Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2016, 3(3):43-49**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Application of Combination Forecasting Model in Forecasting the Epidemic of Ebola

Qi Pengmei, Zhou Lihui

No.1 College of Science, North China University of Science and Technology, China

Abstract Africa broke out in the history of the most serious viral epidemic - Ebola in 2014. The death rate of 90% has been a great harm to people. Accurate prediction of the development of the virus epidemic is the key to control the epidemic. In this paper, the ARMA, SIR (nonlinear dynamics) and combined forecasting models were established to predict the virus spreading trend in Guinea country. The forecasting results show that the precision of the combination forecasting model is higher than that of the single forecasting model. Combined forecasting model is meaningful for the prediction of the epidemic of Ebola.

Keywords Ebora, ARMA, SIR, combined forecasting

Introduction

Ebola virus also translated as the Ebola virus is a very rare virus. The transmission routes include direct contact with the body fluids, or contact with the skin, mucous membrane and so on. The virus has a higher mortality rate than the general infectious virus, between 50% and 90%. The higher mortality and the lack of targeted drugs have exacerbated the spread of the disease.

Currently, Ebola has become a hot topic of discussion. In 2014, West Africa Ebola epidemic situation is a large-scale epidemic situation since its outbreak in West Africa in February 2014. As of December 2, 2014, World Health Organization gave the report on the Ebola epidemic said that, Guinea, Liberia, Sierra Leone, Mali, the United States and Nigeria, Senegal and Spain whose epidemic has ended cumulatively appear Ebola confirmed, probable, and possible infections 17,290 cases, where 6128 people were killed. Figure 1 is three confirmed cases of data in West Africa from August 29, 2014 to January 28, 2015. According to a simple analysis from the figure, we can get that: Guinea, Liberia, and Sierra Leone Ebola epidemic is worsening.

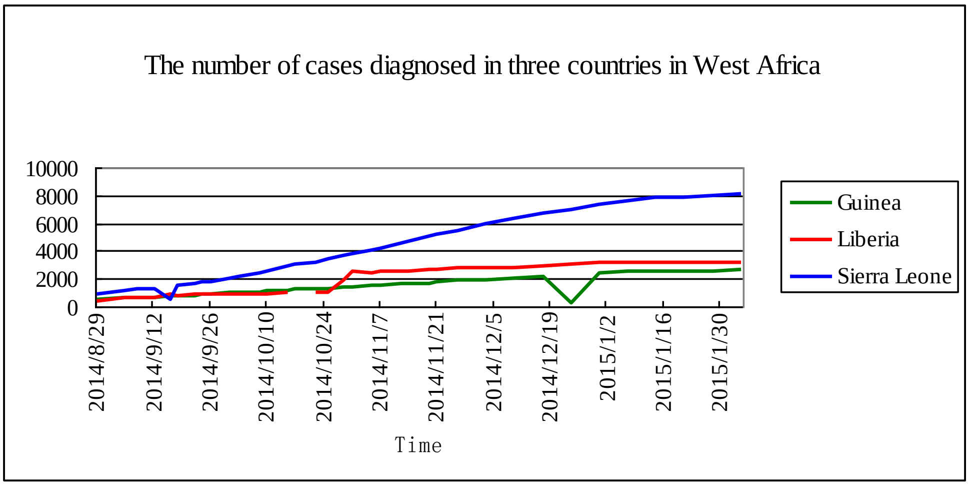

Figure 1: The Epidemic Situation of Guinea, Liberia, and Sierra Leone

Because of Ebola virus's outbreak, outbreak in the region, people stop working, expatriates choose to leave, and the slowdown in economic activity, agricultural activity stagnated. The World Bank estimates that Ebola Guinea will make economic growth fell from 4.5% to 3.5%. Economists believe that in Liberia and Sierra Leone, food prices will rise, other necessities become scarce, this makes vulnerable to the budget deficit face more severe challenges facing in the region. Thus, the negative impact caused by the Ebola virus is very large. The optimization of the treatment of the epidemic of drug transport, accurately predict and control the development, are a priority program of public research institutions. Here we should make some efforts.

