PREDICTIVE MODEL ON THE YEARLY FREQUENCY OF TYPHOON LANDING IN SOUTH OF CHINA[1]

Li Zuoyong(李祚泳), Deng Xinmin (邓新民) and Sang Huamin (桑华民)[2]

Chengdu Institute of Meteorology, Chengdu, 610041

Rreceived 3 March 1997, accepted 3 July 1997

ABSTRACT

The major purpose of this study was to develop a predictive model of the yearly frequency of typhoon. Based on historical data, a predictive model of yearly frequency of typhoon that landed in South China was proposed during the season of typhoon using the basic thinking and its algorithm of Projection Pursuit Regression (PPR). The results show that the predictive precision of PPR model is much better than that of stepwise regression (SR) model.

Key words: typhoon, forecast of typhoon, projection pursuit, projection pursuit regression

Ⅰ. INTRODUCTION

China is one of the countries in the world that suffers from serious effects of typhoons and casualties and economic loss are especially remarkable in the southeast coast of China every year. Large difference exists in the frequency of landing typhoon from season to season. If we can predict the landing frequency on a yearly basis, efforts can be made to reduce or prevent disasters associated with it. Generally, the frequency of the landing typhoon is non-linearly related with the information of prior predictors. If the traditional Certain Data Analysis (CDA) based on "supposition–analogue–prediction" is used in modeling, it is difficult to identify the inherent pattern among the data with satisfactory effects because of the restriction of format and mathematics. Thus, it is necessary to work on a new method of analysis and modeling.

Since the 1970's, the development of computer technique has made it more and more popular the technique of Exploratory Data Analysis (EDA), which works by "data review–analogue by computer–prediction", and the Projection Pursuit (PP) approach well represents the new idea (Friedman and Tukey, 1974). It has been successfully applied to many fields of research due to its suitability for analyzing and processing of multi-dimensional, nonlinearity and abnormal problems (Liu Daxiu, Zheng Zuguo and Xu Jiang, 1995; Cheng Qingjun, 1996). As applications in studies of catastrophe and meteorology have been rarely reported, attempts are being made here to address the basic idea of applying PP in the prediction and modeling for meteorological catastrophe and its algorithm. The application also extends to comparison of annual landing frequency as predicted by PP and SR models.

Ⅱ. BASIC THINKING AND ALGORITHM OF PP

---

[1] The work was supported by Applied Fundamental Research Fund of Sichuan Province.

[2] Sang is a '92 graduate of Chengdu Institute of Meteorology.

The technique of PP is both new and sophisticated and is used to analyze and process data of multi-dimensions, abnormality and nonlinearity, combining applied mathematics, modern statistics and computer sciences. The basic idea is to project multi-dimensional data onto certain planes at certain projection directions, and then incessantly search for, through computational techniques, significant projection planes which reveal the structures and features of multi-dimensional, non-linearity data; low-dimensional space is then useful for analysis of the structure of data with multi-dimension and non-linearity (Diaconis, Friedman, 1984). Many new analysis methods can be obtained by combining this thought with traditional statistical analysis methods, of which, the PPR model is obtained as follows (Huber, 1985).

Assuming that $y=f(x)$ and $\overline{X}=(x_1, x_2, \cdots x_k)$ are random variables of one dimension and k dimensions, respectively. In order to reveal the true features of multi-dimensional and non-linear data, PPR employs the sum of a series of ridge functions (also known as numerical functions) $G_m(z)$ to approximate the regression function

$$f(x) \sim \sum_{m=1}^{M} \beta_m G_m(z) = \sum_{m=1}^{M} \beta_m G_m(\bar{\alpha}.\bar{x})$$

$$= \sum_{m=1}^{M} \beta_m G_m \left(\sum_{m=1}^{k} \alpha_{jm} x_j \right) \tag{1}$$

where β_m, α_{jm} are coefficients; $G_m(z)$ is the mth ridge function; $z=\bar{\alpha}\cdot\bar{x}$ is an independent variable of the ridge function, which is the projection of the random variable of k dimensions in the direction of $\bar{\alpha}$ vector, a k dimensional variable itself; M is the number of ridge functions. In Eq.(1), the model error can be decreased by increasing the number of the ridge function $G_m(z)$, which is obtained by continuously smoothing and approximating data in each of the projection directions using step-linear functions. More objectively, the PPR model is then able of displaying the inner structure and feature of data and the result leads to the enhancement of model stability.

