Available online at www.sciencedirect.com
ScienceDirect
Trans. Nonferrous Met. Soc. China 23(2013) 2194−2200
Transactions of Nonferrous Metals Society of China
www.tnmsc.cn

Modeling dam deformation using independent component regression method

Wu-jiao DAI[1], Bin LIU[1], Xiao-li DING[2], Da-wei HUANG[1]

1. School of Geosicences and Info-Physics, Central South University, Changsha 410083, China;
2. Department of Land Surveying and Geo-informatics, Hong Kong Polytechnic University, Hung Hom 999077, Kowloon, Hong Kong, China

Received 10 November 2012; accepted 30 May 2013

Abstract: In the application of regression analysis method to model dam deformation, the ill-condition problem occurred in coefficient matrix always prevents an accurate modeling mainly due to the multicollinearity of the variables. Independent component regression (ICR) was proposed to model the dam deformation and identify the physical origins of the deformation. Simulation experiment shows that ICR can successfully resolve the problem of ill-condition and produce a reliable deformation model. After that, the method is applied to model the deformation of the Wuqiangxi Dam in Hunan province, China. The result shows that ICR can not only accurately model the deformation of the dam, but also help to identify the physical factors that affect the deformation through the extracted independent components.

Key words: dam deformation analysis; independent component regression; principal component regression; ill-condition problem; interpreting of dam deformation

1 Introduction

Large dams are generally monitored for both boundary conditions control and structural response. Analysis of the monitoring data plays an important role in the assessment of the safety of a dam [1−4]. Two methods, statistical analysis and structural identification, are mainly applied in the analysis of dam deformation [5]. The statistical model has the advantages of simplicity for formulation, fast execution and suitability to any kind of correlation between the governing and dependent parameters. Regression analysis is a widely used method in the statistical modeling of dam deformation. However, there is a possible problem of ill-condition in the coefficient matrix, the multicollinearity of the variables used in regression may result in an inaccurate or even wrong model [6]. Stepwise regression is a conventional method to build the dam deformation model, but some environmental factors of deformation may be neglected. Principal components regression (PCR) was used to model the monitoring data of a dam by LI et al [7], and partial least-squares regression (PLSR) was applied to modeling the deformation of an earth dam by DENG et al [8]. Both PCR and PLSR, based on the theory of multivariate statistical projection, are proposed to solve the problem due to the multicollinearity of the variables. The main drawback of those statistical models above is that they make little contribution to the physical interpreter of dam deformation. So, independent component regression (ICR) model is proposed to solve these problems.

Independent component analysis (ICA) is a method of blind source separation. It transforms the observed mixed signals into a series of signals, the components of which are mutually independent in statistical sense. POPESCU [9] used the second order blind identification (SOBI) algorithm, a method of blind source separation (BSS), to find out the contributions of external loads to the displacements of the dam without a priori knowledge on the generator phenomena or the propagation environment. ICR is a method combining independent component analysis and linear regression. It has been effectively applied into some fields such as chemistry, medicine and structural engineering [10−12].

In this work, ICR was applied firstly in the field of deformation analysis to building a horizontal displacement model of the dam, and the independent

Foundation item: Project (41074004) supported by the National Natural Science Foundation of China; Project (2013CB733303) supported by the National Basic Research Program of China
Corresponding author: Wu-jiao DAI; Tel: +86-731-88879515; Fax: +86-731-88830573; E-mail: wjdai@csu.edu.cn
DOI: 10.1016/S1003-6326(13)62717-X

components in regression model were used to identify the physical factors contributed to the displacement of dam.

2 Independent component regression

2.1 Independent component analysis

ICA is a useful method for blind source separation. Its basic logistics is shown in Fig. 1. It is supposed that there are M observations $\boldsymbol{X}$, $\boldsymbol{X}(t)=[\boldsymbol{X}_1(t), \cdots, \boldsymbol{X}_M(t)]^{\mathrm{T}}$, from N independent components $S_i(t)$, $i=1, 2, \cdots, N$.

