and high-frequency components obtained from the first decomposition of the original data. $Y = (2, 0)$ and $Y = (2, 1)$ represent the low-frequency component obtained in the first decomposition, and then the low-frequency and high-frequency components are obtained in the second decomposition, and finally the data are decomposed into a low-frequency trend sequence and a high-frequency trend sequence.

Least squares support-vector machines (LS-SVM) is a simplified and improved version of the standard SVM model. Its greatest advantage is dealing with non-stationary series and overcoming the problem of overestimated prediction results from traditional prediction methods (27). In this paper, Matlab 2018a software (MathWorks, Natick, MA, USA) and LS-SVM lab Toolbox version 1.8 (<http://www.esat.kuleuven.be/sista/lssvmlab>) were used to model the positive sequence data of avian influenza virus. The algorithm principle is as follows:

Suppose there are n groups of data points, and a sample set

$$S = \{ (x_1, y_1), \dots, (x_l, y_l), x_i \in R^n, y_i \in R \}_{i=1}^l$$

In the formula, l is the number of samples, x_i is the input vector, and y_i is the corresponding output value. The linear regression function of the sample set is

$$f(x) = \omega^T \phi(x) + b$$

In the formula, ω is the weight vector, $\omega_i \in R^n$; the offset constant is $b \in R$; and $\phi(\cdot)$ is a kernel function for solving nonlinear problems.

Based on the principle of structural risk minimisation, the optimisation problem is defined as

$$\min_{J_{LS}}(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} \gamma \sum_{i=1}^l \xi^2$$

In the formula, $\|\omega\|^2$ and γ represent model complexity and normalisation parameters, respectively.

According to the Karush–Kuhn–Tucker conditions and Mercer's condition, the optimisation problem of LS-SVM is transformed into the solution of the linear equation, and the nonlinear regression equation of the LS-SVM model at the x moment is obtained which is

$$f(x) = \sum_{i=1}^l a_i K(x, x_i) + b$$

In the formula, x is the support vector; x_i is the input of the i support vector; $f(x)$ is the predicted output value; and $K(x, x_i)$ is the kernel function mapping the sample to the feature vector. The kernel function adopts the RBF formula as follows:

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma}}$$

In the formula, σ is the hyperparameter of the radial basis function kernel.

Particle swarm optimisation algorithm. The prediction effect of the LS-SVM model is affected by the penalty factor C and the kernel function σ . If the values of C and σ are improper, then overlearning can easily occur, which is not conducive to sample regression. Particle Swarm Optimisation (PSO) is an intelligent algorithm based on population search, which is used to optimise the parameters of the LS-SVM model due to its simplicity and lack of many parameters that need to be adjusted. Our study intended to use the PSO algorithm source code of Matlab 2018a software to optimise the parameters of the LS-SVM model. The algorithm

principle of PSO is as follows (13): a group of random particles is first initialised, and then the characteristics of the particles are represented by position, speed and fitness. The particles update the speed and position according to the formulas, and they are substituted into the optimisation objective function to calculate the corresponding fitness value in order to evaluate the advantages and disadvantages. In each iteration, particles are updated by the individual extremum $Pbest$ and the population extremum $Gbest$. The individual extremum $Pbest$ is the position with the best fitness value which the particle has had, while the population extremum $Gbest$ is the best position which all particles in the group have had. These two extrema are used for the particle to find the optimal position in each iteration process, and then the particle will update the speed and fitness.

$$V_i^{k+1} = \omega V_i^k + c_1 rand() (Pbest_i^k - X_i^k) + c_2 rand() (Gbest_i^k - X_i^k)$$

$$X_i^{k+1} = X_i^k + V_i^{k+1}$$

In the formulas, $i = 0, 1, \dots, N$, N is the total number of particles; $k = 0, 1, \dots, M$, M is the maximum number of iterations; V_i^k is the velocity of the i particle in the k iteration; X_i^k is the position of the i particle in the k iteration; ω is the inertia factor; $rand()$ is a random number between 0 and 1; and c_1 and c_2 are learning factors.

Autoregressive integrated moving average model. In this article, the autoregressive integrated moving average (ARIMA) model was constructed by Eviews 10.0 software (IHS Markit, London, UK) combined with the time series of positivity rates of avian influenza virus. The model is a linear combination of past errors and past values of a stationary time series. It is a classical time series analysis method with high short-term prediction accuracy (21). The general mathematical expression is as follows:

$$\hat{y}_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t - (\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q})$$

In the formula, \hat{y}_t is the time series of t moment; $\varphi_1, \varphi_2, \dots, \varphi_p$ is the autoregressive coefficient; $\theta_1, \theta_2, \dots, \theta_q$ is the moving average coefficient; ε_t is the time residual sequence; p is the autoregressive order; and q is the moving average order.

Results

Situation of H7N9 avian influenza epidemic in China. The first case of H7N9 avian influenza came to notice in Inner Mongolian chickens in June 2017, and a total of 10 outbreaks of H7N9 avian influenza had occurred by April 2020 (Table 1). Among them, there were 4, 5 and 1 outbreak of H7N9 avian influenza in 2017, 2018 and 2019, respectively. The epidemic in 2017 was the most serious, with the largest number of cases, fatalities and culled animals. In terms of the number of provinces, the most were affected in 2018. From the perspective of time, H7N9 avian influenza occurrences were mainly concentrated in a March-to-June period, in which June was the month with the largest number of outbreaks, cases, fatalities and culled animals. Chickens were vulnerable to avian influenza, and 90% of the outbreaks were related to this species. Once the poultry got sick, the mortality rate was relatively high, usually approximately 70%.

The prevalence of H7N9 avian influenza virus in China and spatial autocorrelation. The high-risk regions of H7N9 avian influenza were mainly located in the southeast of China from 2013 to 2020 (Table 2). Among them, Guangdong, Fujian and Zhejiang were the three provinces with the most detections of H7N9 avian influenza virus. Except for Liaoning province with a cumulative number of virus positive months of more than three, the cumulative number of virus-positive months in other northern regions was less than three.

