International Journal of u- and e- Service, Science and Technology
Vol.8, No.2 (2015), pp.101-108
http://dx.doi.org/10.14257/ijunesst.2015.8.2.10

Research on the Forest Disease Forecasting Based on Markov Process and Gray Model

Yanrong Zhang[1]

1. College of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China
zhangyanrong_5@163.com E-mail

Abstract

In order to improve the accuracy of forecasting forests diseases and the number of insect number, the paper makes conclusion by using gray model and Markova chain model. It takes Hongxing forestry bureau as a demonstration site and also forecasts the insect number according to the historical data from it. By demonstrating the disease of larch that fall early among the sites in 10 years, the result shows that the forecasts are coincided with the practical case. And the rate of coincidence can be up to 90%.

Keywords: *Markov, Gray model, Forest disease forecasting*

1. Introduction

The amount of forest disease occurs prediction mainly forecast the forest disease incidence severity, incidence and susceptibility index of the main indicators. Prediction of the Period of the forest disease occurs mainly forecast the years of some kind of forest disease occurs and a specific time that the invasion period, the Incubation period, the period of onset and the dormancy period of the forest disease to appear.

This paper takes yichun hongxing forestry bureau for demonstration site, and derives the prediction models of the emergence period and the amount of occurrence by using the Markov chain and the gray model, based on the data that Hongxing Forestry collected during the past years. Thus, the paper is able to forecast the emergence period and the amount of occurrence of the forest disease.

2. Gray Model and Markov Process

2.1 Markov Process

Markov process refers to the random process that the state of the moment t and before it is irrelevant with the moment after it, which is a random process without aftereffect. The Markov process of discrete time and discrete state is called Markov Chain.

Markov process of discrete time and discrete State referred to markov chain. The definition is as follows:

E is the discrete state space of a random sequence $\{X(n), n = 1, 2...\}$. If, for any m nonnegative integers, $\{n_1, n_2, \ldots n_n (0<= n_1< n_2<\ldots< n_n)\}$ and any natural number k and arbitrary $i_1, i_2 ... i_n$, $j \in E$ If you meet

$$P\{X(n_n+k)=j \mid X(n_1)=i_1, X(n_2)=i_2, ..., X(n_n)=i_n\} = P\{X(nn+k)=j \mid X(n_n)=i_n(1)\}$$

Random sequence $\{X(n), n = 1, 2...\}$ is markov chains.

2.2 Gray Model

Grey System Forecasting Model is to use the fewer behavioral characteristics of the original data sequence generated after transformation to generate the data sequence to establish differential equation. Due to environmental disturbances on the system, so that the original data series presents state of chaos, chaos and the number of columns is the number of columns of gray or gray process, the process to establish the model of the process is called Gray Model. Gray system model is a kind of model to reveal the process of things ‘changing and continuous development in internal system, Therefore, the model of gray system is generally used to describe the differential equations. One of the most typical is GM (1, 1) model.

Set $X^{(0)}$ is a negative sequence:

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n))$$

The first-order accumulative generation sequence of $X^{(0)}$ is $X^{(1)}$:

$$X^{(1)} = (x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(n))$$

Among them, $X^{(1)}(k) = \sum^{k} x^{(0)}(i)(k = 1, 2, ..., n)$, called $X^{(0)}(k) + aX^{(1)}(k) = b$ primitive form of GM (1, 1) model.

3. Application of Markov Process Analysis in Forest Disease Forecasting

The number of all possible states of a forest disease is denoted as $S_i = S_1, S_2, ..., S_n$, the possible transfer time is denoted as $t_i = t_1, t_2, ..., t_n$.The probability of the number of states transferring from S_j to S_k can be recorded as: P_{jk} , $(j = 1, 2, ..., n)$ $(k = 1, 2, ..., n)$, P_{jk} arranged in a matrix:

$$P = \begin{bmatrix} P_{11} & P_{12} & ... & P_{1n} \\ P_{21} & P_{22} & ... & P_{2n} \\ ... & ... & ... & ... \\ P_{n1} & P_{n2} & ... & P_{nn} \end{bmatrix}$$

A step transferring from one state to another state is called transfer step. Transition probability is essentially conditional probability. Step transition probability matrix for element is called a first-order transition matrix. Step transition probability formula is:

$$P_{jk} = \frac{m_{jk}}{m_j}$$

(1)

Wherein, m_j = the number of occurrences of the state S_j ; m_{jk} = the number of times of step transition from state S_j to state S_k ; P_{jk} = Step transition probability of step transition from state S_j to state S_k .