$$\boldsymbol{X}(t)=\boldsymbol{AS}(t);\ M\geq N \tag{1}$$

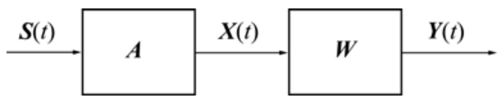

Fig. 1 Basic logistics of ICA

Unknowing any other priori information about matrix $\boldsymbol{A}$ or source signals, ICA aims to obtain a separating matrix $\boldsymbol{W}$ to separate the original signals $\boldsymbol{S}(t)$ in Eq. (1) based on some optimization criteria and learning methods. Generally, the process of calculating $\boldsymbol{W}$ can be divided into two steps: 1) Whiten the observed signals $\boldsymbol{X}(t)$ by a whitening matrix $\boldsymbol{B}$, to let $\boldsymbol{Z}=\boldsymbol{BX}$ and $\boldsymbol{E}(\boldsymbol{ZZ}^{\mathrm{T}})=\boldsymbol{I}$ ($\boldsymbol{I}$ is a unit matrix); 2) Calculate the rotation matrix by the specific independence optimize rule, let $\boldsymbol{Y}(t)=\boldsymbol{UZ}$, where $\boldsymbol{Y}(t)$ is the best approximation vector of $\boldsymbol{S}(t)$.

2.2 FastICA algorithms

ICA algorithms can be divided into two main categories, and both of them are based on the nongaussianity and independence of the source signals. FastICA [13,14] is a fast optimization iterative algorithm with a good stability. It is based on the negentropy that is a common quantitative measure of the nongaussianity of a random variable. The stronger the nongaussianity of a random variable is, the greater the negentropy is. The detailed steps are as below:

1) Center and whiten the observed data;

2) Choose an initial weight vector of unit norm (random) $\boldsymbol{w}$;

3) Update $\boldsymbol{w}$ through

$$\boldsymbol{w}(k+1)=E[\boldsymbol{x}g(\boldsymbol{w}^{\mathrm{T}}(k)\boldsymbol{x})]-E[g'(\boldsymbol{w}^{\mathrm{T}}(k)\boldsymbol{x})]\boldsymbol{w};$$

4) Normalizate $\boldsymbol{w}$ by

$$\boldsymbol{w}(k+1)=\frac{\boldsymbol{w}(k+1)}{\|\boldsymbol{w}(k+1)\|};$$

5) Go back to step 3) if not converged.

2.3 Independent component regression

Independent component regression (ICR) is a method that extracts mutually independent components from arguments first and then builds the regression model with the independent components instead of the observed arguments [15,16].

The regression model adopted is

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\cdots+\beta_mX_{im}+\varepsilon_i\ (i=1, 2, 3, \cdots, n) \tag{2}$$

where Y is the dependent variable; X is the observed arguments and β is the regression coefficient and $\beta=[\beta_0, \beta_1, \cdots, \beta_m]^{\mathrm{T}}$. The process of ICR modeling is as follows.

1) Extract the independent components $\boldsymbol{C}$ of the observed arguments $\boldsymbol{X}$ by FastICA algorithms through

$$\begin{cases}[\boldsymbol{A},\boldsymbol{W}]=\text{FastICA}(\boldsymbol{X})\\ \underset{n\times l}{\boldsymbol{C}}=\underset{n\times m}{\boldsymbol{X}}\boldsymbol{W}^{\mathrm{T}}\end{cases} \tag{3}$$

2) Build the regression model with $\boldsymbol{C}$ and $\boldsymbol{Y}$, and calculate the coefficients with least squares method by

$$\begin{cases}\boldsymbol{Y}=d_0+d_1\boldsymbol{C}_1+d_2\boldsymbol{C}_2+\cdots+d_l\boldsymbol{C}_l\\ \boldsymbol{D}=[d_0,d_1,\cdots,d_l]^{\mathrm{T}}\end{cases} \tag{4}$$

3) Calculate the original coefficient $\boldsymbol{\beta}$ with the result $\boldsymbol{D}$ in Eq.(4) and the separating matrix $\boldsymbol{W}$ in Eq.(3).

The advantage of ICR is that the components are mutually independent, so the problem of ill-condition in coefficient matrix is avoided. At the same time, the independent components extracted by ICR can be used to identify the implicit factors relative to the dependent variable $\boldsymbol{Y}$.

