Seismic comprehensive forecast based on modified project pursuit regression*

Anxu Wu^{1,*} Xiangdong Lin¹ Changsheng Jiang² Yongxian Zhang³
Xiaodong Zhang³ Mingxiao Li³ and Ping'an Li¹

¹ Earthquake Administration of Beijing Municipality, Beijing 100080, China

² Institute of Geophysics, China Earthquake Administration, Beijing 100081, China

³ China Earthquake Networks Center, Beijing 100045, China

Abstract In the research of projection pursuit for seismic comprehensive forecast, the algorithm of projection pursuit regression (PPR) is one of most applicable methods. But generally, the algorithm structure of the PPR is very complicated. By partial smooth regressions for many times, it has a large amount of calculation and complicated extrapolation, so it is easily trapped in partial solution. On the basis of the algorithm features of the PPR method, some solutions are given as below to aim at some shortcomings in the PPR calculation: to optimize project direction by using particle swarm optimization instead of Gauss-Newton algorithm, to simplify the optimal process with fitting ridge function by using Hermitian polynomial instead of piecewise linear regression. The overall optimal ridge function can be obtained without grouping the parameter optimization. The modeling capability and calculating accuracy of projection pursuit method are tested by means of numerical emulation technique on the basis of particle swarm optimization and Hermitian polynomial, and then applied to the seismic comprehensive forecasting models of poly-dimensional seismic time series and general disorder seismic samples. The calculation and analysis show that the projection pursuit model in this paper is characterized by simplicity, celerity and effectiveness. And this model is approved to have satisfactory effects in the real seismic comprehensive forecasting, which can be regarded as a comprehensive analysis method in seismic comprehensive forecast.

Key words: particle swarm optimization; Hermitian polynomial; projection pursuit; numerical modeling; forecasting regression model

CLC number: P315.75

Document code: A

1 Introduction

As a principal means of seismic prediction, the comprehensive forecasting methods play an important role in this field (Mei et al, 1993; Shi et al, 1997; Zhang et al, 2003; Wu et al, 2003, 2005a, 2008a, b). Projection pursuit (PP) technique is one of the most important seismic comprehensive forecasting methods (Zhu et al, 1994; Zhou et al, 1995; Zhao and Zhou, 1999). PP is a data-processing method combining the applied mathematics, modern statistics and computer technology. Instead of any human intervention, such as assumptions and threshold segmentation to the statistic data, it makes optimization by dimensionality reduction of projection

for the high-dimensional data, and makes objective research on the data structure so as to fully obtain the useful information from the data. In this way, it accomplishes the description of its structure with numerical functions. The knotty problem of the analysis on non-normal, non-linear high-dimensional data is solved by means of linear methods theoretically and practically, and the problem of mutual influence among sample factors is also solved (Cheng and Li, 1986; Li, 1998). Therefore, the projection pursuit technique and its research have been widely used in many forecasting fields, such as earthquake, meteorology, hydrology, atmospheric pollution, pathology, and economic trend since the projection pursuit regression, classification and density estimation are put forward by Friedman and Stuetzle (1981).

The principle of projection pursuit is to search for the projection direction which can reflect the structural

* Received 27 April 2008; accepted in revised form 13 March 2009; published 10 October 2009.

* Corresponding author. e-mail: wu-anxu@163.com.

characteristics of high dimensional data by projecting the high-dimensional data by lattice onto the low-dimensional ($K=1-3$ dimension) sub-space, and then to analyze the data structures or characteristics in a low-dimensional space to realize research and analysis of the high-dimensional data. And the projection pursuit regression method is to make projection of the input variable in a number of one-dimensional directions, and to fit the projection values of different directions with one-dimensional function, and then to approach the regression function with the sum of these one-dimensional functions. The expression is as below (Friedman and Stuetzle, 1981; Zhu et al, 1994):

$$Y = \bar{Y} + \sum_{m=1}^M \beta_m G_m \left(\sum_{j=1}^P \alpha_{mj}^T X \right). \quad (1)$$

In above equation, (X, Y) is a pair of random variables; $\bar{Y}$ is the average value of Y ; P is the input space dimension; M is the number of ridge functions; α_{mj} is the j th component of the m th projection direction; β_m is the weight value, which stands for the contribution amount of the m th ridge function to the output of the model; G_m is the m th smooth ridge function, which meets

$$\sum_{j=1}^P \alpha_{mj}^2 = 1, \quad E(G_m) = 0, \quad E(G_m^2) = 1.$$

In the research of seismic comprehensive forecast, although the projection pursuit regression has been widely used with remarkable results (Zhu et al, 1994; Zhou et al, 1995; Zhao and Zhou, 1999), it is not difficult to find some negative factors in this method while using equation (1). The ridge function G has no a specific function expression, and must be expressed by a rather extensive set of functions. Difference is used instead of differential coefficient in calculating derivative. The well-learned set of function must be applied to make interpolation in extrapolative forecasting by projection pursuit regression method, which is rather complex to calculate; α is generally solved in optimization by means of the Gauss-Newton method, which easily leads to the local solutions. So it is certainly difficult to obtain the optimal solution. When optimization parameters α , β , and the ridge function G (Friedman and Stuetzle, 1981) are inverted in group, two of these parameters are firstly fixed while the other one being optimized, and after that another two parameters are fixed as well while the other one being optimized. They are carried out one by one. The optimality of optimized parameters will be lost in the process of next