Table 1. Situation of H7N9 avian influenza epidemic from April 2013 to April 2020

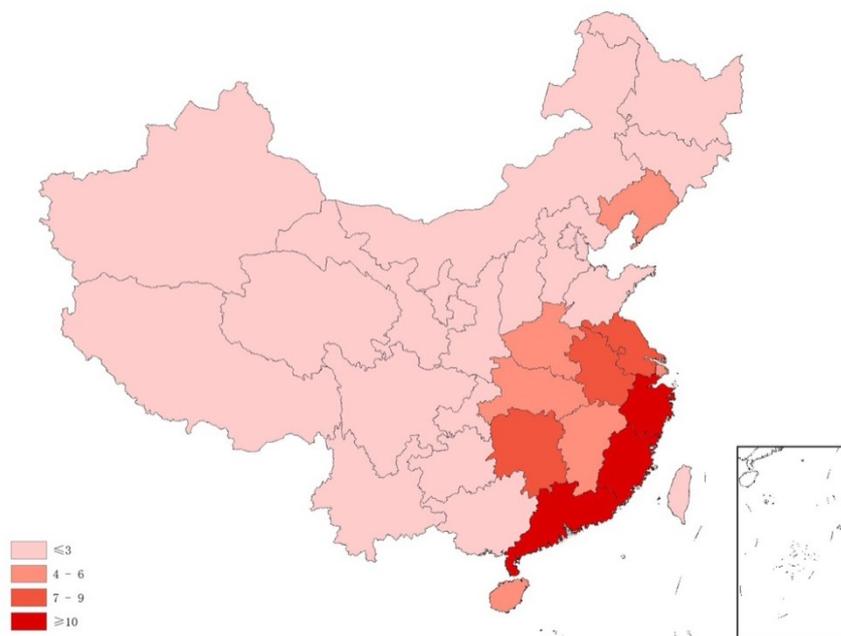
Time	Province	Number of outbreaks	Poultry	Number of cases	Number of fatalities	Number of animals culled
June 2017	Inner Mongolia	2	Chicken	63,406	37,582	424,197
June 2017	Heilongjiang	1	Chicken	20,150	19,500	16,610
August 2017	Anhui	1	Chicken	1,368	910	74,463
March 2018	Shaanxi	1	Chicken	1,000	810	1,000
April 2018	Shanxi	1	Chicken	812	699	6,374
April 2018	Ningxia	1	Chicken	1,200	585	13,578
May 2018	Liaoning	1	Chicken	11,000	9,000	8,000
May 2018	Ningxia	1	Chicken	3,000	2,210	86,000
March 2019	Liaoning	1	Peacock	9	9	191

Data source: *Veterinary Bulletin*

Table 2. Spatial autocorrelation analysis of H7N9 avian influenza virus detections in China

Variable	Moran's <i>I</i>	Z score	P value
From 2013 to 2020	-0.014	0.387	0.350
In 2013	-0.019	0.259	0.398
In 2014	-0.046	-0.362	0.359
In 2015	0.042	1.783	0.037
In 2016	-0.064	-0.992	0.161
In 2017	-0.034	-0.075	0.470
In 2018	-0.007	0.607	0.272
In 2019	-0.021	0.239	0.406
In 2020	—	—	—
In January	-0.008	0.507	0.306
In February	-0.008	0.526	0.299
In March	-0.006	0.563	0.287
In April	-0.004	0.586	0.279
In May	-0.040	-0.210	0.417
In June	-0.045	-0.350	0.363
In July	0.009	1.103	0.135
In August	—	—	—
In September	-0.073	-0.996	0.160
In October	-0.036	-0.164	0.435
In November	-0.046	-0.366	0.357
In December	0.072	2.320	0.010
Before comprehensive animal H7N9 immunisation	-0.020	0.255	0.399
After-comprehensive animal H7N9 immunisation	-0.028	0.058	0.477

No H7N9 avian influenza virus detections occurred in China in 2020 or in August of each year, so there are no spatial autocorrelation analysis results presented for them

**Fig. 2** Regional distribution of H7N9 avian influenza virus detections in China

The annual regional distributions of H7N9 avian influenza virus detections in different years are shown in Fig. 3. In 2013, the virus was mainly detected in southeast China, and by 2016 the positive rate had reduced. Due to the large number of live poultry markets and frequent migration of wild birds, the time span of H7N9 virus detections in animals in Guangdong province exceeded six months in each year from 2013 to 2016. In 2017, owing to the coexistence of low

pathogenicity and high pathogenicity H7N9 influenza virus strains in various animals, the area of virus detection expanded, and the number of areas with virus detections recorded for more than six months increased to an unprecedented level. In 2018, only three southern provinces were found to be positive for the virus, while in 2019, the detection areas were mainly in Inner Mongolia and Liaoning provinces, and after 2020, no virus was detected.

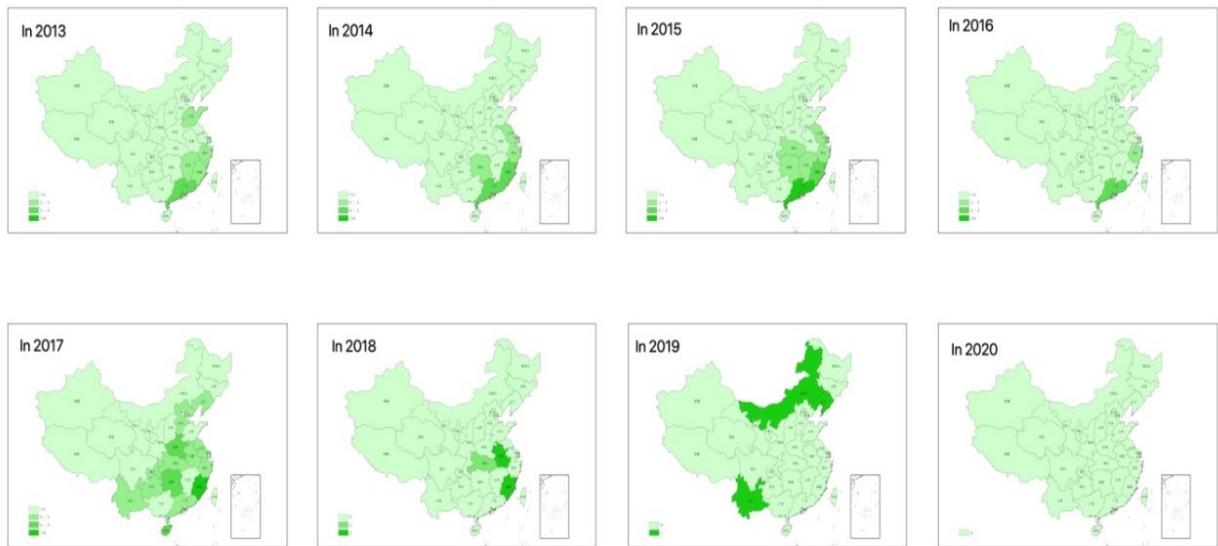


Fig. 3. Regional distribution of H7N9 avian influenza virus in China from 2013 to 2020