3 Simulation experiment

3.1 Data simulation

To demonstrate the performance of ICR, a simulation test is conducted to compare it with PCR. The regression model in the test is assumed as

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\cdots+\beta_nX_{in}+\varepsilon_i\ (n=6; i=1, 2, 3, \cdots, 400) \tag{5}$$

In the simulation test, two independent components S_1 and S_2 are constructed as the arguments $\boldsymbol{X}$. S_1 is a sinusoidal signal and S_2 is a random signal. They are used as source signals that have physical meaning hidden in the observed data. The waveforms of these signals are shown in Fig. 2. The 6 arguments, which are linear transformations of the 2 signals, have been generated and shown in Fig. 3. The dependent variable $\boldsymbol{Y}$ is given by

$$\boldsymbol{Y}=\boldsymbol{X}_1+\boldsymbol{X}_2+\boldsymbol{X}_3+\boldsymbol{X}_4+\boldsymbol{X}_5+\boldsymbol{X}_6 \tag{6}$$

where $\boldsymbol{X}_i$ is standardized. As it is shown, the dependent variable is actually affected by the 2 independent original signals. A small amount of noise is added into $\boldsymbol{X}$ and dependent variable $\boldsymbol{Y}$ in the regression model.

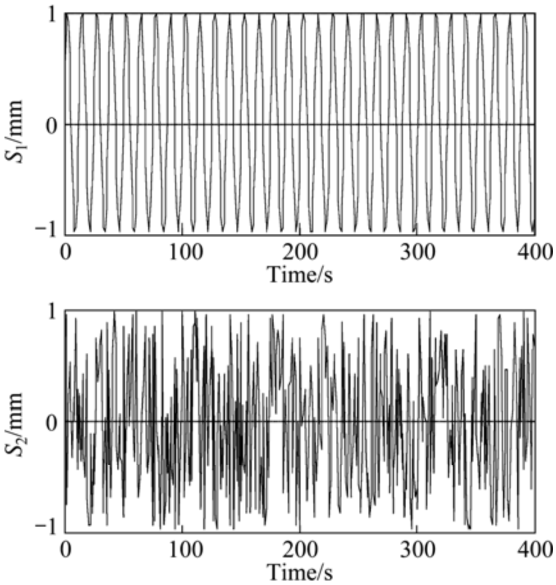

Fig. 2 Original analog signals

3.2 Regression results and analysis

The methods of least squares regression (LSR), PCR and ICR are separately used to solve the regression model and the results are listed in Tables 1 and 2. Table 1 indicates that the LSR has a great discrepancy with the actual situation and gets wrong coefficients due to the problem of ill-condition in the coefficient matrix. PCR and ICR solve the problem of multicollinearity of the variable and the derived coefficients are more realistic.

The principal components (P_s, s=1, 2) and independent components (I_s, s=1, 2) extracted by PCR and ICR are respectively shown in Figs. 4 and 5. It is clear that the numbers of principal and independent components are both two, which are consistent with the simulation settings. The first two principal components include almost 100% information about the observables. In the ICA processing, PCA is done as a pretreatment for dimensionality reduction. So P_s and I_s contain the same information when the dimensions are the same, and I_s can be obtained by a linear transformation from P_s. So, the same result is obtained from them.

Figure 5 illustrates that the ICR separates the hidden components which affects $\boldsymbol{Y}$ actually, but with a large uncertainty. On the other hand, Fig. 4 illustrates that the two components extracted by PCR have no obvious practical significance.