Cai Haiyang *et al*., [1] through establishment of the combined model to simulate the city's incidence of hepatitis B in 2012. The conclusion is that the combination of the model reduces the prediction error of hepatitis B incidence. Yang Jiaqi and Zhang Yinjuan [2] through the establishment of BP neural network model to predict the three Ebola hemorrhagic fever outbreak in West Africa in 2015. Concluded prediction accuracy is 94.00%. BP neural network can be used to predict the Ebola epidemic. Comes [3] etc used the Gleam model to analyze the global spread of the Ebola virus in West Africa which produced an overall forecast three months under no reduction in disease outbreaks with control measures in 2014. Projections of the contents include the local spread of the Ebola virus in West Africa and the risk of international spread of ebola.

Model Hypothesis and Symbol Description

Model Assumptions

1. Assume that all of the statistical data is true and without omission phenomenon.
2. Without considering the birth rate and natural growth rate of the population in the period of infection.
3. Assume that the person who is infected with an effective contact with the patient who has not been affected by the disease will be infected.
4. Deal with the dead patients in accordance with the provisions by the strict protection of professional personnel.
5. Assume patients obtaining immunity after cured will not be secondary infection, and will not become a source of infection.

Symbol description

Symbol	Symbol description
p	order autocorrelation function
q	partial self-order correlation function
s	seasonal cycle
$S(t)$	the susceptible population accounted for the proportion of the total population in t
$I(t)$	the proportion of the total population has been infected in t
N	total population of Guinea
$E(t)$	the proportion of the total population in the latent period in t
$R(t)$	quitters accounted for the total population proportion in t
$\lambda(t)$	daily contact rate
$\alpha(t)$	day incidence in incubation period
$\beta(t)$	daily exit rate

Prediction model of Guinea Ebola epidemic

Establishment and solution of ARMA forecasting model

ARMA (p,q) model is usually expressed by the following formula:

$$x_t = r_1 x_{t-1} + r_2 x_{t-2} + \cdots + r_p x_{t-p} + a_t - Z_1 a_{t-1} \cdots - Z_q a_{t-q} + U \quad (1)$$

Among them, a_t represents a mean of 0, Random Impulse variance $\partial^2{}_a$ of identically distributed, $r_1, r_2, \cdots r_p$ and $Z_1, Z_2, \cdots, Z_p$ represent the coefficients of since regression and the moving average, U shows a model of constant. Here's x_t possible that can be from p previous randomized $x_{t-1}, x_{-2}, \cdots, x_{t-p}$ and q previous impulse $a_{t-1}, a_{t-2}, \cdots, a_{t-q}$ and constant U and the error term a_t to represent [4].

Based on the data of Guinea Ebola cases, the software [5] is used to solve the time series and the first order difference of the diagnosis.

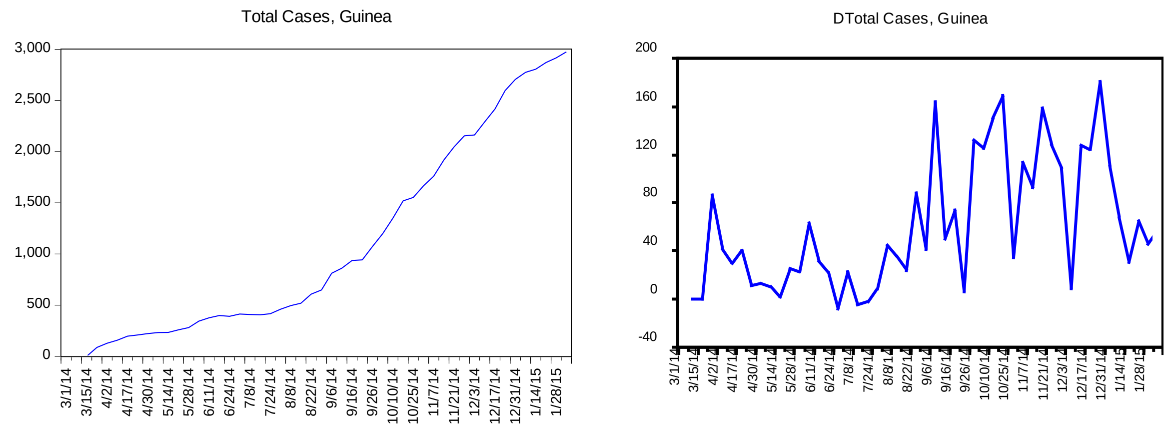

Figure 2: The Timing Diagram of Guinea's Patients (left) and Sequence Diagrams (right)

Step 1: Confirmation processing with model data

The interval from August 29, 2014 to February 4, 2015 is different, but the interval is about 7 days, so we can think of this time interval approximately equal. However, the time series is not stationary, we were taking the natural logarithm of each sequence and its first-order differential, the sequence shown in the right of Figure 2. After the conversion, the data is relatively stable [6].