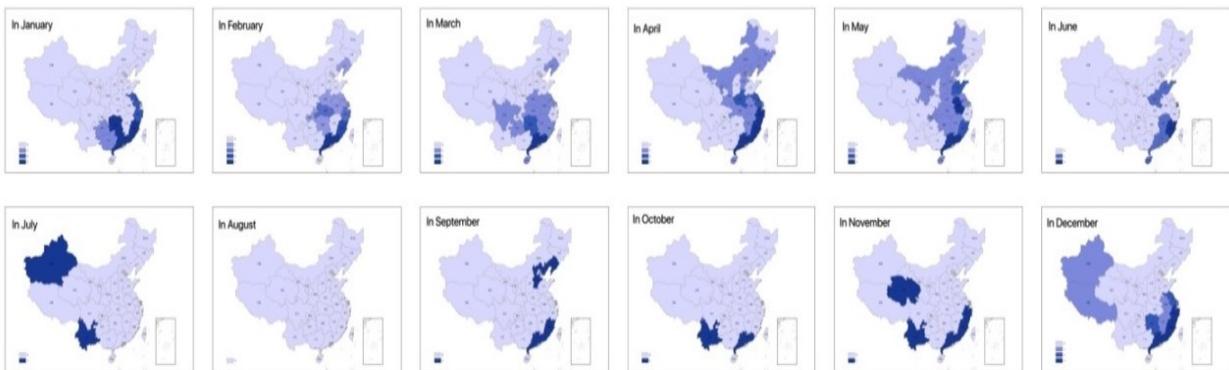


Fig. 4. Regional distribution of cumulative H7N9 avian influenza virus detections in China from January to December between 2013 and 2020

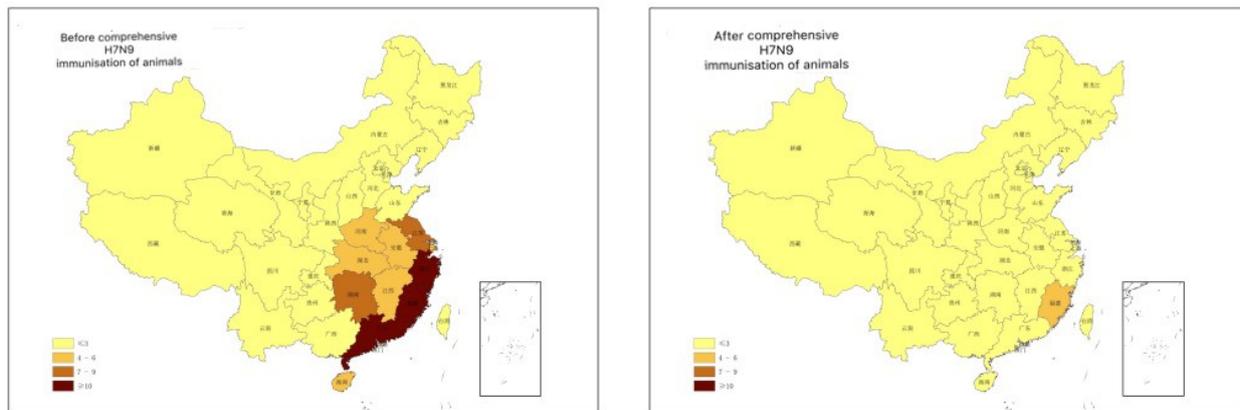


Fig. 5. Regional distribution of H7N9 avian influenza virus detections before and after comprehensive H7N9 immunisation of at-risk animals in China

Fig. 4 shows that March, April and May were the months in which samples often tested positive in many areas, while in August, September and October, the virus could be detected in fewer areas. This pattern indicates that the H7N9 virus was most prevalent in late spring and early summer. Although the temperature increased in this period, there was still a high risk of H7N9 outbreak in animals.

As can be seen in Fig. 5, before the national animal comprehensive H7N9 immunisation programme began in September 2017, most regions were positive for

H7N9 virus, especially Guangdong, Fujian and Zhejiang. The cumulative positive rate of H7N9 avian influenza virus was more than 10 months in these provinces. After immunisation, the poultry produced antibodies, which reduced the positivity rate. Other than Fujian province, there were no areas where H7N9 avian influenza virus had been detected for more than four months. This indicates that the comprehensive H7N9 immunisation effort had a significant effect on reducing the positivity rate of H7N9 avian influenza.

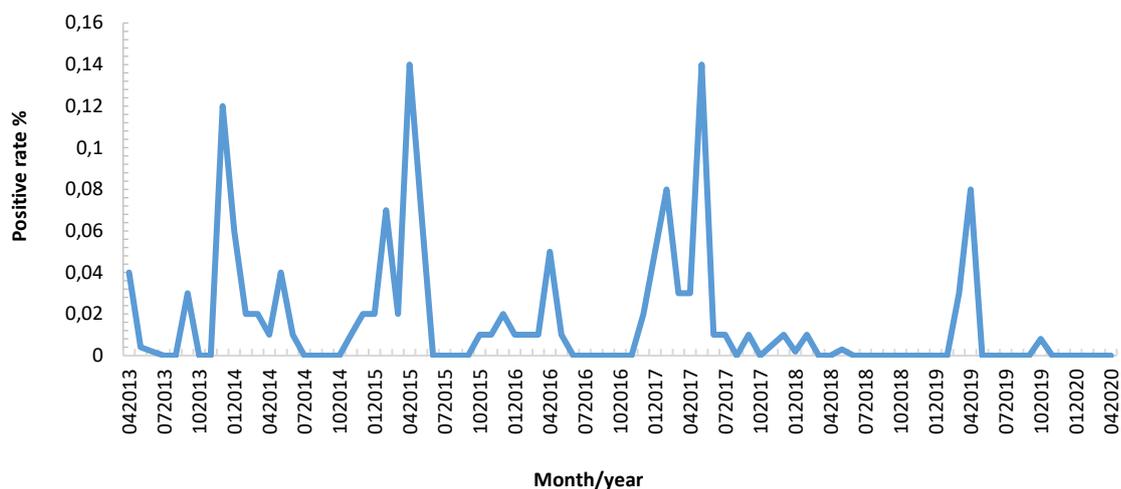


Fig. 6. Temporal distribution of positive rate of animal H7N9 avian influenza virus detections in China from 2013 to 2020