parameter optimization, and also in danger of local solutions. It is not easy to obtain the best overall optimal solution. The aforementioned questions will complicate the learning process of projection pursuit regression, cumber the extrapolation technique, lower the calculation efficiency, which makes it difficult to determine the effect. Aiming at these negative factors in the application, a projection pursuit regression algorithm (abbreviated as PH-PPR) can be re-built on the basis of the particle swarm optimization algorithm (abbreviated as PSO) (Kennedy and Eberhart, 1995) and the recursive Hermitian polynomial (Feng et al, 1978; Weaver, 1992) so as to simplify the process of modeling, change the model structure and learning strategy, and improve the efficiency and capacity of modeling. In order to test the modeling capability and the practical application effect of the seismic forecast by means of PH-PPR method, first of all, we carry out multi-dimensional numerical simulation of the theoretical sample, and then apply it to the experiment of the two types of seismic comprehensive forecast using multi-dimensional data, and finally explore its effectiveness and practicality in the seismic comprehensive forecast.

2 Construction of project pursuit regressive (PPR) model for seismic comprehensive forecast based on particle swarm optimization and structure of Hermitian polynomial

2.1 Solution for the parameter optimization of the model on the basis of particle swarm optimization algorithm

In the traditional learning process of projection pursuit regression, Gauss-Newton algorithm is generally used to make optimization (Friedman and Stuetzle, 1981; Zhu et al, 1994) by replacing the differential coefficient with difference in calculation. It is not easy to get the overall optimal solution. On the contrary, the particles swarm optimization algorithm is a newly-developed stochastic optimization algorithm with high efficiency, which is unnecessary to make the differential calculation any longer. It has efficient searching ideas with rather simple parameters. Thus the particle swarm optimization algorithm can be used to replace the Gauss-Newton algorithm. The basic principle of particle swarm optimization (Kennedy and Eberhart, 1995) is introduced as follows.

Particle swarm optimization algorithm was developed by Kennedy and Eberhart (1995) who were enlightened by foraging behavior of bird flock. This algorithm simulates the flying foraging behavior of bird flock, which reaches to the group optimization by their mass collaboration. The particle swarm optimization is an optimization algorithm based on the theory of swarm intelligence. All particles in this particle swarm optimization algorithm have fitness decided by the fitting function and a displacement of iteration in the search space decided by the velocity $V_i=(v_{i1}, v_{i2}, \dots, v_{id})$. Among them, the fitting function is defined by the optimization objectives. Firstly, the particle swarm optimization algorithm is initialized to a group of random particles (random solutions), in which the i th particle's location is expressed as $X_i=(x_{i1}, x_{i2}, \dots, x_{id})$, and then it produces the optimal solution through the iterative search. Particles update their own speed and location through two extremes in each iteration. One is the optimal solution and the individual extreme generated from the initiation of iteration algorithm of the particles to the current iterative search $p_i=(p_{i1}, p_{i2}, \dots, p_{id})$, and the other is the overall extreme which is the current optimal solution for the entire particle group overall $g_i=(g_1, g_2, \dots, g_d)$. Particles update the speed and location according to the following equations:

$$v_{id} = w \cdot v_{id} + c_1 \cdot r \cdot (p_{id} - x_{id}) + c_2 \cdot r \cdot (g_{id} - x_{id}), \quad (2)$$

$$x_{id} = x_{id} + v_{id}, \quad (3)$$

where v_{id} is the velocity of the particle i in the d th dimension space, and $v_{id} \in [-v_{dmax}, v_{dmax}]$, while x_{id} is the current location of particle i in the d th dimension; w is the inertia weight, which keeps the particles in inertia motion and plays a role of adjusting the overall and local algorithmic contractilities. The algorithm can be strongly capable of overall search when w is a large value, and the algorithm would be inclined to the local search when w is a small one. Generally, w is initiated to 0.9 and reduced to 0.4 with increase of the iterations so as to converge the search space to a certain region rapidly, and then to get the high-precision solution by local fine search. r is the random number that uniformly distributes between $(-1, +1)$; c_1 and c_2 , learning factors, respectively adjust the maximum steps of the directions to the best individual particles and to the best overall particles. If the step is too short, the particles may be far away from the target area, and if too long, they will suddenly fly to the target area, or over the target area. We gener-

ally set $c_1=c_2=2$. The flying speed of the particles is maximally limited as v_{dmax} . If the value is too large, the particles will be beyond the best solution, and if too small, it will be caught in local optimal solution. Assuming the d th dimension of search space is defined as $[-x_{dmax}, x_{dmax}]$, it would be $v_{dmax} = kx_{dmax}$ ($0.1 < k < 1.0$), each dimension is set up by the same method.

The first term in the right of equation (2) can be interpreted as the previous speed or inertia of particles, and the second term as the “cognitive” behavior of the particles which expresses the elaborative faculty of the particles themselves, and the third term as the “social” behavior of the particles which demonstrates information sharing and mutual cooperation among the particles. The particles would not stop running search by continuously tracking individual and overall extremes in the solution space until they reach to the specified maximum number of iterations or are less than the specified standard error.