Friedman and Stuelzle (1981) suggest a multiple smoothing algorithm for PPR technique. Zheng et al. develop three kinds of applied procedure of PPR, PPAR and PPMR. In PPR with multiple smoothing, the key of realization is to employ a method of alternative iteration and optimization by the stratification and group to determine the parameters of α_{jm}, β_m, $G_m(z)$ and the optimal item M_u. The least square method is used as the minimum criterion for PPR model, i.e. selecting the suitable combination of 4 parameters mentioned above to make the following Eq. (2) a minimal.

$$L_2 = E\left[y - \sum_{m=1}^{Mu} \beta_m G_m \left(\sum_{j=1}^{k} \alpha_{jm} x_j \right) \right]^2 . \tag{2}$$

The steps are as follows:

First, divide all parameters into a number of groups and assign all but one of them different initial values for optimization. Then take the extremes in this set of parameters as the initial values with which another set of parameters are optimized. The process is repeated in looping until to a point at which a set of values is found with which the trend of decrease in Eq.(2) no longer appears. Specifically, parameters of $\alpha_{jm}(j=1,2,\cdots k)$, β_m and G_m are included in a group where $m=1,2,\cdots,M$, totaling M sets. The M-1 set is first fixed when α_{jm}, β_m and G_m are optimized for derivation. The set is then divided into 3 subsets and two of them are fixed so that the third is searched repeatedly until convergence.

During a numerical computation with the compiled PPR applied procedure, no parameters are needed to adjust operationally except for the following ones: (1)smooth coefficient $S \in (0,1)$, which determines the sensitivity of the model, the smaller the S, the more sensible the model will be, (2) the number of sample N, and (3) the maximum number M and optimal number M_u of the ridge functions which determine the accuracy with which the inner structure of data is searched for the model, and which generally meets $M_u \leqslant M \leqslant 9$. Its final selection is determined by the fitting of computational results.

III. PREDICTION OF FREQUENCY OF TYPHOON LANDING IN SOUTH OF CHINA USING PPR

The data in the case study is referred to Li and Deng (1995). The frequency of landing typhoon on a yearly basis (1954-1983) in the south of China (y) and the values of two predictors are listed in Table 1. The two predictors are as follows:

Table 1. Landing frequency of typhoon on a yearly basis (1954-1983) in south of China and the values of two predictors as well as the comparisons of two predicted results

Year	x_1	x_2	y	y_{PP}	$\lvert R_{PP} \rvert$	y_{SR}	$\lvert R_{SR} \rvert$
1954	17.5	9.9	9	9.10	1.1%	8.05	10.6%
1955	15.5	7.5	5	4.78	-4.3%	5.17	3.7%
1956	18.5	10.2	10	9.82	-1.8%	8.85	11.5%
1957	18.0	8.6	7	6.80	-2.9%	7.43	6.1%
1958	17.5	8.7	8	8.19	2.3%	7.21	9.9%
1959	15.5	8.7	5	6.53	30.6%	6.03	20.7%
1960	18.0	9.8	10	9.87	-1.3%	8.27	17.3%
1961	18.5	10.5	10	10.15	1.5%	9.06	9.4%
1962	16.0	8.8	6	6.16	2.7%	6.40	6.6%
1963	15.5	8.7	8	6.53	-18.4%	6.03	24.6%
1964	20.5	12.9	11	10.73	-2.5%	11.93	8.4%
1965	20.0	11.8	8	8.48	6.0%	10.86	35.7%
1966	15.0	6.9	4	3.48	-13.0%	4.47	11.8%
1967	17.5	8.6	8	10.97	-0.4%	7.14	10.8%
1968	14.0	7.5	5	4.67	-6.6%	4.31	13.9%
1969	14.5	6.7	2	2.24	12.0%	4.04	101.8%
1970	17.5	9.0	9	8.83	1.9%	7.42	17.6%
1971	20.0	12.7	11	11.26	2.3%	11.49	4.5%
1972	15.5	7.3	4	4.33	8.2%	5.04	26.1%
1973	18.0	12.3	10	9.87	-1.4%	10.04	0.4%
1974	17.5	10.2	9	8.93	-0.7%	8.26	8.2%
1975	17.5	11.6	9	9.22	2.4%	9.25	2.8%
1976	16.5	7.1	5	5.01	0.2%	5.49	9.8%
1977	15.0	7.3	4	3.83	-4.3%	4.75	18.8%
1978	14.5	9.8	7	7.25	3.6%	6.22	11.1%
1979*	16.0	8.1	7	6.11	-12.8%	5.90	15.7%
1980*	17.0	11.9	9	10.00	11.1%	9.17	1.9%
1981*	16.5	9.5	7	7.98	13.9%	7.18	2.6%
1982*	15.5	6.9	2	3.96	98.2%	4.76	138.2%
1983*	15.5	7.4	4	4.59	14.8%	5.12	28.0%

x_1: mean lowest temperature at Shati. Liuzhou and Changsha in January and February of the current year:
x_2: mean position of the ridge line of subtropical high in April and May:

In Table 1, R_{PP} and R_{SR} are the relative errors of the fitting and prediction frequency of typhoon per annum using PPR and stepwise regression models, respectively. In this example, a 2-predictor model of projection pursuit regression [PPR (2)] is set up due to selection of two typhoon-related predictors. The data of the former 25 years (1954-1978) are used in modeling while those of the latter 5 years (1979-1983) are used for verification. Setting the number of training samples N=25, and S, M and M_u are adjusted repeatedly. Then if the relative error $|R_{PP}| \le 20\%$ is set to be qualified, the fitting and predicting of the model have the optimums of up to 96% and 80%, respectively, as S=0.1, M=9 and M_u=6 are specified. The results of verification of the fitting and prediction using PPR model as well as SR model are shown in Table 1. When the significant level is set as 5%, the SR equation is as follows:

$$y=-9.18+0.5863x_1+0.7050x_2.$$

The results of the fitting and predicting are 80% and 60%, respectively. Obviously, the results of verification of PPR is superior to that of SR model (See Table 2).