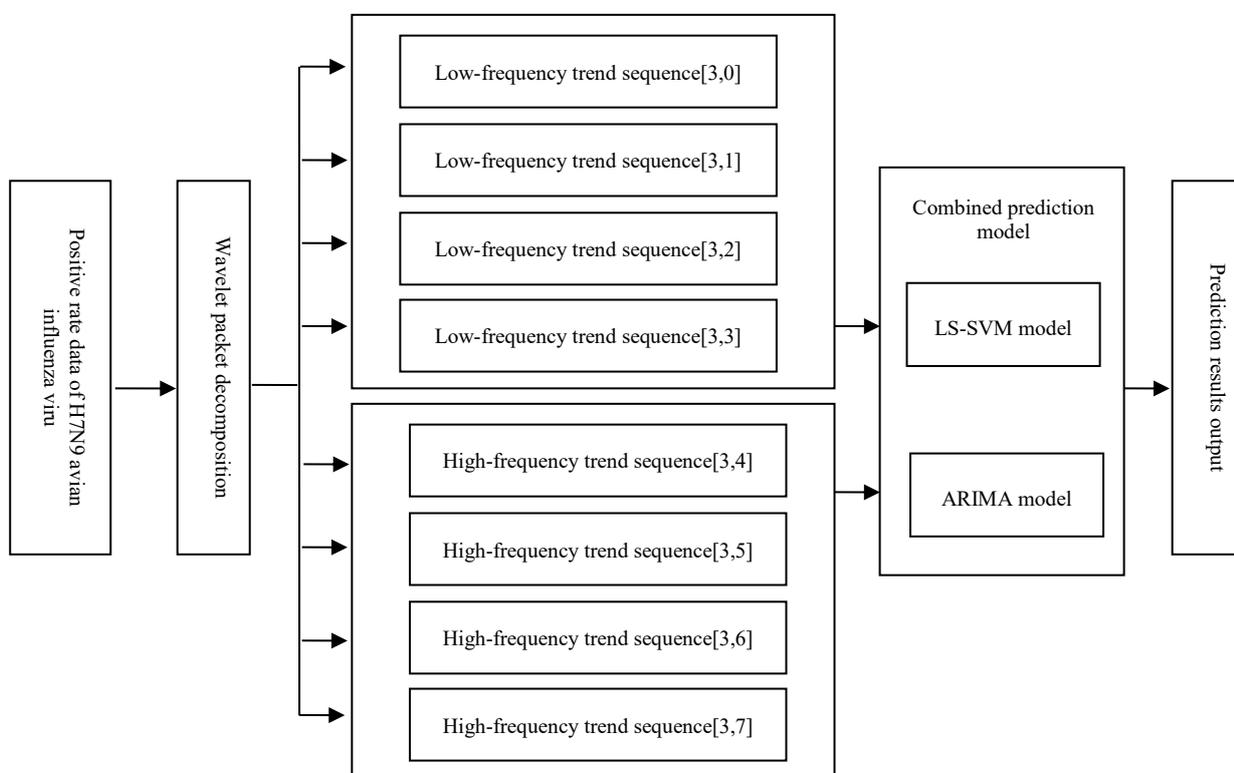


Fig. 7. Steps for constructing the LS-SVM-ARIMA combined model based on WPD
 LS-SVM – least squares support-vector machines; ARIMA – autoregressive integrated moving average

The MI values of H7N9 avian influenza virus detections are shown in Table 2. There was no significant spatial autocorrelation among such detections in 31 provinces from 2013 to 2020. Moreover, the positive spatial correlation of avian influenza virus detections before and after the comprehensive H7N9 immunisation programme was not significant. Among different years, only virus detections in 2015 had significant spatial autocorrelation ($P < 0.05$); and among different months, only December had ($P < 0.05$).

WPD of time series data of H7N9 avian influenza virus detections. Fig. 6 shows that the overall

positivity rate of the virus detections fluctuated dramatically, with a maximum of 0.14% (from April 2015 to May 2017). However, all H7N9 avian influenza virus tests were negative for six consecutive months from November 2019 to April 2020, indicating that no animal was infected with H7N9 avian influenza virus in China during this period.

The augmented Dickey–Fuller (ADF) test results showed that the time series of the positive rate of H7N9 avian influenza virus detections failed the significance test of the coefficient at the significance levels of 1%, 5%, and 10%. Since this indicated that the time series

was not smooth, an LS-SVM-ARIMA combined model based on WPD was then constructed. The specific modelling steps are shown in Fig. 7.

Given the limited data, our study used the monthly data from April 2013 to April 2019 to build the combined model, and the data from May 2019 to April 2020 as the prediction data. The monthly data from April 2013 to April 2019 were decomposed by a three-layer

wavelet packet by Db4 wavelet, and the positive rate data of eight decomposition spaces were obtained. Among them, the four low-frequency trend sequences (*i.e.* [3,0], [3,1], [3,2] and [3,3]) and waveforms are shown in Fig. 8, while the four high-frequency trend sequences (*i.e.* [3,4], [3,5], [3,6] and [3,7]) and waveforms are shown in Fig. 9.

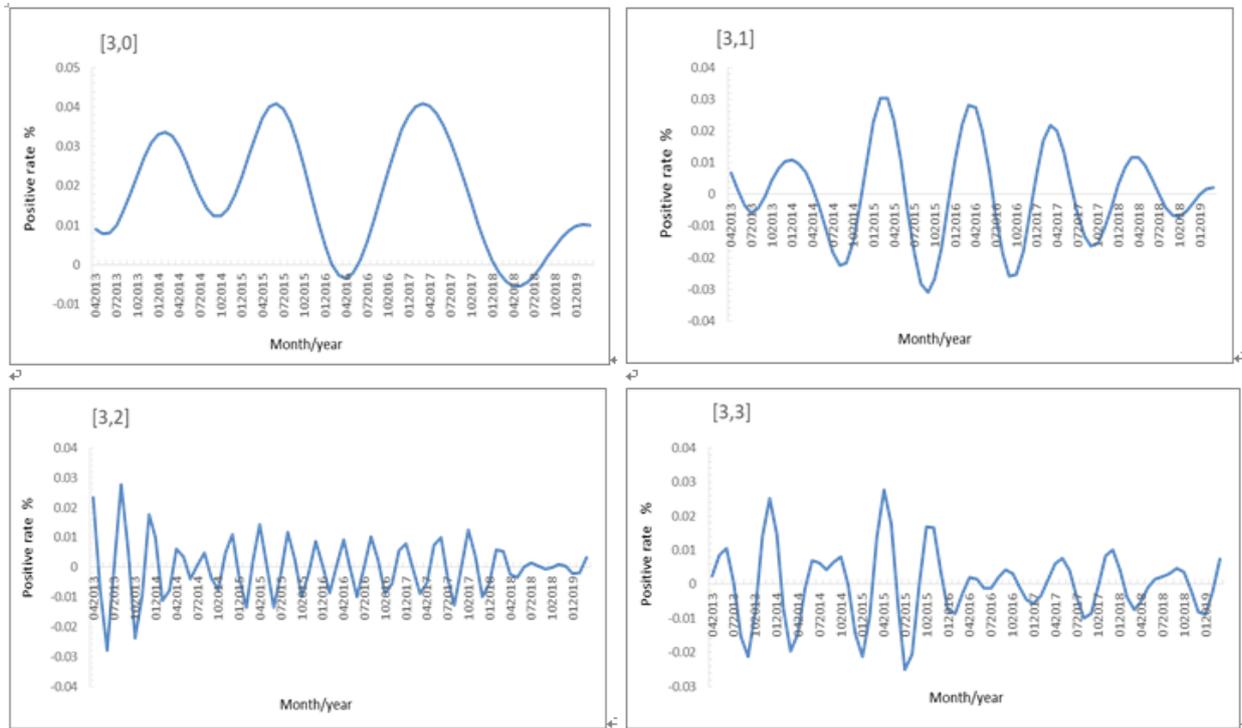


Fig. 8. Low-frequency trend sequence decomposition of positive rates of H7N9 avian influenza virus from April 2013 to April 2019