2.2 Ridge function fitting technique in projection pursuit regression

In traditional projection pursuit regression model, the specific expression of ridge function G was not explicitly given (Friedman and Stuetzle, 1981; Zhu et al, 1994), instead of using large collections of simple functions to smoothly approximate, which would bring difficulties in choosing what kind of ridge functions structure to users in the actual construction of projection pursuit regression model and the forecast. In order to avoid using a large collections of simple functions, as well to ensure fitting capability and smooth effect of the ridge function G , a signal mixed with amplitude modulation (AM), frequency modulation (FM) and the sine wave $x(t)=[1+0.2\sin(2\pi 7.5t)] \times \cos[2\pi 30t+0.5\sin(2\pi 15t)] + \sin(2\pi 120t)$ (Figure 1) could be chosen as an experiment signal of the theoretical model, which could be superimposed by a fundamental frequency of 30 Hz, and a modulating frequency of 15 Hz, and a sine wave of 120 Hz (Wu et al, 2005b). Only Hermitian polynomial has been demonstrated to have stronger describing abilities in smoothness, details, trends, dynamics and nonlinear fitting by analyzing various types of polynomials (Dunkl and Xu, 2001) and by the actual check of AM and FM based on a simple non-linear theoretic data produced by this equation (Figure 1). Its ability of describing the complexity is continuously increasing with the degree of polynomials, and the calculation becomes quick and easy due to adopting the recursive way. Therefore, the recursive variable-order orthogonal polynomial h could

be used for one-dimensional ridge function fitting with expression as (Feng et al, 1978; Weaver, 1992):

$$h_n(x) = (n!)^{-1/2} \pi^{1/4} 2^{-(n-1)/2} H_n(x) \phi(x), \quad -\infty < x < \infty, \quad (4)$$

where $x = \alpha^T X$; $n!$ stands for the factorial of polynomial degree n ; $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$; $H_n(x)$ is the recursive Hermitian polynomial, that is, $H_0(x)=1$, $H_1(x)=2x$, ..., $H_n(x)=2[xH_{n-1}(x)-(n-1)H_{n-2}(x)]$.

In this way, the ridge function of projection pursuit regression can be expressed in specific form as the following fitting polynomial:

$$G(x) = \sum_{n=1}^N c_n h_n(x), \quad (5)$$

in which, c is the polynomial coefficient, and N is the polynomial degree.

Figure 1 A known signal and its 11 fractional orthogonal Hermitian polynomials (h).

2.3 Projection pursuit regression and the forecast process on the basis of the particle swarm optimization algorithm and Hermitian polynomial

The learning and the forecast process of PH-PPR can be separated into the following main steps.

1) To set the number of ridge functions as M which could be small initially, such as 1 and 2, the fitting number of Hermitian polynomials as N , which is equivalent to a structure-fixed neural network, and the fitting precision required in equation (1), and to randomly select M groups of the initial projection directions based on particle swarm optimization algorithm to complete the parameter initialization of the model.

2) To calculate one-dimensional projection series x of X for any group of the projection direction, to calculate N -order recursive Hermitian polynomial $h_n(x)$ by means of equation (4), and to figure out polynomial coefficient c_n in equation (5) by means of the image method of contradiction equation set (Feng et al, 1978) according to the principle of minimizing variance. Then

to determine the ridge function which is best fitted with the pairs of data (x, Y) . In the same way, to have an ergodic setting of M ridge functions, and calculate all the Hermitian polynomial fitting coefficients c_n . All the processes mentioned above are not in any particular order. The sum fitting error of M ridge functions and equation (1) would be calculated according to all the fitting results of the Hermitian polynomial.

3) The overall projection directions are optimized according to the rule of the minimum gross error of particle swarm optimization algorithm until finding all of the projection directions with the minimum gross error as the current best projection direction. The fitting sum of all ridge functions and the gross error in equation (1) are recalculated. If the given conditions of the fitting accuracy are met, the final model parameters will be output, and then the learning process of the model is over.

4) Otherwise, another ridge function (Hermitian fitting polynomial) is added on the original basis. All tasks of projection direction optimization and the polynomial

coefficients calculation are re-started in the same way of equation (1) till the increasing numbers of ridge functions can meet the criteria of the fitting precision of equation (1). The inversion parameters obtained here are the optimization parameters of the model. The model described by the parameters can be used as the final model of learning samples and the best forecast model of the extrapolation of new samples.

5) The corresponding predictive value can be calculated on the basis of the best forward model obtained by using the new data samples.

3 Projection pursuit regression model with multi-dimensional theoretical values

Assume the trfunction as follows (Wang, 1998):

$$y = (1.0 + x_1^{1/2} + x_2^{-1} + x_3^{-3/2})^2. \quad (6)$$

The PH-PPR model can be established according to 17 prior training samples and 17 rear test samples (Figure 3) run from equation (6) to test the modeling learning and forecast capability of multi-dimensional samples. The 10th order Hermitian polynomial is adopted as the ridge function for the training samples (Figure 3) generated from the learning process in section 2.3. PH-PPR modeling study is done by the particle swarm optimization to ultimately obtain the best projection pursuit regression model which is composed of three ridge functions (Figure 2). The coefficient c of the first ridge function is $[-0.55, 24.16, 5.68, 81.40, 13.65, 125.63, 13.12, 96.26, 4.64, 29.65, 1.40]$, and the coefficient