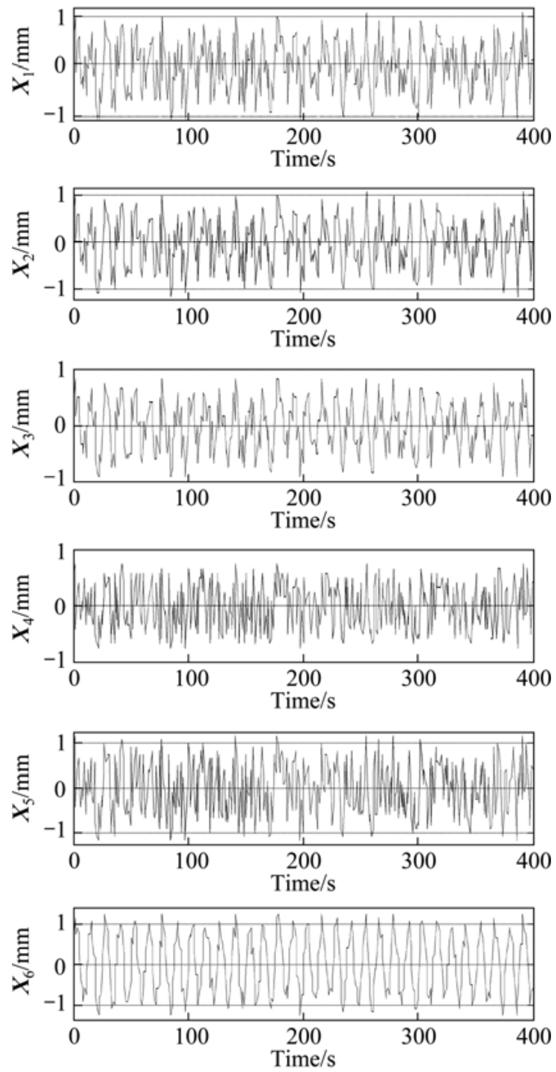

Fig. 3 Mixed signals in regression model

The correlation coefficient between the components extracted by PCR and ICR and $\boldsymbol{Y}$ are listed in Table 3. According to the principle of PCA, the two principal components extracted by PCR represent the information of observed arguments as more as possible. But not every component can have a good explanation about the dependent variable. For example, the second principal component P_2 in the simulation experiment, the correlation coefficient with $\boldsymbol{Y}$ is just 0.0396, which cannot be used to explain the dependent variable. Actually, the principal components are generally short of the explanatory capability to the dependent variable.

Table 1 Estimated model coefficients and root-mean-square (RMS) values of regression for three methods

Coefficient	β_0	β_1	β_2	β_3	β_4	β_5	β_6	RMS (10^{-3})
Genuine value	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	−
LSR	1.0010	−14.6985	62.3705	−149.4208	−29.8778	48.5626	42.0112	0.5880
PCR	1.0010	1.0467	1.0137	0.8052	0.7750	1.1942	1.0353	0.5904
ICR	1.0010	1.0467	1.0137	0.8052	0.7750	1.1942	1.0353	0.5904

Table 2 Principal and independent components extracted by PCR and ICR

Coefficient	β_0	P_1	P_2	I_1	I_2
PCR	1.0050	2.4113	0.2360		
ICR	1.0050			1.9237	2.2189

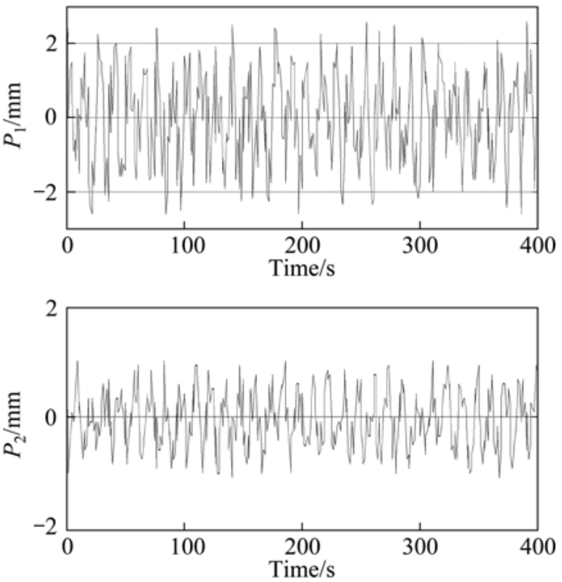

Fig. 4 P_s extracted by PCR

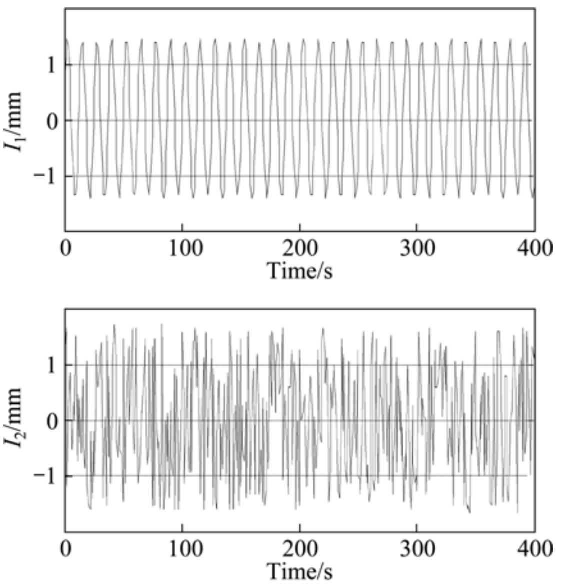

Fig. 5 I_s extracted by ICR

Table 3 Correlation coefficients between P_s, I_s and $\boldsymbol{Y}$

Correlation	P_1	P_2	I_1	I_2
$\boldsymbol{Y}$	0.9992	0.0396	−0.6551	0.7556

4 Real dataset experiment

4.1 Statistical model of Wuqiangxi Dam

The Wuqiangxi Dam, built in 1994, is located in the main stream of Yuanshui River in Hunan province, China. The river is about 73 km going through the city of Yuanling. The dam is equipped with the automated monitoring system of wire alignment, inverted plumb, hydrostatic leveling, seepage monitoring, uplift pressure monitoring, and water level measuring, and so on. The data of horizontal displacement in the dam center, daily temperature and water level of upstream and downstream are used to model the dam deformation in this study.