Similarly, there are 2, 3 order, n order of higher order transition probability and higher order transition matrix, to remember: $P_{jk}^{(1)}, P_{jk}^{(2)}, ..., P_{jk}^{(n)}$.

$$P_{jk}^{(n)} = \sum_{r=1}^{t} P_{jr}^{(m)} \bullet P_{rk}^{(n-m)}$$

(2)

According to the theory of markov process, state S_j via n order transferred to the state S_k can be seen as a first- transferring via $m(0 < m < n)$ order to the state S_r , by state S_r after $n-m$ order arrived the state S_k. Since no aftereffect, these two transfers are independent; the recursive formula calculated by using the total probability formula is as follows:

$$P_{jk}^{(n)} = \sum_{r=1}^{t} P_{jr}^{(m)} \bullet P_{rk}^{(n-m)}$$

(3)

The extent of the early fall larch disease occurred in Red Star Forestry is divided into heavy, medium and light, these three states are respectively denoted with S_1, S_2 and S_3, the occurrence of historical data is shown in Table 4-14.

Table 1. Occurrence Degree of Mycosphaerella

order	state	order	state	order	state	order	state
1	S2	6	S1	11	S2	16	S3
2	S2	7	S2	12	S2	17	S1
3	S3	8	S1	13	S3	18	S2
4	S2	9	S3	14	S1	19	S1
5	S1	10	S1	15	S3	20	S2

The first-order, second-order, third-order and fourth-order transfer matrix are Calculated as:

$$P^{(1)} = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} = \begin{bmatrix} 0.14 & 0.57 & 0.29 \\ 0.43 & 0.29 & 0.28 \\ 0.60 & 0.20 & 0.20 \end{bmatrix};$$

$$P^{(2)} = \begin{bmatrix} 0.44 & 0.31 & 0.25 \\ 0.35 & 0.38 & 0.27 \\ 0.29 & 0.44 & 0.27 \end{bmatrix};$$

$$P^{(3)} = \begin{bmatrix} 0.34 & 0.39 & 0.27 \\ 0.37 & 0.36 & 0.27 \\ 0.39 & 0.36 & 0.25 \end{bmatrix};$$

$$P^{(4)} = \begin{bmatrix} 0.37 & 0.37 & 0.26 \\ 0.37 & 0.37 & 0.26 \\ 0.37 & 0.37 & 0.26 \end{bmatrix}.$$

It is clear that the probability of $P^{(4)}$ has reached its limits. Prediction results are as follows:

Predicting the extent of the onset of the early fall larch disease for the coming year, we start the investigation from $P^{(1)}$. If the degree is heavy in this year, we'll get $P_{12}^{(1)}$=0.57 in the first line for the biggest and predict the extent of the onset of the early fall larch disease next year is medium.

Predicting the extent of the onset of the early fall larch disease for the year after, we start the investigation from $P^{(2)}$. If the degree is medium in the next year, we'll get

$P_{22}^{(2)}$ =0.38 in the second line for the biggest and predict the extent of the onset of the early fall larch disease year after is medium.

4. Application of Gray Model in Forest Disease Forecasting

4.1 Utilize the Gray Model to Simulate the Pathogenesis Process and Predict the Time Point of the Onset of the *Mycosphaerella* Laricileptolepis Lto, *et al.*

Some kind of disease onset process simulation and forecasting of onset time are not quantity dynamic simulation and prediction, but the original sequence number in the state of catastrophe values mapped to the catastrophe in the sequence, which constitute a cataclysm time-series data column, and on this basis, to simulate the process of onset and predict the GM (1, 1) model, thus the onset point in time of the disease can be forecasted in the future though the extensionality of the model.