c of the second ridge function is $[1.30, 3.49, 6.09, 12.69, 11.46, 22.39, 9.41, 19.84, 2.81, 7.06, 0.15]$, and while that of the third function is $[-0.84, 8.27, -3.64, 26.97, -6.37, 40.99, -5.08, 30.31, -1.57, 9.16, 0.46]$, and the projection directions are respectively as $A_1 = [-0.854, 0.359, 0.247]$, indicating that component parameter x_1 plays the most important role in its projection direction; $A_2 = [-0.078, 0.146, 0.986]$, indicating that component parameter x_3 plays the most important role in its projection direction; while $A_3 = [-0.385, -0.894, 0.229]$, indicating that component parameter x_2 plays the most important role in its projection direction. It can be seen that all these three components take effect from these three groups of projection directions and the fitting results of ridge functions. It indicates that the homogeneous distribution of theoretical sample is in line with the actual situation. The actual prediction value of the 17 rear forecast samples can be calculated according to the three groups of obtained projection directions and the polynomial coefficient parameter set of the three ridge functions on the basis of equations (5) and (1). The learning and testing results of all the 34 samples are eventually obtained (Figure 3) as follows: the error of learning samples is 0.730 with the correlative coefficient of measured value and predictive value of 0.992; the error of test sample is 0.945 with the correlative coefficient of 0.976; the error of all samples is 0.815 with the correlative coefficient of 0.987. The accuracy rate of the inner check results can reach to 100% if calculated by the threshold error of ± 1.0 , while the accuracy rate would be 90% with the relative error of 10%.

Figure 2 Three ridge functions (G) of simulation data. The discrete dots stand for the original results about projection directions and ridge function values, while the continuous solid lines for the fitting results.

Figure 3 Measured value and prediction value series obtained from PH-PPR studying and testing with simulation data. Solid lines stand for the measured value while the dotted lines for the prediction value, y' for the predicted value, and $y'-y$ for the absolute prediction error, $(y'-y)/y$ for the relative prediction error.

The results of the learning of the theoretical samples and the inner coincident test by PH-PPR model prove that the projection pursuit regression model constructed on the basis of particle swarm optimization and Hermitian polynomial has a strong ability in fitting, high training speed and fast convergence error. There are less influences of the spatial distribution of training samples on system. The overall optimal solution can be obtained basically and the obtained regression model has a strong ability of expression, generalization and extrapolation. Therefore, this method can be used to comprehensive modeling analysis on the actual multi-dimensional seismic sample data, and to study their actually comprehensive capabilities for seismic data processing.

4 Projection pursuit regression model for earthquake forecast

4.1 Complexity of precursors affecting the level of seismic magnitude in the process of seismic gestation and the comprehensive seismic forecast

The comprehensive seismic forecast is one of the important methods in seismic forecast; however the

seismic precursors are very complex with hundreds of precursor's abnormalities having their unique characteristics. It is hard to find a relative characteristic parameter with a definite physical significance between the precursors and the future seismic magnitude levels. It is crucial and difficult as well to extract the parameters with common characteristics from the complex precursors so as to establish the criterion of comprehensive seismic forecast. A useful probe has been carried out in traditional projection pursuit regression with the remarkable effects (Zhu et al, 1994; Zhao and Zhou, 1999). In this paper, the multi-dimensional comprehensive seismic forecast model is established by means of particle swarm optimization algorithm and the projection pursuit regression model constructed by Hermitian polynomial to probe into the non-linear reflection relationship among a variety of precursory abnormalities, multi-dimensional time series of seismicity parameters and the magnitude of moderate earthquake. It will be very meaningful to make the most of seismic precursory abnormalities or to describe the characteristics of historical seismicities evolution in a certain region as the comprehensive prediction criterion for forecasting the future seismic magnitude level in a specific region and

period, especially to know the possible magnitude level that will occur after a number of abnormalities appearing or to be sure of the magnitude level of a certain region in the next year generally at the annual seismic meeting. It is of positive exploration significance that the earthquake occurring mechanism is resolved. We will make the comprehensive analysis on the two types of multi-dimensional seismic samples with PH-PPR models as follows, and probe into the effectiveness of PH-PPR model in solving the practical problems.

4.2 Model analysis on seismic comprehensive prediction of seismic precursors with multi-dimensional disordered samples

4.2.1 Screening of seismic precursory anomalies in Chinese earthquake cases

A large number of moderately strong earthquakes took place in Chinese mainland since 1970s, and we have accumulated abundant seismic data. It is greatly significant to apply most of these seismic data to the future seismic forecast. Wang et al (1999) carried out some important work in this field by artificial neural network. In this paper, we will directly use the factors of Chinese earthquake cases that sorted by Wang et al (1999) as the input of PH-PPR model. These 15 precursor abnormal indexes (Wang et al, 1999) include seismic belt, seismic gap, strain release, seismic frequency, value b , seismic window, wave velocity ratio, short-range leveling, Earth tilt, Earth electrical resistivity, water radon, water level, stress, macro-abnormality, and numbers of abnormality and so on. And we will take the real seismic magnitude as the final output of the system to establish a PH-PPR model for multi-dimensional comprehensive seismic forecast, and to forecast the seismic magnitude level which may occur in a certain area according to the precursor anomalies.