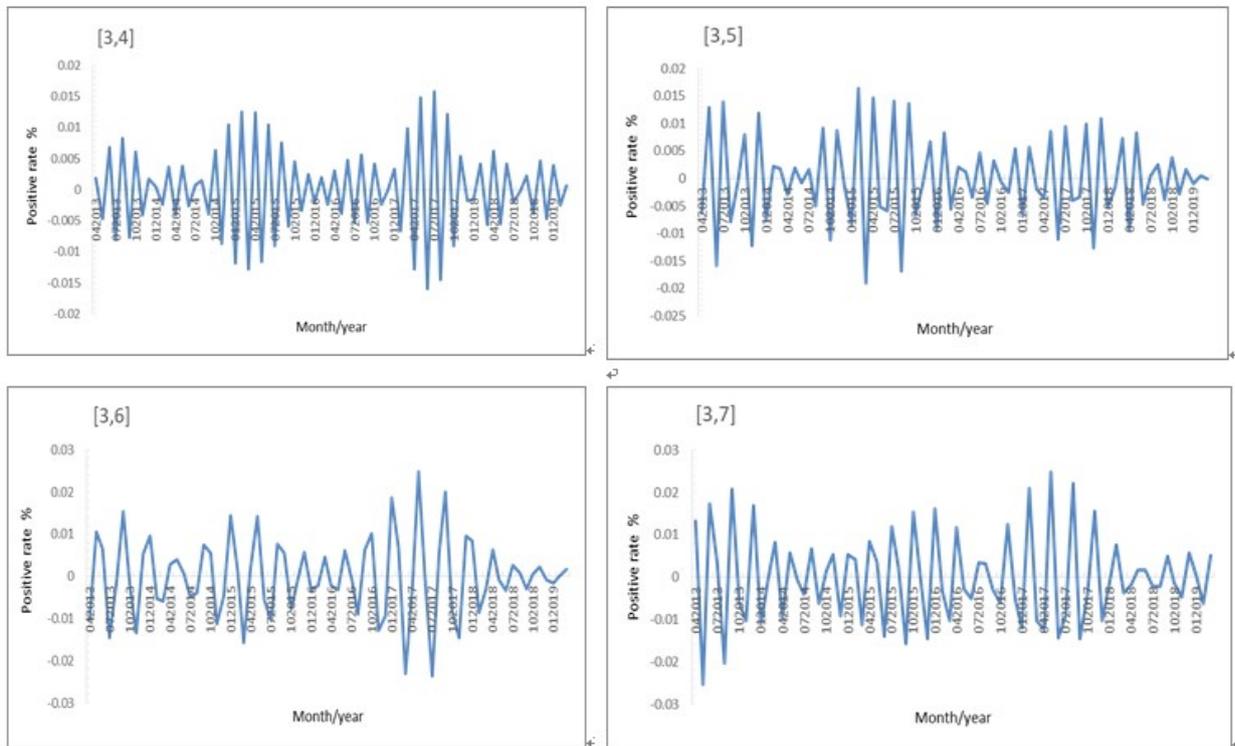


Fig. 9. High-frequency trend sequence decomposition of positive rates of H7N9 avian influenza virus detections from April 2013 to April 2019

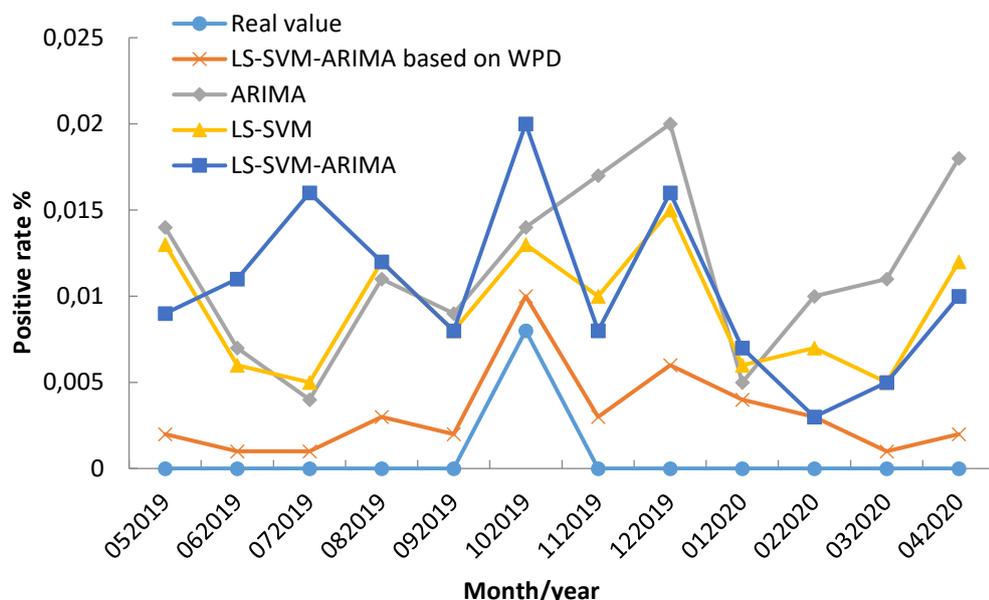


Fig. 10. Comparisons of predicted values from each model with the true values
LS-SVM – least square support-vector machines; WPD – wavelet packet decomposition; ARIMA – autoregressive integrated moving average

Table 3. Optimisation results of low-frequency trend sequence parameters

Decomposition sequence	Hyperparameters σ	Normalisation parameters γ
[3,0]	9506	7350.8
[3,1]	2514.5	3946.2
[3,2]	9113.6	9740.8
[3,3]	818.709	7242.1

Table 4. ARIMA model parameters

Decomposition sequence	ARIMA model parameters
[3,4]	ARIMA (1,0,0)*(2,0,0) [12]
[3,5]	ARIMA (3,0,0)
[3,6]	ARIMA (0,0,0)
[3,7]	ARIMA (3,0,0)

ARIMA – autoregressive integrated moving average; ARIMA (1,0,0)*(2,0,0) [12] is the multiplicative seasonal model of ARIMA

Modelling of low-frequency trend sequence of positive rates of H7N9 avian influenza virus. The LS-SVM model parameter optimisation calculation was used on the four low-frequency trend sequences of [3,0], [3,1], [3,2] and [3,3] by the PSO algorithm. The specific steps are as follows: first, the order of the model was adopted to determine the influence of the positive rates for the previous months on the positive rate of the current year. The minimum mean square deviation of the predicted results was taken as the standard for further order expansion. By calculation, the final order of the model was 12, indicating that the current data in the sequence were closely related to the data of the previous 12 months (the 12-month data from before April 2019). Therefore, the data from May 2018 to April 2019 were used as the input values of the LS-SVM model to predict the data for May 2019 to April 2020. Then the PSO algorithm was used to optimise the hyperparameters σ

and normalisation parameters γ of the radial basis function kernel of the model. The search ranges of the two parameters were set to 10^{-12} – 10^{12} . The optimal combination of the two parameters was obtained which is shown in Table 3.