The horizontal displacement of a dam is mainly affected by the factors such as hydrostatic load, environment temperatures and time effect. So, the statistical model of dam deformation usually consists of the components of water level, temperature and time. According to the observed data, the statistical model is chosen as follows:

$$\boldsymbol{Y} = a_0 + \sum_{i=1}^{4} a_i H^i + \sum_{i=1}^{4} b_i T_i + c_1\theta + c_2 \ln\theta \tag{7}$$

where $\sum_{i=1}^{4} a_i H^i$ is the component of water level, where H denotes the water head of upstream and downstream; $\sum_{i=1}^{4} b_i T_i$ is the component of temperature, where T_i means the average temperature of 0−1, 2−7, 8−30 and 31−60 d because of the lag effect between the temperature of dam and the environment; $c_1\theta + c_2\ln\theta$ is the component of time, where θ is calculated by the observation date minus the base date then divided by 100. The data used in modeling have been normalized to reduce its impact to numerical calculation.

4.2 Result of modeling and related analysis

According to the procedure of ICR, an ICA about the data including components of water level, temperature and time is made in Eq. (7). Three independent components which include about 94.87% information about the observed data are chosen to build the regression model. At the same time, PCR and LSR are used to build the regression to make a comparison. The results of the three methods are listed in Table 4. It indicates that the coefficients got by LSR are unreasonable obviously because of the multicollinearity among the 1 to 4 items of water head. Simultaneously, the components of water level are not chosen in the regression model because their coefficients cannot be approved by significance testing. Thus it results in the loss of some information and affects the accuracy of the model. The results of the models using PCR and ICR are the same because the ICA in this examination is built on the basis of PCA, and the three independent components are obtained by a linear transformation from the top three principal components. According to the principle of PCR and ICR, P_s and I_s are used to build the regression model

Table 4 Regression coefficients of three methods

Method	a_0	a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4	c_1	c_2
LSR	6.5424	−48.4278	129.5761	−114.1359	33.6932	−0.5537	−2.0281	−0.8362	−1.6754	0.2838	−0.0234
PCR	6.5424	0.1467	0.1465	0.1463	0.1462	−1.2460	−1.2857	−1.3124	−1.2448	0.0530	0.1715
ICR	6.5424	0.1467	0.1465	0.1463	0.1462	−1.2460	−1.2857	−1.3124	−1.2448	0.0530	0.1715

first and the original coefficients are calculated by the separating matrix. Table 5 indicates that although the final results of PCR and ICR are the same, the coefficients of P_s or I_s are different, namely that the P_s and I_s represent different physical factors.

Table 5 Coefficients of P_s and I_s

Method	a_0	P_1/I_1	P_2/I_2	P_3/I_3
PCR	6.5424	1.6446	1.8701	−0.6276
ICR	6.5424	−4.8694	−0.6342	−0.6731

In the model of ICR, the sample determination coefficient R^2=0.9738. It means the fitting is in good condition. The significance tests about the regression equation and regression coefficients are made. The F-criterion value of regression equation is F=7760.0 and t-criterion values of three independent components are 274.1738, 35.6962 and 37.8743, respectively. All the test have been approved by significance testing. The fitting of ICR is shown in Fig. 6, and the RMS is 0.8132 mm. Then ICR model is used to make predictions about the horizontal displacement in the next few months of the dam, as shown in Fig. 7. The RMS is 0.6062 mm, manifesting them as a very stable model and a very good prediction.