4.2.2 Establishment of PH-PPR comprehensive forecast model

We choose the top 30 earthquake cases as learning samples and the last 15 ones as inner check samples according to the sequence of Chinese earthquake samples sorted out by Wang et al (1999) just due to the disorder of seismic precursory samples. The we will learn PH-PPR modeling from the obtained seismic training samples through the particle swarm optimization by taking 10th order of the Hermitian polynomial as the ridge function on the basis of modeling method and the learning process given in the section 2.3. The final best projection pursuit regression model is composed of four ridge functions (Figure 4), the coefficients of the

four ridge functions are respectively $[-0.83, 6.59, -9.86, 15.99, -20.91, 25.82, -21.33, 17.71, -9.35, 4.42, 0.56]$, $[-4.17, -11.68, -19.00, -42.80, -38.80, -72.33, -36.92, -61.38, -13.12, -21.59, -0.55]$, $[3.48, 0.38, 16.53, -7.38, 32.94, -28.77, 31.39, -37.18, 11.98, -15.77, 0.27]$, and $[-79.67, -153.53, -325.86, -449.84, -549.43, -601.06, -427.92, -394.97, -127.33, -103.77, -10.75]$. And the projection directions are respectively $A_1=[0.094, 0.221, 0.009, 0.282, 0.191, 0.104, 0.030, 0.387, 0.474, -0.238, 0.183, 0.474, 0.007, -0.205, -0.287]$, which shows that the tilt and water level component play the primary role in its projection direction; $A_2=[0.155, 0.092, -0.284, -0.051, -0.630, -0.053, 0.084, 0.360, -0.141, -0.444, 0.138, 0.235, -0.200, 0.113, 0.065]$, $A_3=[0.229, -0.501, 0.144, -0.353, -0.059, 0.175, -0.028, -0.236, -0.173, -0.303, 0.441, 0.015, -0.377, -0.026, -0.035]$ and $A_4=[-0.359, -0.431, 0.364, 0.440, -0.021, 0.00157, 0.058, 0.170, 0.206, -0.238, 0.042, 0.014, -0.147, 0.440, -0.105]$, which shows that the value b , seismic gap and macro-component play the key role in their projection direction, respectively. From the four groups of projection directions and the corresponding fitting results of the ridge functions, we can find that four different contributing precursory factors, which demonstrate that the corresponding situation of precursory factors and the seismic occurrence, is complex. The predictive values of the last 15 seismic forecast samples can be easily calculated by means of equations (5) and (1) according to the parameters collection of four groups of the obtained projection directions and polynomial coefficients of four ridge functions. Finally the learning and test results of all 45 samples were obtained (Figure 5). From the statistics of Figure 5, we can conclude that the learning sample error is 0.030 with the correlative coefficient between the measured value and predictive value of 0.939; test sample error is 0.051 with the correlative coefficient of 0.964; the error of all samples is 0.034 with the correlative coefficient of 0.947. If calculated by per threshold level of magnitude error of ± 0.5 , the accuracy rate of the inner check results will be 100%; and the accuracy rate will be 93% with per relative error of 5%. Therefore, the model can be taken as a comprehensive prediction model for seismic precursors. From the results of learning and inner check of the seismic samples of PH-PPR model, we can see that the model has strong abilities in learning, extrapolation and generalization.

Figure 4 Four ridge functions (G) of typical seismic samples in Chinese mainland. See the caption to Figure 2 for an explanation of the dots and lines.

Figure 5 Regression model of real seismic magnitude and predicted magnitude of the typical seismic samples in Chinese mainland. In (a), the solid lines stand for the real seismic magnitude (M) series, and the dashed lines for predicted seismic magnitude (M') series. In (b), vertical axis stands for the absolute predictive errors between the predictive value (y') and measured value (y). In (c), vertical axis stands for the relative prediction error.

4.3 Projection pursuit regression model analysis on the time series of the largest magnitude comprehensive of multi-dimensional order samples

4.3.1 Multi-dimensional time series composed of the annual seismicity parameters and the largest magnitude of Chinese mainland

Mei et al (1993) and Lu et al (2002) indicated the

evolutionary on seismicity indexes plays a very important role in seismic forecast. It is particularly important to forecast the annual seismic level of maximum magnitude in a certain region in the annual conference of seismic forecast. It is well known that the factors which influence the maximum magnitude level of a certain region are quite complex, which might be controlled and

influenced by the historical seismicities, the regional stress field, active tectonics and so on. Since there is no deterministic dependence among these factors, we could choose a number of evolutionary seismicity indexes and the largest magnitude (Lu et al, 2002) to form multi-dimensional time series (Table 1). It can be seen from Table 1 that the annual seismic frequencies with magnitude $M > 5$ are listed in the column of seismic frequency. The other parameters are calculated according to the $M > 4$ earthquake data. The time window is set as one year without mutual overlap and the data of the Chinese mainland is used as that of the space window with removal of deep-source earthquakes. The earthquake catalogues of the quick reports and moderately strong earthquakes are from China Earthquake Networks Center (Gu, 1983; Department of Earthquake Hazard Prevention, State Seismological Bureau, 1995;

Song et al, 1995). In order to predict the future annual maximum magnitude level, it is assumed that the strongest seismic magnitude of the current year is affected by seismicity factors of the last year, which means the maximum seismic magnitude of the next year is predicted by using some seismicity indices of the last year to form learning and testing samples of a projection pursuit regression model. Since the time-series created by evolutionary indexes of seismicity are multi-dimensionally ordered samples, in order to avoid non-uniformity of the sample space and the human impact on learning outcomes, the objective PH-PPR comprehensive forecasting model is established to randomly generate 30 learning samples and seven inner check samples from 37 samples (Table 1) by uniform random number function. Then the learning results obtained in this way would be more objective for the