Base on this sequence of first-order difference, we draw the autocorrelation function and the partial autocorrelation function table.

Table 1: The Autocorrelation Function and Partial Autocorrelation Function of Guinea

<table>
<tr><th>Autocorrelation</th><th>Partial Correlation</th><th></th><th>AC</th><th>PAC</th><th>Q-Stat</th><th>Prob</th></tr>
<tr><td>. |*******</td><td>. |*******</td><td>1</td><td>0.945</td><td>0.945</td><td>47.393</td><td>0.000</td></tr>
<tr><td>. |******|</td><td>. | . |</td><td>2</td><td>0.887</td><td>-0.056</td><td>90.027</td><td>0.000</td></tr>
<tr><td>. |******|</td><td>. | . |</td><td>3</td><td>0.826</td><td>-0.063</td><td>127.75</td><td>0.000</td></tr>
<tr><td>. |******|</td><td>. | . |</td><td>4</td><td>0.765</td><td>-0.030</td><td>160.80</td><td>0.000</td></tr>
<tr><td>. |***** |</td><td>. | . |</td><td>5</td><td>0.701</td><td>-0.060</td><td>189.17</td><td>0.000</td></tr>
<tr><td>. |***** |</td><td>. | . |</td><td>6</td><td>0.635</td><td>-0.051</td><td>213.03</td><td>0.000</td></tr>
<tr><td>. |**** |</td><td>. | . |</td><td>7</td><td>0.570</td><td>-0.035</td><td>232.69</td><td>0.000</td></tr>
<tr><td>. |**** |</td><td>. | . |</td><td>8</td><td>0.508</td><td>-0.010</td><td>248.70</td><td>0.000</td></tr>
<tr><td>. |*** |</td><td>. | . |</td><td>9</td><td>0.448</td><td>-0.027</td><td>261.43</td><td>0.000</td></tr>
<tr><td>. |*** |</td><td>. | . |</td><td>10</td><td>0.389</td><td>-0.029</td><td>271.29</td><td>0.000</td></tr>
<tr><td>. |** |</td><td>.*| . |</td><td>11</td><td>0.327</td><td>-0.075</td><td>278.43</td><td>0.000</td></tr>
<tr><td>. |** |</td><td>. | . |</td><td>12</td><td>0.266</td><td>-0.039</td><td>283.27</td><td>0.000</td></tr>
<tr><td>. |*. |</td><td>. | . |</td><td>13</td><td>0.207</td><td>-0.026</td><td>286.28</td><td>0.000</td></tr>
<tr><td>. |*. |</td><td>. | . |</td><td>14</td><td>0.153</td><td>-0.004</td><td>287.98</td><td>0.000</td></tr>
<tr><td>. |*. |</td><td>. | . |</td><td>15</td><td>0.101</td><td>-0.027</td><td>288.74</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>. | . |</td><td>16</td><td>0.052</td><td>-0.024</td><td>288.95</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>. | . |</td><td>17</td><td>0.002</td><td>-0.061</td><td>288.95</td><td>0.000</td></tr>
<tr><td>. | . |</td><td>. | . |</td><td>18</td><td>-0.043</td><td>-0.008</td><td>289.10</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>19</td><td>-0.084</td><td>-0.015</td><td>289.70</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>20</td><td>-0.122</td><td>-0.023</td><td>290.99</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>21</td><td>-0.156</td><td>-0.009</td><td>293.18</td><td>0.000</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>22</td><td>-0.191</td><td>-0.052</td><td>296.57</td><td>0.000</td></tr>
<tr><td>**| . |</td><td>. | . |</td><td>23</td><td>-0.223</td><td>-0.030</td><td>301.35</td><td>0.000</td></tr>
<tr><td>**| . |</td><td>. | . |</td><td>24</td><td>-0.254</td><td>-0.043</td><td>307.80</td><td>0.000</td></tr>
</table>

The ARMA (3, 2) model is estimated from table 1.