Number of forest disease dynamics under certain conditions depends on whether factors and it often depends on a dominant meteorological factor. For this type of disease, after determining the catastrophic threshold of the leading meteorological factors, you can simulate the process leading to its catastrophic meteorological factors, and predict the catastrophic point diseases indirectly through the forecast of meteorological catastrophe point.

Now, according to the heilongjiang province yichun hongxing forestry bureau forest disease prevention quarantine station of historical data, the forestry bureau of *Mycosphaerella* laricileptolepis lto, *et al.*, pathogenesis process simulation and pathogenesis time point prediction are as follows.

The severity of *Mycosphaerella* laricileptolepis lto *et al.*, is most relevant with the rainfall of July in last year. Under normal circumstances, when the July rainfall $\leq$200mm, the *Mycosphaerella* laricileptolepis lto, *et al.*, index $\geq$40 in the coming year. Therefore, determine the threshold of catastrophe for the july rainfall as 200mm, when the rainfall of july <200mm, the catastrophe of the *Mycosphaerella* laricileptolepis lto, *et al.*, occurred. Heilongjiang Province Yichun City Forestry Bureau Hongxing 1995-2010 July rainfall as shown in Table 1.

Table 2. The July Rainfall in HongXing Forestry Bureau from 1995 to 2010

Order	Years	July precipitation ($x_{(k)}^{(0)} = 200$)
1	1995	119.90mm
2	1996	117.90 mm
3	1997	211.10 mm
4	1998	110.90 mm
5	1999	232.00 mm
6	2000	110.50 mm
7	2001	172.80 mm
8	2002	130.90 mm
9	2003	140.60 mm
10	2004	88.40 mm
11	2005	176.80 mm

Make mapping for the catastrophic threshold in accordance with $x^{(0)}_{(k)} = 200$, $\lambda : x^{(0)}_{(k)} \to x^{(0)}_{\lambda(k^{'})}$ $\lambda = 200$

$$x^{(0)}_{\lambda(k^{'})} = \{119.90,117.90,110.90,110.50,172.80,130.90,140.60,88.40,176.80\}$$
$$= \{x^{(0)}_{\lambda(1^{'})}, x^{(0)}_{\lambda(2^{'})}, x^{(0)}_{\lambda(3^{'})}, x^{(0)}_{\lambda(4^{'})}, x^{(0)}_{\lambda(5^{'})}, x^{(0)}_{\lambda(6^{'})}, x^{(0)}_{\lambda(7^{'})}, x^{(0)}_{\lambda(8)}, x^{(0)}_{\lambda(9^{'})}\}$$
$$= \{x^{(0)}_{(1)}, x^{(0)}_{(2)}, x^{(0)}_{(4)}, x^{(0)}_{(6)}, x^{(0)}_{(7)}, x^{(0)}_{(8)}, x^{(0)}_{(9)}, x^{(0)}_{(10)}, x^{(0)}_{(11)}\}$$

Respectively, make the mapping for cataclysm year order R and catastrophe year S : $P^{(0)} : \{k^{'}\} \to \{r\}$, $P^{(0)'} : \{k^{'}\} \to \{S\}$

Get catastrophe in order in the serial number $P^{(0)}$ and catastrophe year actual number sets $P^{(0)'}$ as follows:

$$P^{(0)}_{(k)} = \{1, 2, 4, 6, 7, 8, 9, 10, 11\}$$
$$P^{(0)'}_{(k)} = \{1995, 1996, 1998, 2000, 2001, 2002, 2003, 2004, 2005\}$$

And once, according to the $P^{(1)}_{(k)} = \sum_{i=1}^{k} P^{(0)}_{(i)}$, accumulating the $\{P^{(0)}\}$, $\{P^{(0)'}\}$ and generating respectively, we get:

$$P^{(1)}_{(k)} = \{1, 3, 7, 13, 20, 28, 37, 47, 58\}$$
$$P^{(1)'}_{(k)} = \{1995, 3991, 5989, 7989, 9990, 11992, 13995, 15999, 18004\}$$