Table 1 Poly-dimensional time series of seismicity in Chinese mainland

Year	C	D	S	$E/10^{13}J$	N	Seismic creep / 10^8J	b	$M_{S-\max}$ of the previous year	$M'_{S-\max}$ of the present year	$M'_{S-\max}$	Type of samples
1971	0.359	0.527	6.377	2 580.499	50	2.876	0.780	7.8	6.7	6.70	Learning
1972	0.127	0.438	5.999	220.751	57	5.065	0.647	6.7	6.2	6.53	Testing
1973	0.052	0.365	5.652	36.056	41	6.096	0.779	6.2	7.6	7.69	Learning
1974	0.137	0.452	6.377	2 340.818	59	9.768	0.676	7.6	7.3	7.28	Learning
1975	0.195	0.455	6.224	816.926	52	12.340	0.718	7.3	7.3	7.52	Testing
1976	0.155	0.513	6.336	779.919	57	15.221	0.824	7.3	7.8	8.05	Learning
1977	0.240	0.821	7.031	6 518.324	154	25.582	0.861	7.8	6.6	6.55	Learning
1978	0.094	0.272	5.924	129.532	51	27.192	0.776	6.6	6.4	6.57	Learning
1979	0.068	0.306	5.609	34.683	31	28.085	0.856	6.4	6.8	6.85	Learning
1980	0.073	0.289	5.890	150.831	37	29.383	0.832	6.8	6.6	6.79	Testing
1981	0.081	0.217	5.746	75.975	28	30.412	0.875	6.6	6.9	6.83	Learning
1982	0.024	0.237	5.775	272.176	16	31.518	0.861	6.9	6.6	6.62	Learning
1983	0.014	0.162	5.503	72.191	23	32.327	0.648	6.6	6.8	6.83	Learning
1984	0.023	0.172	5.588	67.523	25	33.122	0.775	6.8	6.2	6.36	Learning
1985	0.000 1	0.095	4.956	3.670	15	33.393	0.936	6.2	7.4	7.43	Learning
1986	0.289	0.382	6.044	620.684	40	35.226	0.734	7.4	6.5	6.48	Learning
1987	0.132	0.344	5.727	55.692	47	36.444	0.678	6.5	6.4	6.33	Learning
1988	0.096	0.265	5.565	58.259	27	37.468	0.702	6.4	7.6	7.22	Testing
1989	0.086	0.261	6.185	1 626.300	37	40.556	0.648	7.6	6.7	6.63	Learning
1990	0.118	0.292	5.923	266.025	40	42.435	0.727	6.7	7.0	7.03	Learning
1991	0.060	0.217	5.842	248.763	24	43.752	0.798	7.0	6.5	6.22	Testing
1992	0.008	0.196	5.590	48.027	33	44.615	0.713	6.5	6.5	6.78	Learning
1993	0.053	0.250	5.556	45.760	24	45.318	0.874	6.5	6.6	6.68	Learning
1994	0.078	0.262	5.721	147.503	28	46.643	0.681	6.6	6.3	6.46	Learning
1995	0.035	0.189	5.574	42.465	29	47.601	0.728	6.3	7.3	6.93	Learning
1996	0.011	0.227	5.565	61.249	27	48.532	0.668	7.3	7.1	6.92	Learning
1997	0.070	0.374	6.011	708.478	32	50.974	0.639	7.1	7.5	7.36	Testing
1998	0.278	0.507	6.289	919.469	47	53.606	0.848	7.5	6.6	6.73	Testing
1999	0.151	0.476	5.902	127.913	41	55.270	0.775	6.6	5.6	5.59	Learning
2000	0.096	0.314	5.349	13.750	32	55.933	0.747	5.6	6.6	6.58	Learning
2001	0.101	0.374	5.757	85.399	42	57.240	0.711	6.6	8.1	7.87	Learning
2002	0.065	0.402	5.644	43.748	34	58.312	0.768	8.1	5.9	5.74	Learning
2003	0.025	0.212	5.444	16.172	22	58.892	1.033	5.9	6.8	6.80	Testing
2004	0.183	0.481	6.083	226.359	50	61.018	0.789	6.8	6.7	7.06	Learning
2005	0.116	0.380	5.896	128.069	46	62.474	0.790	6.7	6.5	6.25	Learning
2006	0.105	0.291	5.564	40.231	28	63.250	0.871	6.5	5.6	5.32	Testing
2007	0.027	0.210	5.345	10.884	19	63.821	1.013	5.6	6.4	6.43	Learning

Note: C , D , S , E and N represent concentration level of seismic space, seismic risk level, seismicity ratio, seismic energy, and seismic frequency, respectively; $M_{S-\max}$ represents the largest measured magnitude M_S , and $M'_{S-\max}$ denotes the predictive largest magnitude M_S of the present year.

extrapolation forecast of the multi-dimensional time ordered series.