Step 2: Parameter Estimation

Through software, at the level of $\alpha = 0.05$ the goodness of fit of ARMA (3, 2) parameter estimation results are R^2=0.99889, $\bar{R}^2$=0.998514, F statistic=2626.910, the corresponding probability value is very small. The results show that the model is obviously, and the fitting effect is better. In the estimated ARMA (3, 2) model, AR has the 4 reciprocal complex roots. They are -0.02±0.58i,-0.65±0.57i.The modules of the 4 reciprocal complex roots are less than 1. MA has the 4 reciprocal complex roots. They are -0.14±0.99i,-0.73±0.65i. The

modules of the 4 reciprocal complex roots are less than 1. Therefore, it is considered that the ARMA (3, 2) model is stable and reversible.
Written estimation results of model ARMA (3, 2) by using the lag polynomial.

$$(1-0.679L-0.545L^2-0.447L^3)BIN_t = 1338.02+(1+0.76L+0.651L^2)\varepsilon_t \quad (2)$$

R^2=0.99889, R^{-2}=0.998514, AIC criteria=10.27183, SC criteria=10.75843.

Step 3: Generates a prediction

According to the above model, to simulate that on August 29, 2014 to February 2015, 4, then on the back within 21 days (7 days for a cycle) Ebola to predict the number of cases, get the predictive results of Guinea as shown in figure 3.

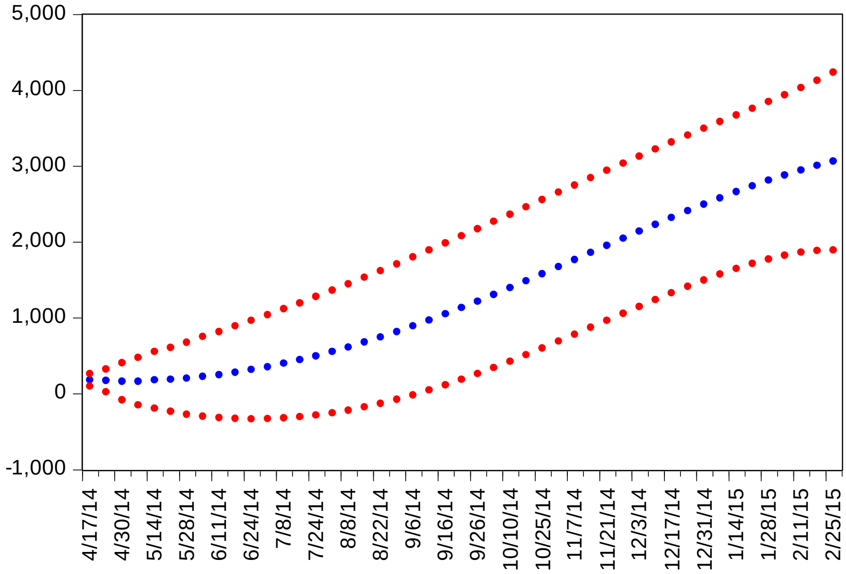

Figure3: The Number of Forecast Patients with Ebola in Guinea

Establishment and solution of SIR model

The SIR model divides the population into 3 categories: S (susceptible population), I (infectious group), and R (the group of people who are out of the crowd, including the dead and the cure). Due to the long period of the Ebola virus, in the study of its development, the population should be divided into 4 categories: S, I and R class and E class (in the incubation period) [7].

According to the related knowledge of the general epidemiology, we can see the daily rate $\alpha(t)$ and the daily exit rate $\beta(t)$ in latent period as constant α and β. The differential equation model of Ebora's law of transmission is

$$\begin{cases} \frac{dS(t)}{dt} = -\lambda(t)S(t)I(t) \\ \frac{dE(t)}{dt} = \lambda(t)S(t)I(t) - \alpha E(t) \\ \frac{dI(t)}{dt} = \alpha E(t) - \beta I(t) \\ \frac{dR(t)}{dt} = \beta I(t) \\ S(t)+E(t)+I(t)+R(t)=1 \\ S(0)=S_0, E(0)=E_0, I(0)=I_0, R(0)=R_0 \end{cases} \quad (3)$$

In order to solve the model (3), the expression of daily contact rate $\lambda(t)$ on time t should be obtained.

$A(t)$ indicates the impact of the epidemic indicators. It is affected by new deaths $d(t)$ the number of new confirmed cases $b(t)$ and new cases of suspected cases $v(t)$ per day. That is,

$$A(t) = q_1 \frac{d(t)}{\max\limits_{T\le t}\{d(t)\}} + q_2 \frac{b(t)}{\max\limits_{T\le t}\{b(t)\}} + q_3 \frac{v(t)}{\max\limits_{T\le t}\{v(t)\}} \quad (4)$$

Among which, q_1, q_2, q_3 are $d(t)$, $b(t)$, $v(t)$ the relative influence weight of $A(t)$.