5 Deformation interpretation based on ICR model

The results above demonstrate that the method of PCR and ICR can both solve the problem of ill-condition

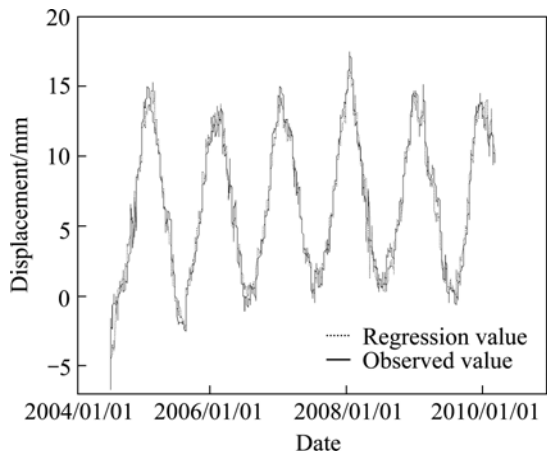

Fig. 6 Displacement fitting diagram

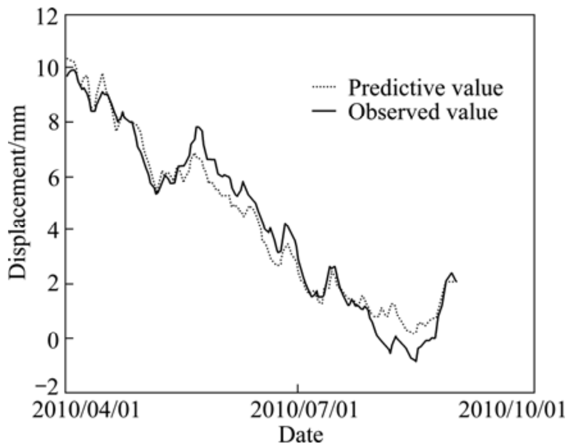

Fig. 7 Displacement forecasting diagram

and get the same regression model. A further analysis about P_s and I_s extracted in the regression model is made to illustrate the differences between the two methods. The three components extracted by PCR and ICR are shown in Figs. 8 and 9, respectively. The statistical models of the dam show that the main factors affecting the displacement of the dam are hydrostatic load, environment temperatures and time. So, the displacement components of water level, temperatures and time in the model are respectively calculated, and the results are shown in Fig. 10.

Comparing the curves in Figs. 8, 9 and 10, it is seen that the three dependent components extracted by ICR are basically the same with the curves in Fig. 10, but the principal components (Fig. 8) have no obvious similarity with the displacement components of each factor (Fig. 10). The correlation analysis between P_s, I_s and the displacement components of each factor are conducted and listed in Tables 6 and 7. The results show that the correlations between I_1 and the displacement component of temperature, I_2 and the displacement component of water level, and I_3 and the displacement component of time are −0.9946, −0.9790 and −0.9823, respectively. As a result, I_1, I_2 and I_3 can reflect the varying pattern of the factor of temperature, water level and time, respectively. Thus, it is concluded that the independent components represent the displacement components of obvious physical origin, while the principal components only show a bit connection to each factor, preventing from being interpreted as clear physical excitation.