4.3.2 PH-PPR comprehensive prediction model of the multi-dimensional time series for the strongest seismic magnitude in Chinese mainland

We will adopt Hermitian polynomial of the 10th order as a ridge function for the generated training samples and do PH-PPR modeling study by means of the particle swarm optimization on the basis of the specific learning process given in the section 2.3. The best projection pursuit regression model is composed of three ridge functions (Figure 6) and their coefficients are respectively $[-1\ 694.54, -2\ 206.67, -6\ 111.21, -4\ 802.36, -9\ 102.18, -4\ 335.54, -6\ 274.18, -1\ 633.04, -1\ 661.70, -162.21, -202.09]$, $[-10.45, -26.73, -61.39, -100.67, -138.36, -147.88, -137.82, -92.53, -51.36, -18.50, -1.19]$ and $[1.84, -1.53, 7.53, -1.49, 9.74, 11.12, 1.29, 22.49, -3.97, 12.19, -0.33]$. And their projection directions are respectively as follows: $A_1 = [-0.281, 0.681, 0.234, 0.328, -0.060, 0.475, -0.172, 0.191]$, which shows that value D , seismic risk level, plays a primary role in its projection direction; $A_2 = [0.462, 0.083, 0.038, -0.544, -0.023, 0.394, 0.219, 0.528]$, which shows that the measured components of the strongest magnitude of the former year, play a primary role in its projection direction; $A_3 = [0.488, 0.269, -0.297, 0.218, -0.418, -0.111, -0.562, -0.227]$, which shows that the value b plays a primary role in its projection direction. From the three groups of fitting results of projection directions and the corresponding ridge functions, we can see that

the time parameter related to the seismic intensity is the main seismicity parameter related to the seismic intensity, maximum magnitude, value b and value D . On the one hand, it demonstrates there is a non-linear mapping relationship between other seismicity factors and the occurrence of maximum earthquake, and on the other hand, it also shows that the time-series of the largest magnitude in China are mainly under the control of its own factors. All the aforementioned results play a suggestive significance to the evolution and control of the largest magnitude sequence, and are helpful for the guidance in choosing the seismicity parameters as well. The real forecast value of the seven prediction samples in Table 1 can be calculated by means of equations (5) and (1) according to the obtained three groups of projection directions and polynomial coefficient set of three ridge functions. Eventually we obtain all of the learning and testing results from 37 samples (Figure 7 and Table 1). From the statistics of Figure 7 and Table 1, we can get the error of the learning sample as 0.026 with the correlative coefficient of the measured value and predictive value of 0.957; the error of the testing sample as 0.066 with the correlative coefficient of the measured value and predictive value of 0.924; the error of all samples as 0.034 with the correlative coefficient of the measured value and predictive value of 0.947. If we calculate with per threshold level of magnitude error of ± 0.5 , the accuracy rate of the inner check results will be 100%, while if we calculate with per relative error of 5%, the accuracy rate will be 97%. Therefore, the model

Figure 6 Three ridge functions of the maximum seismic magnitude in Chinese mainland. The discrete dots stand for the original results about projection directions and ridge function (G) values, while the continuous solid lines for the fitting results.

Figure 7 Annual maximum real seismic magnitude and predicted magnitude series in Chinese mainland. In subfigure (a), the solid lines stand for the real seismic magnitude (M) series, and the dotted lines for predicted seismic magnitude (M') series. In subfigure (b), vertical axis stands for the absolute predictive errors between the predictive value (y') and measured value (y). In subfigure (c), vertical axis stands for the relative prediction error.

can be used as a comprehensive prediction model of the multi-dimensional time series for the strongest magnitude in Chinese mainland. From the results of learning and inner check, we can conclude once again that PH-PPR model has a high learning accuracy and strong ability in its extrapolation and generalization.

5 Discussions and conclusions

1) In the learning of PH-PPR model, the Hermitian fitting polynomial with fixed orders is adopted as the ridge function just similar to the neural network with the invariant structure (Wang, 1998; Wang et al, 1999; Wu et al, 2005a). In fact, we can adopt the dynamic function as the form of Hermitian fitting polynomial. But it will undoubtedly bring more complexity in programming and more calculation amounts in the learning process.

2) Theoretically, the ridge function weight of the equation (1) could be inverted alone, however, the weight in the PH-PPR model is recognized without independence according to equations (1) and (5). In order to reduce the unknown number of inversion parameters, the weight of the ridge functions can actually be melt with each other without individual calculation by multiplying the polynomial coefficient c of equation (5).

3) The parameters of the inversion solutions are requested to be completed at one time simultaneously without being grouped in PH-PPR model. It could avoid the coupling of the interiorly-circulating fitting offset,

accelerating the speed of calculation, and getting the overall optimal solution theoretically.

4) It has the advantages of the smooth interpolation and extension by using Hermitian polynomial function as the specific fitting form of the ridge function. Compared with the traditional projection pursuit regression method, its calculation process is simpler without recording the ridge functions in the parameter table, and it can also expose more non-linear information of high-dimensional seismic data.

5) The complex derivation calculation is unnecessary in optimizing the projection directions with the simple particle swarm optimization, which enhances the practicability of the PPR method. The application shows that particle swarm has a strong ability in parameter optimization, fast speed and high precision. The overall solution can be quickly obtained if the parameter is appropriately set. Therefore, the particle swarm optimization method can be used as a main optimization tool of the parameter inversion based on some seismic models.