By modeling the data, we simulate the catastrophe in order and catastrophe year, and get the structure of GM (1, 1) model:

$$\overline{P}^{(1)'}_{(k+1)} = 3.47e^{1.635k} - 2.47 \quad (4)$$

Decrease progressively and generate the formula:

$$\overline{P}^{(0)'}_{(k+1)} = \overline{P}^{(1)'}_{(k+1)} - \overline{P}^{(1)'}_{(k)} \quad (5)$$

Analyze and simulate the rainfall data ≤200mm in July of the years from 2006 to 2011 recorded by Hongxing Forestry, and predict the year that occurred *Mycosphaerella* laricileptolepis lto, *et al.*, as shown in Table 2, predicted results conforms well to the actual situation.

Table 3. Predicted Results

The rainfall ≤200mm	Virtual year	2006	2007	2008	2009	2010
	virtual rainball	186.70 mm	105.20 mm	119.60 mm	128.00 mm	122.80 mm
	Predicted year	2006	2007	2008	2009	2010

3.2 Utilize Gray Model to Analyze and Predict the Disease Index of *Mycosphaerella* Laricileptolepis Lto, *et al.*

Cumulate the observational data of Table 1, and get modeling data columns:

$$X^{(1)} = (X^{(1)}_{(t)} \mid t = 1, 2, 3, 4, 5, 6)$$
$$= (X^{(1)}_{(1)}, X^{(1)}_{(2)}, X^{(1)}_{(3)}, X^{(1)}_{(4)}, X^{(1)}_{(5)}, X^{(1)}_{(6)})$$

$= (3.8, 8.2, 15.7, 28.0, 43.8, 59.4)$

According to the dynamic simulation in number of *Mycosphaerella* lariciieptolepis lto *et al.*, from 2005 to 2010, the model is obtained:

$$\overline{X}^{(1)}_{(t+1)} = 22.8540e^{0.2498t} - 19.0540$$

(6)

Table 4. 2005-2010 *Mycosphaerella* Lariciileptolepis Lto, *et al.*, Observation Data

order	Year	$X^{(0)}_{(t)}$	Disease index
1	2005	$X^{(0)}_{(1)}$	3.8
2	2006	$X^{(0)}_{(2)}$	4.4
3	2007	$X^{(0)}_{(3)}$	7.5
4	2008	$X^{(0)}_{(4)}$	12.3
5	2009	$X^{(0)}_{(5)}$	15.8
6	2010	$X^{(0)}_{(6)}$	15.6

Predict the number dynamics with model (3) for 2011, we the result of the disease index as $\overline{X}^{(0)}_{(7)} = \overline{X}^{(1)}_{(7)} - \overline{X}^{(1)}_{(6)}$ =29.0291, while virtual disease index is 29.59.Obviously, the predicted and measured results are basically consistent.

Prediction results compared with actual condition of Mycosphaerella shown in Figure 1.