Table 2. Parameters and verified results of prediction for different models of landing typhoon by PPR

Yr	Case freq.	N S M M_u 20 0 1 5 3	Rela-tive Error	N S M M_u 21 0 5 5 3	Rela. error	N S M M_u 22 0.5 5 3	Rela error
		Prediction freq.		Prediction freq.		Prediction freq.	
1974	9	8.63	-4.1%				
1975	9	9.63	7.0%	9 65	7.2%		
1976	5	5.58	11 6%	5.54	9 0%	5 21	4.2%
1977	4	3.42	-14 5%	3.97	-0 8%	3 78	-5.5%
1978	7	9.14	30.6%	7.75	10.7%	8.18	16.9%
1979	7	6.05	-13.6%	6.00	-14.3%	5.78	-17 4%
1980	9	10.76	19.5%	10.14	12.7%	10.52	16.9%
1981	7	7.47	6.7%	7.68	9.7%	7.94	13.4%
1982	2	3.77	88 5%	4.53	126 5%	4.20	110.0%
1983	4	4 78	19.5%	4 77	19.2%	4.64	16 0%
		Fitting rate: 90% Prediction accuracy: 80%		Fitting rate: 90% Prediction accuracy: 88%		Fitting rate: 91% Prediction accuracy: 88%	

Yr.	Case freq.	N S M M_u 23 0.5 5 3	Rela-tive error	N S M M_u 24 0.5 5 3	Rela. error	N S M M_u 25 0.1 5 3	Rela. error
		Predition freq.		Predition freq.		Predition freq.	
1977	4	4.06	1.5%				
1978	7	6.47	-7 6%	6.53	-7.0%		
1979	7	6.22	-11.1%	6.07	-13.6%	6.11	-12 7%
1980	9	9.10	1.1%	9 02	0.002%	10.00	11 1%
1981	7	8.26	18.0%	7.92	13.1%	7.98	14.0%
1982	2	3.76	88 0%	3 87	93.5%	3 96	98 0%
1983	4	4.63	15.8%	4.60	15.0%	4.59	14.8%
		Fitting rate 87% Prediction accuracy: 85%		Fitting rate: 88% Prediction accuracy: 83%		Fitting rate: 96% Prediction accuracy: 80%	

To verify the stability of the PPR model, PPR is once again used in the modeling of the former samples (1954-1973), (1954-1974), ⋯, and (1954-1977) while the samples in the latter periods (1974-1983), (1975-1983), ⋯, and (1978-1983) are retained for verification. The model parameters N, S, M, and M_u as well as the verified results are seen in Table 2. It shows a fitting rate of more than 90% for these models and a prediction accuracy of more than 80%, suggesting a loose link (stable model nature) between the sample size N for the modeling and that for the verification.

Ⅳ. CONCLUDING REMARKS

a. It is difficult to predict the frequency of landing typhoon accurately due to notable difference from year to year. As better prediction is with PPR against SR, it is found suitable for nonlinear data analysis and modeling.

b. No prior assumption is needed for the data structure in the analysis of data and modeling with PPR. Instead, the data is directly reviewed and optimized repeatedly by the computer. Good objectivity and stability are found in a case study of the model.

c. In operational use of the applied PPR procedure in real cases, only a few parameters need to be specified and adjusted, making it a tool ready to use and extensive to fit.

d. The attempts to predict the annual frequency of landing typhoons with PPR are only preliminary and much remains to address in more work that is intensive.

REFERENCES

Chen Qingjun, Zheng Zuguo, Xu Jiang, 1996. Application of projection pursuit regression technique (PPR) in analysis of cucumber photosynthitic rate in solar energy greenhouse (in Chinese). *J. Xinjiang Agricul. Univ.* 19: 28-32.

Diaconis P, Friedman D. 1984. Asymptotics of graphical projection pursuit. *Ann. statis.*, **12**: 793-815.

Friedman J H, Tukey J W. 1974. A projection pursuit algorithm for exploratory data analysis, IEEE Trans. Computers C-23: 881-889.

Friedman J H, Stuetzle W, 1981. Projection pursuit regression. *J. Amer. Statis.*, **76**: 817-823.

Huber P J, 1985. Projection pursuit (with discussion). *Ann. Statis.*, **13**: 435-525.

Liu Daxiu, Zheng Zuguo, Ge Yixion, 1995. The applied researches of projection pursuit regression in the experiment design analysis (in Chinese). *Appl. Statis.*, Manage., **14**: 47-51.

Liu Zuoyong, Deng Xinmin, 1995. A primary study on the application of artificial neural network to the forecasting of typhoon (in Chinese). *J.Natural Disasters*, **4**: 86-90.