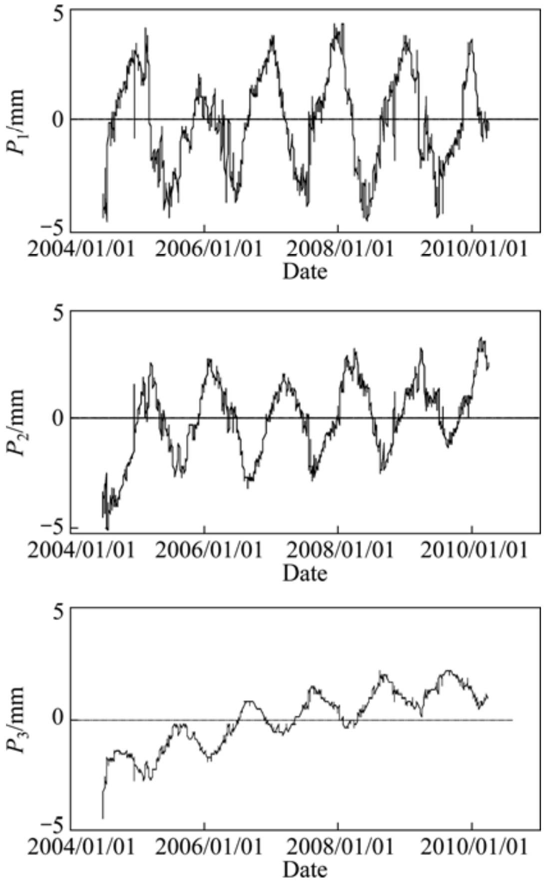

Fig. 8 Three components extracted by PCR

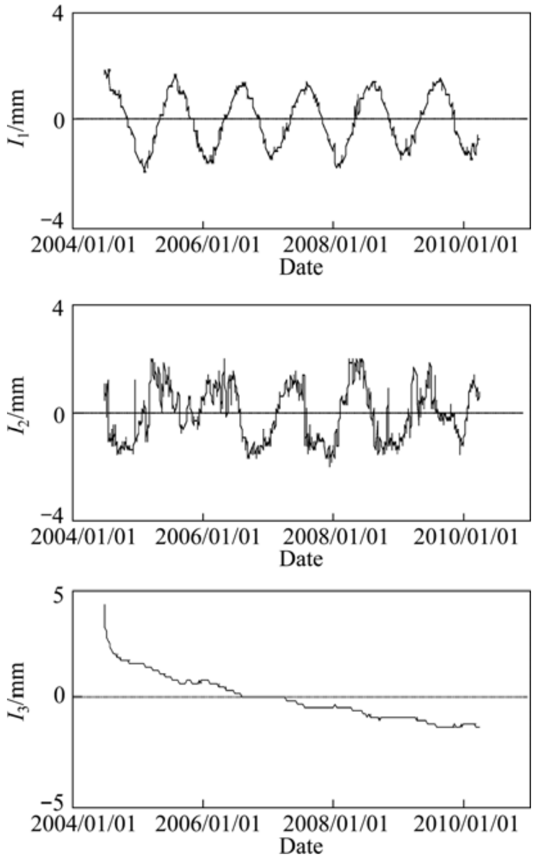

Fig. 9 Three components extracted by ICR

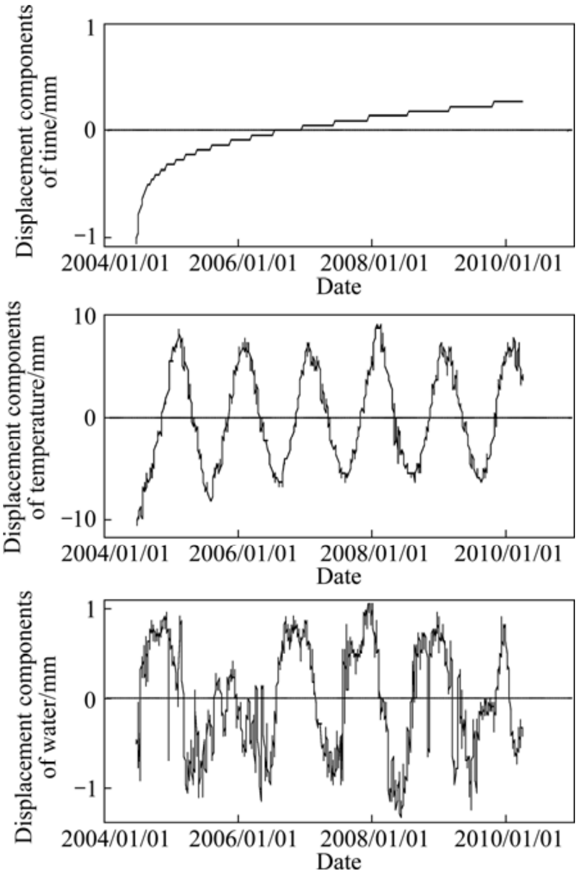

Fig. 10 Displacement components of three factors

Table 6 Correlation coefficients between P_s and impact factors

Correlation	Water level	Temperature	Time
P_1	0.8830	0.6356	0.0569
P_2	−0.4540	0.7392	0.4995
P_3	0.1089	−0.2227	0.8540

Table 7 Correlation coefficients between I_s and impact factors

Correlation	Water level	Temperature	Time
I_1	−0.1921	−0.9946	−0.1163
I_2	−0.9790	−0.0157	0.0591
I_3	0.0488	−0.1022	−0.9823

6 Conclusions

1) ICR can compress the dimension of observed arguments and make a stable regression model, avoiding the problem of ill-condition in coefficient matrix due to the multicollinearity of the variables.

2) Compared with the principal components derived from the PCR, the independent components extracted by ICR can better represent the displacement components of obvious physical excitation.

3) ICR outperforms the PCR and similar methods, and has a great potential to analyze and characterize the deformation of a dam or similar deformation body.