6) The comprehensive forecast shows that the PH-PPR seismic comprehensive forecasting model, which is constructed on the basis of the particle swarm optimization algorithm and Hermitian polynomials, is reasonable for relatively simple thought, intuition, objective modelling process, high fitting precision, robust calculation results, strong generalization ability and certain self adaptability to data.

Acknowledgments The authors would like to express their thanks to Ms. Zhifang Yin for improving English-writing of this paper.

References

- Dunkl C F and Xu Y (2001). *Orthogonal Polynomials of Several Variables*. Cambridge University Press, Cambridge, 1–408.
- Cheng P and Li G Y (1986). Projection pursuit: A rising statistical method. *Chinese Journal of Applied Probability and Statistics* **11**(3): 138–143 (in Chinese with English abstract).
- Department of Earthquake Hazard Prevention, State Seismological Bureau (1995). *Catalogue of Strong Earthquakes in Chinese History*. Seismological Press, Beijing, 3–471 (in Chinese).
- Feng K, Zhang J Z, Zhang Q X, Yang Z Q and Cao W L (1978). *Numerical Method*. National Defense Industry Press, Beijing, 207–210 (in Chinese).
- Friedman J H and Stuetzle W (1981). Projection pursuit regression. *Amer Statist Assoc* **76**: 817–823.
- Gu G X (1983). *China Earthquake Catalogue*. Science Press, Beijing, 1–894 (in Chinese).
- Kennedy J and Eberhart R (1995). Particle swarm optimization. *IEEE International Conference on Neural Networks* **4**(27): 1942–1948.
- Li Z Y (1998). The theory of projection pursuit and its application development. *Discovery of Nature* (1): 48–51 (in Chinese).
- Lu Y Z, Li S L, Deng Z H, Pan H W, Che S and Li Y L (2002). *The Earthquake Analysis and Prediction System Based on GIS*. Chengdu Map Publishing House, Chengdu, 11–72 (in Chinese).
- Mei S R, Feng D Y, Zhang G M, Zhu Y Q, Gao X and Zhang Z C (1993). *Introduction to Earthquake Prediction in China*. Seismological Press, Beijing, 328–427 (in Chinese).
- Shi Z L, Wang L M, Fu Z X and Jin X S (1997). *The Method Research of China's Medium and Long-term Earthquake Prediction in Chinese Mainland*. China Ocean Press, Beijing, 90–139 (in Chinese).
- Song Z P, Mei S R and Xue Y (1995). Information development and using of earthquake historical materials in China. *Earthquake* **15**(1): 12–19 (in Chinese with English abstract).
- Wang S T (1998). *Neuro-Fuzzy System and its Application*. Beijing Aerial and Spaceflight University Press, Beijing, 179 (in Chinese).
- Wang W, Jiang C X, Zhang J, Zhou S K and Wang C M (1999). BP neural network in the earthquake prediction application. *Earthquake* **15**(2): 118–126 (in Chinese with English abstract).
- Weaver H J (1992). *Applications of Discrete and Continuous Fourier Analysis*. Krieger Pub Co., Malabar, Fla., 1–375.
- Wu A X, Wu P Z, Gu T Y, Qu L J, Zhao W Z and Lin X D (2005a). Fuzzy modular neural network and applied in estimation of earthquake risk in Guanting reservoir and its around zone. *Journal of Geodesy and Geodynamics* **25**(Special issue): 32–36 (in Chinese with English abstract).
- Wu A X, Wu P Z, Lan C X, Xu P and Lin X D (2005b). Hilbert-Huang transform and time-frequency analysis of seismic signal. *Earthquake Research in China* **21**(2): 207–215 (in Chinese with English abstract).
- Wu A X, Wu P Z, Zhao W Z, Zhang L F and Wang K Y (2003). Combinatorial statistic forecast of seismic activity trend in Guanting reservoir and it around zone. *Seismological and Geomagnetic Observation and Research* **24**(1): 14–21 (in Chinese with English abstract).
- Wu A X, Zhang X D, Zhang Y X, Lin X D, Li P A, Lu Y J and Mu H Y (2008a). Fuzzy system on forecast and precaution of seismic activity based on learning from examples (FSLE). *Earthquake* **28**(2): 101–107 (in Chinese with English abstract).
- Wu A X, Zhang Y X, Zhang X D, Li P A, Mu H Y and Lu Y J (2008b). Study on model of support vector machine about seismic premonitory comprehensive prediction. *Earthquake* **28**(3): 55–60 (in Chinese with English abstract).
- Zhang X D, Fu Z X, Zhang Y X, Niu A F, Huang F Q, Peng K Y and Lu J (2003). Studies and experiments on earthquake prediction during 1999–2002. *Acta Seismologica Sinica* **16**(5): 508–521.
- Zhao C P and Zhou S Y (1999). Study on PP regression model about seismic synthetical prediction in Xinjiang main zone. *Northwestern Seismological Journal* **21**(1): 37–43 (in Chinese with English abstract).
- Zhou S Y, Zhu L R and Deng C L (1995). Applied research of PP cluster in seismic swarm analysis. *Acta Seismologica Sinica* **17**(3): 312–321 (in Chinese with English abstract).
- Zhu L R, Zhou S Y and Deng C L (1994). The new comprehensive seismic prediction method: Projection pursuit regression. *Acta Seismologica Sinica* **16**(Suppl.): 1–9 (in Chinese with English abstract).