The intensity preventive measures $x(t)$ （$0 \leqq x(t) \leqq 1$） and extent of people's vigilance against Ebola $y(t)$ ($0 \leqq y(t) \leqq 1$) were changed with the change of $A(t)$. $x(t)$ and $y(t)$ influence people's preventive measures. Their effects are roughly the same. So $z(t) = 0.5x(t) + 0.5y(t)$.

$\lambda(t)$ is the number of people who are active in the average daily number of patients is in a healthy population. Therefore, $\lambda(t)$ is a function of $z(t)$. And when $z(t)$=0, the maximum value of $\lambda(t)$ is λ_0. When $z(t)$ increases, $\lambda(t)$ will be reduced. When $z(t) \to 1$, and $\lambda(t) \to 0$. Through the analysis of the relationship between $\lambda(t)$ and $z(t)$, $\lambda(t)$ is a monotonically decreasing function of $z(t)$, and the specific expression of $\lambda(t)$ can be obtained by fitting the actual data.

According to Guinea's data $\alpha = 3.73$, β =0.3, q_1 =0.2041, q_2 =0.6690, q_3 =0.5859.

Using the 36 groups of data fit the total number of cases of images, such as Figure 4. From the figure can be seen that the fitting effect is better.

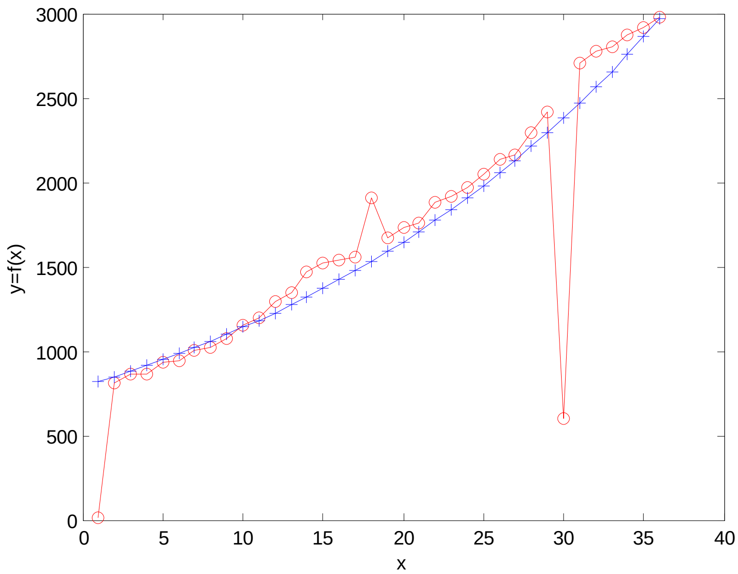

Figure 4: The Curve Fitting of the Total Number of Cases

$S(t)$、$I(t)$、$R(t)$、$E(t)$ were more difficult to resolve by model (3).So they are resolved by numerical calculation. Getting i～s curves as shown in Figure 5. So the total number of cases in the next 3 cycles is obtained.

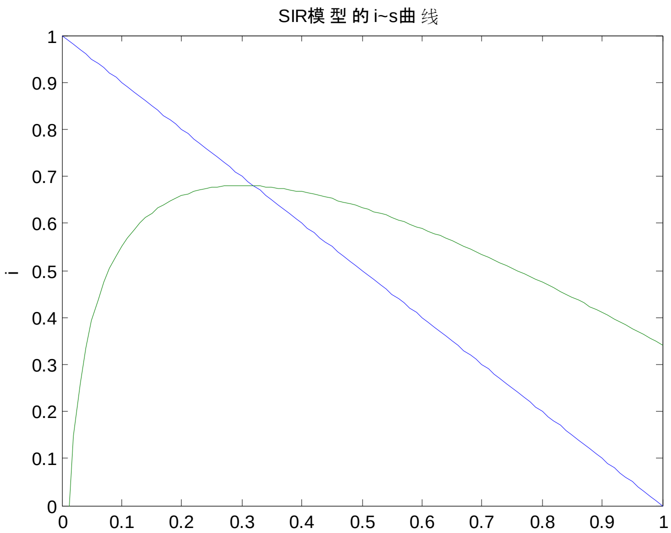

Figure 5: The Numerical i～s Curve Diagram of SIR Model

Establishment and solution of combination forecasting model

Assumptions for the same prediction problem have k (k≧2) prediction method. Prediction error square sum of the combination forecasting method is ${e_c}^2$. Then, [8]

$$ {e_c}^2 = \sum_{t=1}^{n} {e_{ct}}^2 = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(\omega_i \omega_j \left(\sum_{i=1}^{n} e_{it} e_{jt}\right)\right) = \omega^T E\omega \qquad (5) $$

Among which $\omega = (\omega_1, \omega_2, \cdots, \omega_k)^T$ is combined weight vector. $E = (c_{ij})_{k\times k}$ is prediction error information matrix. It is calculated by errors of k types of single item forecast, which $c_{ij} = \sum_{i=1}^{n} e_{it} e_{jt}$.