Modelling of high-frequency trend sequence of positive rates of H7N9 avian influenza virus. For the four high-frequency trend sequence data of [3,4], [3,5], [3,6] and [3,7], the unit root test of each time period sequence was first conducted by ADF to verify whether the sequence was stable. If not, the model needed to be stabilised, and the P and Q values of the model were determined according to the autocorrelation function, partial autocorrelation function, minimum Bayesian information criterion and maximum R^2 after stabilisation. As shown in Table 4, the ARIMA model of the four high-frequency trend sequences (*i.e.* [3,4], [3,5], [3,6] and [3,7]) were ARIMA (1,0,0) * (2,0,0) [12], ARIMA (3,0,0), ARIMA (0,0,0) and ARIMA (3,0,0).

Evaluation of the prediction effectiveness of different models. The prediction value of each sequence was obtained by modelling with each decomposed time sequence. The final prediction value of the positive avian influenza virus sequence was the linear superposition of the predicted values of each decomposition sequence. The formula is as follows:

$$\hat{y} = \hat{y}_{[3,0]} + \hat{y}_{[3,1]} + \hat{y}_{[3,2]} + \hat{y}_{[3,3]} + \hat{y}_{[3,4]} + \hat{y}_{[3,5]} + \hat{y}_{[3,6]} + \hat{y}_{[3,7]}$$

In the formula, \hat{y} is the predicted value of combined model; and $\hat{y}_{[3,i]}$ is the sequence data of time interval [3, i], $i = 0, 1, 2, \dots, 7$.

According to the formula above, the predicted values of each decomposed sequence were accumulated. As shown in Table 5, the prediction results of the LS-SVM-ARIMA combined model based on WPD were obtained.

Table 5. Prediction results of the LS-SVM-ARIMA combined model based on WPD

Date	Real value	[3,0]	[3,1]	[3,2]	[3,3]	[3,4]	[3,5]	[3,6]	[3,7]	Predictive value
05/2019	0	0	0	0.001	0	0	0.001	0	0	0.002
06/2019	0	0	0	0	0	0	0	0	0.001	0.001
07/2019	0	0	0.001	0	0	0	0	0	0	0.001
08/2019	0	0.001	0	0	0.001	0.001	0	0	0	0.003
09/2019	0	0	0	0	0.001	0	0.001	0	0	0.002
10/2019	0.001	0.001	0.002	0.001	0.001	0.001	0.002	0	0.001	0.010
11/2019	0	0.001	0.001	0	0	0.001	0	0	0	0.003
12/2019	0	0.001	0	0.001	0.001	0	0.001	0	0.002	0.006
01/2020	0	0	0.001	0	0.001	0.001	0.001	0	0	0.004
02/2020	0	0.001	0	0	0	0.002	0	0	0	0.003
03/2020	0	0	0.001	0	0	0	0	0	0	0.001
04/2020	0	0	0.001	0	0	0.001	0	0	0	0.002

LS-SVM-ARIMA – least square support-vector machines–autoregressive integrated moving average, WPD – wavelet packet decomposition

Table 6. Prediction results based on the three models: ARIMA, LS-SVM and ARIMA-LS-SVM

Date	Real value	ARIMA model	LS-SVM model	LS-SVM-ARIMA model
05/2019	0	0.014	0.013	0.009
06/2019	0	0.007	0.006	0.011
07/2019	0	0.004	0.005	0.016
08/2019	0	0.011	0.012	0.012
09/2019	0	0.009	0.008	0.008
10/2019	0.008	0.014	0.013	0.020
11/2019	0	0.017	0.010	0.008
12/2019	0	0.020	0.015	0.016
01/2020	0	0.005	0.006	0.007
02/2020	0	0.010	0.007	0.003
03/2020	0	0.011	0.005	0.005
04/2020	0	0.018	0.012	0.010

ARIMA – autoregressive integrated moving average; LS-SVM – least square support-vector machines

Table 7. Evaluation of the prediction results of the different models

Evaluation index	LS-SVM-ARIMA combined model based on WPD	ARIMA model	LS-SVM model	LS-SVM-ARIMA model
MAE	0.024	0.119	0.098	0.077
RMSE	0.020	0.067	0.059	0.041

LS-SVM – least square support-vector machines; ARIMA – autoregressive integrated moving average; WPD – wavelet packet decomposition

In order to verify the predictive effectiveness of the combined model, this study adopted the same data to predict the ARIMA model, the LS-SVM model and the LS-SVM-ARIMA model. The LS-SVM model and the LS-SVM ARIMA model were optimised by the PSO algorithm. Table 6 and Fig.10 show that the LS-SVM-ARIMA combined model based on WPD had the highest degree of data fitting.

In this section, the mean absolute error (MAE) and root mean square error (RMSE) of the true and predicted values were employed as evaluation indices of the prediction effectiveness of each model. The MAE and RMSE formulas are as follows:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

In both formulas, y_i is the true value; and \hat{y}_i is the predicted value, $i = 1, 2, 3, \dots, n$.

The values in Table 7 show that the MAE and RMSE of predictions from the LS-SVM-ARIMA combined model based on WPD were the smallest, which indicated that neither the single ARIMA model nor the LS-SVM model was likely to capture the linear

and nonlinear characteristics in the data of avian influenza virus detections. Although the LS-SVM-ARIMA model had the advantages of both the ARIMA model and the LS-SVM model, the LS-SVM-ARIMA combined model based on WPD could more comprehensively extract the detailed information contained in the time series, so it better reflected the law of the positive rate change with time, and predicted the positive rate more accurately.