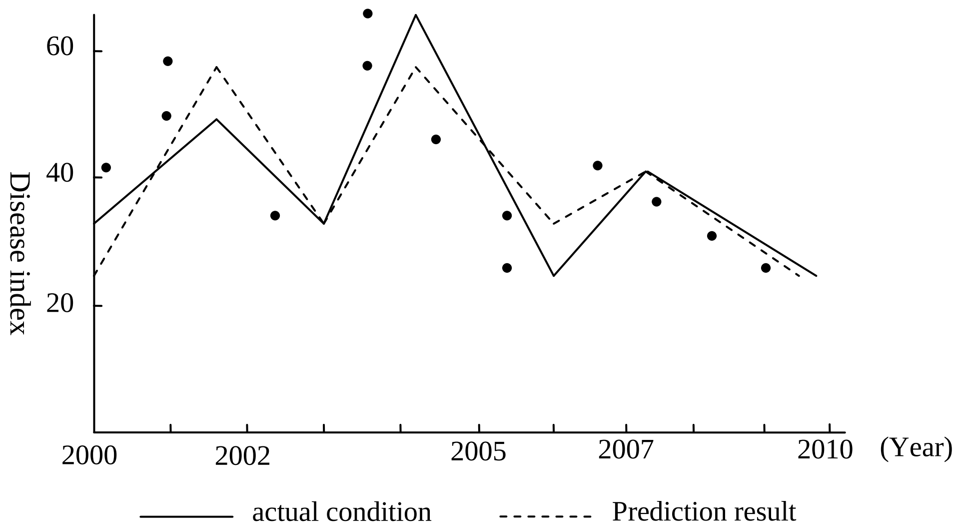


Figure 1. Prediction Results Compared with Actual Condition

5. Conclusion

This paper is aimed at hongxing forestry bureau forest disease that happens often and analyzes them in the use of gray-scale model. Mycosphaerella lariciileptolepis lto, *et al.*, for example, the author simulates the process of its cataclysm and predicts the time point of its cataclysm, observes its number state in a pathogenesis season on the same time and in the same place, by which the author simulated the time state of Mycosphaerella lariciieptolepis lto, *et al.*, and according to the dynamic number of the disease, the author simulates and predicts the disease index, through the model validation, the predicted results with measured data is basically consistent .

Acknowledgements

This work was supported by Educational Science and Technique Research Program of Heilongjiang Province (12541214), Heilongjiang postdoctoral scientific research projects (LBH-213127) and Doctoral Scientific research Project of Harbin University of Commerce (Analysis of e-commerce consumer behavior and its influencing factors).

References

[1]. Y. Zhang, "The Research on the Model and Algorithm of Forest Disease and Pest Forecast based on the Rough Set Theory, Northeast Forestry University", **(2012)**.

[2]. Z. Liu, Q. Zeng and Y. Li, "Research of equipment selection and matching expert system in fully mechanized caving face based on ontology", Key Engineering Materials, **(2010)**, pp**.** 419-420,117-120.

[3]. L. F. Zhang and J. Ai, "Method for Bridge Bearing Capacity Assessment Based on Analytic Hierarchy Process", Transactions of Nanjing University of Aeronautics &Astronautics, vol. 26, no. 3, **(2009),** pp. 236-241.

[4]. Z. Liu, Q. Zeng and Y. L, "Research of equipment selection and matching expert system in fully mechanized caving face based on ontology", Key Engineering Materials, **(2010)**.

[5]. X. Hao, D. Zhang, X. Liu and H. Zhao, "Algorithm and Application Research of Data Mining based on Rough Set Theory", Computer engineering and Applications, vol. 1, no. 9, **(2007).**

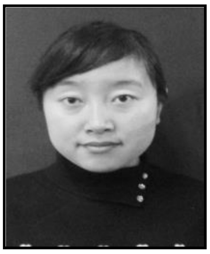
[6]. S. Wu, Q. –s. Lou, Q. Zhijie and X. Li, "Study of Setting up The Forest Resources Management Information System Based on WebGIS", Chinese Geographical Science, vol. 1, **(2003),** pp. 56-61.

[7]. W. Hsiao and J. C. Trappey, "An integrated platform of collaborative project management and silicon intellectual property management for IC design industry", Information Sciences, vol. 197, no. 15, **(2009),** pp. 2576-2590.

[8]. D. Zhang and W. S. Lee, "A Web2based Question Answering System", In: Proceedings of the SMA Annual Symposium**, (2003)** January, NUS, Singapore.

[9]. Y. Li, Cui and Duwu, "Improvement and application of MVC design patterns", Jisuanji Gongcheng, Computer Engineering, **(2006)**, pp. 62-64.

Author



Zhang Yan-rong, Doctor, Lecturer, teacher of School of the Computer and Information Engineering, Harbin University of Commerce. Her main research fields are E-commerce and Artificial Intelligence.