At last, the optimal combination weight is obtained by the combination of the minimum of the combination forecasting error.

$$ \begin{cases} \min {e_c}^2 = \omega^T E\omega \\ s.t. R_k^T \omega = 1, \\ \omega > 0 \end{cases} \qquad (6) $$

$R_k = (1, 1, \cdots 1)_{1\times k}^T$ is k vectors which all elements are 1. As long as the error of the single forecasting method is known, the most effective weight vector ω can be calculated, and then multiplied by the single forecast value. A combined forecasting result is it.

The formula (6) is transformed into a nonlinear programming problem.

$$ \min = e_1^2 + e_2^2 + e_3^2 $$

$$ s.t. \begin{cases} e_1 = 3180 - 2939\omega_1 - 3379\omega_2 \\ e_2 = 3120 - 3000\omega_1 - 3467\omega_2 \\ e_3 = 3190 - 3056\omega_1 - 3221\omega_2 \\ \omega_1 + \omega_2 = 1 \\ \omega_1 > 0, \omega_2 > 0 \end{cases} \qquad (7) $$

Here, ω_1, ω_2 are the weights of two forecasting methods. Getting the solution by software, $\omega_1 = 0.7430407, \omega_2 = 0.2569593$.

Here randomly selected 3 cycles of Guinea's cases, and the results of the three forecasting methods are in table 2.

Table 2: The Prediction Value of Guinea's Cases

	The prediction value of ARMA model	the prediction value of SIR model	the prediction value of combined forecasting model	actual total number of cases
2015/2/11	2939	3379	3052.062	3180
2015/2/18	3000	3467	3120	3120
2015/2/25	3056	3221	3098.398	3190

The relative errors of the above three methods are shown in Table 3.

Table 3: Relative Error of Three Kinds of Forecasting Methods

	2015/2/11	2015/2/18	2015/2/25
ARMA model	0.07	0.04	0.04
SIR model	0.06	0.11	0.01
Combined forecasting model	0.04	0	0.028

The error of the combined forecasting model is better than that of a single forecasting model. Combined forecasting can effectively improve the forecasting results.

Concluding remarks

Ebora, whose high fatality rate and the greater cure difficulty, is a great threat to people's lives, health and psychology. Accurate prediction of the development of the epidemic will be conducive to make a scientific scheme to prevent and control the epidemic situation. The combined forecasting method takes more factors into consideration, and makes full use of the effective information of single forecast. So it is better than other models. And the result is more close to reality. It can be applied to the prediction of infectious diseases such as Ebola.

Reference

[1]. Cai Haiyang, Wu Qinghui, Lu Jingqiao. (2012). Application of Combined Model in Forecasting of Infectious Diseases [J]. Computer Simulation, 29(4):238-242.
[2]. Yang Jiaqi, Zhang Yinjuan. (2015). Application of BP Neural Network to Predict Three Ebola Hemorrhagic Fever Outbreak in West Africa [J]. Everyone Healthy, 9(12):31-31.
[3]. Gomes M F, Piontti A, Rossi L, et al. (2014). Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak[J]. PLoS Curr Outbreaks.
[4]. Tian Zhengyi. (2001). The Theory and Method of Time Series. Higher Education Press, Beijing.
[5]. Yi Danhui. (2002). Data Analysis and Application EVIEWS. China Statistics Press, Beijing.
[6]. Tian Xin, Chen Chao, etc. (2008). Using Time Series Model to Predict Measles Outbreaks. China for Vaccines and Immunization, 14 (3).
[7]. Jiang Qiyuan. (1993). Mathematical Model. Higher Education Press, Beijing, 3rd Ed..
[8]. Cheng Linghua, Chen Huayou. (2007). Validity of Weighted Geometric Mean Combination Forecasting Model Based on the Logarithmic Gray Correlation Degree [J]. Operations and management, 16 (6): 69-73.