Discussion

The LS-SVM-ARIMA combined model based on WPD was a beneficial enhancement for the more accurate prediction of the positivity rate for H7N9 avian influenza virus. However, the prediction results of LS-SVM-ARIMA based on WPD model were not fully consistent with the real values. This difference arises because although the monitoring data of positivity rates are applicable to predicting the presence of avian influenza virus, they are limited by the number of monitoring sites and the effectiveness of sample collection nationwide. In addition, the virus-positive rate is cross-influenced by many factors, so it is difficult for a single data source to reflect the full complexity of the epidemic facts. Astill *et al.* (2) suggested that monitoring data, internet data and environmental data should be integrated to support the construction of an avian influenza prediction model. Therefore, we can propose a few improvements to the predictive ability for the positivity rate (1). Although the combined model has a better predictive ability than the single model generally, more combined models need to be produced and the prediction results compared among them, so as to further verify the effectiveness of the combined model based on WPD (2). Expanding the scope of monitoring, increasing the number of monitoring points, and improving the efficiency of sample collection and detection are necessary to make the published data more accurate (3). Predicting positive rates for the virus in all provinces will gradually facilitate the elimination of H7N9 avian influenza in different regions.

According to the monitoring data from April 2013 to April 2020, there was a greater risk of epidemics in late spring and early summer. After livestock at risk was comprehensively immunised against H7N9, the positive rates of infection with the virus were significantly reduced. Except for 2015 and December of each year, the positive samples did not have spatial clustering and were randomly distributed. Therefore, it is recommended that live poultry markets be closed in late spring and early summer, comprehensive animal H7N9 immunisation be continued, and monitoring programs strengthened. Although the tested positivity rate of the virus during 2020 was lower than in recent years, H7N9 avian influenza may still threaten the poultry industry in China. Even with the measures of market closure, immunisation, and vigilant monitoring, H7N9 will not cease to be a latent threat and prediction capability is

highly desirable for preparedness for this threat. Therefore, our epidemiological survey proposed a combined LS-SVM-ARIMA model based on WPD to improve predictive accuracy. The MAE of 2.4% and RMSE of 2.0% prove the validity of this new prediction method.

Supplementary Material: Raw H7N9 avian influenza virus positivity rate data from the Veterinary Bulletin of the Ministry of Agriculture and Rural Affairs of the People's Republic of China.

Conflict of Interests Statement: The authors declare that there is no conflict of interests regarding the publication of this article.

Financial Disclosure Statement: This study was supported by the National Natural Science Foundation of China Youth Program (71804078).

Animal Rights Statement: None required.

Acknowledgements: We thank Yuan Wang and Yuyang Li for assistance in data collection and the support from the University of International Business and Economics, China.

References

- Abolnik C.: A current review of avian influenza in pigeons and doves. *Vet Microbiol* 2014, 170, 181–196, doi: 10.1016/j.vetmic.2014.02.042.
- Astill J., Dara R.A., Fraser E.D.G., Sharif S.: Detecting and predicting emerging disease in poultry with the implementation of new technologies and big data: a focus on avian influenza virus. *Front Vet Sci* 2018, 5, 263, doi: 10.3389/fvets.2018.00263.
- Bi Y., Chen Q., Wang Q., Chen J., Jin T., Wong G., Quan C., Liu J., Wu J., Yin R., Zhao L., Li M., Ding Z., Zou R., Xu W., Li H., Wang H., Tian K., Fu G., Huang Y., Shestopalov A., Li S., Xu B., Yu H., Luo T., Liu L., Xu X., Luo Y., Liu Y., Shi W., Liu D., Gao G.F.: Genesis, evolution and prevalence of H5N6 avian influenza viruses in China. *Cell Host Microbe* 2016, 20, 810–821, doi: 10.1016/j.chom.2016.10.022.
- Chowell G., Simonsen L., Towers S., Miller M.A., Viboud C.: Transmission potential of influenza A/H7N9, February to May 2013, China. *BMC Med* 2013, 11, 214, doi: 10.1186/1741-7015-11-214.
- de Vries R.D., Herfst S., Richard M.: Avian influenza a virus pandemic preparedness and vaccine development. *Vaccine (Basel)* 2018, 6, 46, doi: 10.3390/vaccines6030046.
- Feare C.J.: The role of wild birds in the spread of HPAI H5N1. *Avian Dis* 2007, 51, 440–447, doi: 10.1637/7575-040106R1.1.
- Gao R., Cao B., Hu Y., Feng Z., Wang D., Hu W., Chen J., Jie Z., Qiu H., Xu K., Xu X., Lu H., Zhu W., Gao Z., Xiang N., Shen Y., He Z., Gu Y., Zhang Z., Yang Y., Zhao X., Zhou L., Li X., Zou S., Zhang Y., Li X., Yang L., Guo J., Dong J., Li Q., Dong L., Zhu Y., Bai T., Wang S., Hao P., Yang W., Zhang Y., Han J., Yu H., Li D., Gao G.F., Wu G., Wang Y., Yuan Z., Shu Y.: Human infection with a novel avian-origin influenza A (H7N9) virus. *N Engl J Med* 2013, 368, 1888–1897, doi: 10.1056/NEJMoa1304459.
- Getis A.: Reflections on spatial autocorrelation. *Reg Sci Urban Econ* 2007, 37, 491–496, doi: 10.1016/j.regsciurbeco.2007.04.005.
- Gilbert M., Golding N., Zhou H., Wint Z.G.R., Robinson T.P., Tatem A.J., Lai S., Zhou S., Jiang H., Guo D., Huang Z., Messina J.P., Xiao X., Linard C., Van Boeckel T.P., Martin V., Bhatt S., Gething P.W., Farrar J.J., Hay S.I., Yu H.: Predicting the risk of