References

[1] ARDITO R, MAIER G, MASSALONGO G. Diagnostic analysis of concrete dams based on seasonal hydrostatic loading [J]. Engineering Structures, 2008, 30(11): 3176−3185.

[2] SZOSTAK-CHRZANOWSKI A, CHRZANOWSKI A, MASSIÉRA M. Use of deformation monitoring results in solving geomechanical problems-case studies [J]. Engineering Geology, 2005, 79(1−2): 3−12.

[3] XI Guang-yong, YUE Jian-ping, ZHOU Bao-xing, TANG Pu. Application of an artificial immune algorithm on a statistical model of dam displacement [J]. Computers and Mathematics with Applications, 2011, 62(10): 3980−3986.

[4] XIAO Jie, ZHANG Jin. Analysis and application of automatic deformation monitoring data for buildings and structures of mining area [J]. Transactions of Nonferrous Metals Society of China, 2011, 21(3): 516−522.

[5] SORTIS A D, PAOLIANI P. Statistical analysis and structural identification in concrete dam monitoring [J]. Engineering Structures, 2007, 29(1): 110−120.

[6] WU Zi-an. The independent variables-snooping in the regression analysis of dam deformation [J]. Journal of Wuhan Technical University of Surveying and Mapping, 1993, 18(1): 20−26. (in Chinese)

[7] LI Hong-xiang, YUE Dong-jie, LI Li-ren. Dam displacement model based on principal component regression [J]. Hydropower Automation and Dam Monitoring, 2008, 32(5): 61−64. (in Chinese)

[8] DENG Nian-wu, WANG Jian-guo, SZOSTAK-CHRZANOWSKI A. Dam deformation analysis using the partial least squares method [C]//Proceedings of 13th FIG Symposium on Deformation Measurement and Analysis & 4th IAG Symposium on Geodesy for Geotechnical and Structural Engineering. Lisbon: LNEC, 2008: 1−10.

[9] POPESCU T D. A new approach for dam monitoring and surveillance using blind source separation [J]. International Journal of Innovative Computing, Information and Control, 2011, 7(6): 3811−3824.

[10] ZHANG Ying-wei, ZHANG Yang. Complex process monitoring using modified partial least squares method of independent component regression [J]. Chemometrics and Intelligent Laboratory Systems, 2009, 98(2): 143−148.

[11] WESTAD F. Independent component analysis and regression applied on sensory data [J]. Journal of Chemometrics, 2005, 19(3): 171−179.

[12] SHIRALI S, HAZRA B, NARASIMHAN S, PANDEY M D. Independent component regression for predicting the responses of biaxial base-isolated buildings [J]. Earthquake Engineering and Structural Dynamics, 2010, 39(5): 583−590.

[13] HYVÄRINEN A. Fast and robust fixed-point algorithms for independent component analysis [J]. IEEE Trans on Neural Networks, 1999, 10(3): 626−634.

[14] HYVÄRINEN A, OJA E. Independent component analysis: Algorithms and applications [J]. Neural Networks, 2000, 13(4−5): 411−430.

[15] SHAO Xue-guang, WANG Wei, HOU Zhen-yu, CAI Wen-sheng. A new regression method based on independent component analysis [J]. Talanta, 2006, 69(3): 676−680.

[16] KANEKO H, ARAKAWA M, FUNATSU K. Development of a new regression analysis method using independent component analysis [J]. J Chem Inf Model, 2008, 48(3): 534−541.

利用独立分量回归建立大坝形变模型

戴吾蛟[1]，刘 斌[1]，丁晓利[2]，黄大伟[1]

1. 中南大学 地球科学与信息物理学院，长沙 410083；
2. 香港理工大学 土地测量与地理咨讯系，香港九龙 红磡 999077

摘　要：在利用回归分析对大坝形变进行建模时，回归方程常因为自变量间的多重共线性而产生病态问题，从而不能建立准确的模型。提出利用独立分量回归(ICR)建立大坝形变模型，并确定大坝形变的物理响应。模拟实验表明：ICR 可以有效地解决病态问题，建立一个可靠的回归模型。将 ICR 用于中国湖南省的五强溪大坝建模，结果表明：ICR 不仅可以建立准确的坝体形变模型，而且通过其提取的独立分量可以确定大坝变形影响因素。

关键词：大坝变形分析；独立分量回归；主成分回归；病态问题；大坝变形解释

(Edited by Jing-hua FANG)