- avian influenza A H7N9 infection in live-poultry markets across Asia. *Nat Commun* 2014, 5, 4116, doi: 10.1038/ncomms5116.
10. Guo X., Yang D., Liu R., Li Y., Hu Q., Ma X., Li Y., Zhang H., Zhang X., Zhao B., Chen T.: Detecting influenza and emerging avian influenza virus by influenza and pneumonia surveillance systems in a large city in China, 2005 to 2016. *BMC Infect Dis* 2019, 19, 825, doi: 10.1186/s12879-019-4405-5.
 11. Han J., Jin M., Zhang P., Liu J., Wang L., Wen D., Wu X., Liu G., Zou Y., Lv X., Dong X., Shao B., Gu S., Zhou D., Leng Q., Zhang C., Lan K.: Epidemiological link between exposure to poultry and all influenza A (H7N9) confirmed cases in Huzhou city, China, March to May 2013. *Euro Surveill* 2013, 18, 20481.
 12. He Y., Liu P., Tang S., Chen Y., Pei E., Zhao B., Ren H., Li J., Zhu Y., Zhao H., Pan Q., Gu B., Yuan Z., Wu F.: Live poultry market closure and control of avian influenza A (H7N9), Shanghai, China. *Emerging Infect Dis* 2014, 20, 1565–1566, doi: 10.3201/eid2009.131243.
 13. Kennedy J., Eberhart R.: Particle swarm optimization. In: *Vol. 4, Proceedings of the IEEE International Conference on Neural Networks*, Perth, Australia, IEEE Press, Hoboken, 1995, pp. 1942–1948, doi: 10.1109/icnn.1995.488968.
 14. Kim S., Kim Y.-I., Pascua P.N.Q., Choi Y.K.: Avian influenza A viruses: evolution and zoonotic infection. *Semin Respir Crit Care Med* 2016, 37, 501–511, doi: 10.1055/s-0036-1584953.
 15. Liò P.: Wavelets in bioinformatics and computational biology: state of art and perspectives. *Bioinformatics* 2003, 19, 2–9, doi: 10.1093/bioinformatics/19.1.2.
 16. Li R., Bai Y., Heaney A., Kandula S., Cai J., Zhao X., Xu B., Shaman J.: Inference and forecast of H7N9 influenza in China, 2013 to 2015. *Eurosurveill* 2017, 22, 30462, doi: 10.2807/1560-7917.ES.2017.22.7.30462.
 17. Millman A.J., Havers F., Luliano A.D., Davis C.T., Borann S., Sovann L., Chin S., Corwin A.L., Vongphrachanh P., Douanggeun B., Lindblade K.A., Chittaganpitch M., Kaewthong V., Kile J.C., Nguyen H.T., Pham D.V., Donis R.O., Widdowson M.-A.: Detecting spread of avian influenza a (H7N9) virus beyond China. *Emerging Infect Dis* 2015, 21, 741–749, doi: 10.3201/eid2105.141756.
 18. Ministry of Agriculture and Rural Affairs, PRC: 农业部关于印发《动物 H7N9 禽流感紧急监测方案》和《动物 H7N9 禽流感应急处置指南（试行）》的通知。(Guidelines for Emergency Handling of H7N9 Avian Influenza in Animals (Trial) (in Chinese)). Livestock and Veterinary Bureau of the Ministry of Agriculture and Rural Affairs of the People's Republic of China, 2013. http://www.moa.gov.cn/govpublic/SYJ/201304/t20130409_3428338.htm.
 19. Ministry of Agriculture and Rural Affairs, PRC: 兽医公报 (Veterinary Bulletin (in Chinese)), Ministry of Agriculture and Rural Affairs of the People's Republic of China, 2020. <http://www.moa.gov.cn/gk/sygb>.
 20. Moran P.: The interpretation of statistical maps. *Stat Soc B* 1948, 37, 243–251, doi: 10.1111/j.2517-6161.1948.tb00012.x.
 21. Pai P.F., Lin C.S.: A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* 2005, 33, 497–505, doi: 10.1016/j.omega.2004.07.024.
 22. Poovorawan Y., Pyungporn S., Prachayangprecha S., Makkoch J.: Global alert to avian influenza virus infection: from H5N1 to H7N9. *Pathog Glob Health* 2013, 107, 217–223, doi: 10.1179/2047773213Y.0000000103.
 23. Qiang X., Kou Z.: Predicting interspecies transmission of avian influenza virus based on wavelet packet decomposition. *Comput Biol Chem* 2019, 78, 455–459, doi: 10.1016/j.compbiolchem.2018.11.029.
 24. Shi J., Deng G., Ma S., Zeng X., Yin X., Li M., Zhang B., Cui P., Chen Y., Yang H., Wan X., Liu W.L., Chen P., Jiang Y., Guan Y., Liu J., Gu W., Han S., Song Y., Liang L., Qu Z., Hou Y., Wang X., Bao H., Tian G., Li Y., Jiang L., Li C., Chen H.: Rapid evolution of H7N9 highly pathogenic viruses that emerged in China in 2017. *Cell Host Microbe* 2018, 24, 558–568, doi: 10.1016/j.chom.2018.08.006.
 25. Su S., Gu M., Liu D., Cui J., Gao G.F., Zhou J., Liu X.: Epidemiology, evolution, and pathogenesis of H7N9 influenza viruses in five epidemic waves since 2013 in China. *Trends Microbiol* 2017, 25, 713–728, doi: 10.1016/j.tim.2017.06.008.
 26. Suarez D.L.: Avian influenza: our current understanding. *Anim. Health Res Rev* 2010, 11, 19–33, doi: 10.1017/S1466252310000095.
 27. Suykens J.A.K., Van Gestel T., De Brabanter J., De Moor B., Vandewalle J.: Least squares support vector machines. World Scientific Publishing, Singapore, 2002.
 28. Walsh D.P., Ma T.F., Ip H.S., Zhu J.: Artificial intelligence and avian influenza: using machine learning to enhance active surveillance for avian influenza viruses. *Transbound Emerg Dis* 2019, 66, 2537–2545, doi: 10.1111/tbed.13318.
 29. Wang X., Liu S., Mao H., Yu Z., Chen E., Chai C.: Surveillance of avian H7N9 virus in various environments of Zhejiang province, China before and after live poultry markets were closed in 2013–2014. *PLoS One* 2015, 10, e0135718, doi: 10.1371/journal.pone.0135718.
 30. Wu J., Ke C., Lau E.H.Y., Song Y., Cheng K.L., Zou L., Kang M., Song T., Peiris M., Yen H.-L.: Influenza H5/H7 virus vaccination in poultry and reduction of zoonotic infections, Guangdong province, China, 2017–2018. *Emerg Infect Dis* 2019, 25, 116–118, doi: 10.3201/eid2501.181259.
 31. Yang J., Müller N.F., Bouckaert R., Xu B., Drummond A.J.: Bayesian phylodynamics of avian influenza A virus H9N2 in Asia with time-dependent predictors of migration. *PLoS Comput Biol* 2019, 15, e1007189, doi: 10.1371/journal.pcbi.1007189.
 32. Yang L., Zhu W., Li X., Chen M., Wu J., Yu P., Qi S., Huang Y., Shi W., Dong J., Zhao X., Huang W., Li Z., Zeng X., Bo H., Chen T., Chen W., Liu J., Zhang Y., Liang Z., Shi W., Shu Y., Wang D.: Genesis and spread of newly emerged highly pathogenic H7N9 avian viruses in mainland China. *J Virol* 2017, 91, e01277-17, doi: 10.1128/JVI.01277-17.
 33. Young S.G., Carrel M., Malanson G.P., Ali M.A., Kayali G.: Predicting avian influenza co-infection with H5N1 and H9N2 in Northern Egypt. *Int J Environ Res Pub Health* 2016, 13, 886, doi: 10.3390/ijerph13090886.
 34. Zhou L., Li L., Lei L.: Avian influenza, non-tariff measures and the poultry exports of China. *Aust J Agric Resour Econ* 2019, 63, 72–94, doi: 10.1111/1467-8489.12285.