Research Journal of Applied Sciences, Engineering and Technology 6(6): 1035-1039, 2013
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: October 22, 2012 Accepted: December 28, 2012 Published: June 30, 2013

Application of Partial Least-Squares Regression Model on Temperature Analysis and Prediction of RCCD

Yuqing Zhao and Zhenxian Xing
North China University of Water Source and Electric Power, Zhengzhou 450011, China

Abstract: This study, based on the temperature monitoring data of jiangya RCCD, uses principle and method of partial least-squares regression to analyze and predict temperature variation of RCCD. By founding partial least-squares regression model, multiple correlations of independent variables is overcome, organic combination on multiple linear regressions, multiple linear regression and canonical correlation analysis is achieved. Compared with general least-squares regression model result, it is more advanced and accurate, had more practical explanation. It is proved feasible and practical, so, it can be used to predict concrete temperature. By calculating, the result shows that rock temperature is the most important factor which affects RCCD temperature. RCCD temperature is decreasing with rock temperature. We suggest that rock temperature should be monitored as emphasis in the future; this can provide some scientific basis for temperature controlling and preventing RCCD crack.

Keywords: Multiple linear regressions, partial least-squares regression, RCCD, temperature analysis and prediction

INTRODUCTION

In modern building, the cracks of the concrete is a universal engineering problem, It is so universal and difficult to solve (Malm and Ansell, 2010; Peng, 2005; Wu, 2000; Wu, 2002). There are many reasons that cause the cracks of the concrete, for example, deformation caused by improper maintenance, chemical reaction. Temperature variation, contraction, expansion and differential settlement can also cause deformation (Yang *et al*., 2012; Xu and Li, 2012). In addition to differential settlement and chemical reaction have nothing to do with the temperature, other concrete cracks, for example, shrinkage cracks, plastic shrinkage cracks, temperature cracks have something to do with the temperature. So it is very important to monitor temperature during dam surveillance (Lahmer and Tom, 2011; Chen *et al*., 2011; Yu *et al*., 2012, Bayagoob *et al*., 2010). And it is very necessary for dam safe surveillance and preventing concrete cracks to predict the trend of RCCD temperature based on monitoring data.

RCCD temperature is mainly affected by rock temperature, air temperature and water temperature and so on Kuzmanovic *et al*. (2010). These factors are multiple correlated, in order to improve accuracy, a lot of research works have been done by dam workers, a variety of calculating methods have been tried (Jin *et al*., 2012; Popescu and Theodor, 2011; Wang *et al*., 2011). The regression analysis has wide applications when predicting the temperature of the concrete. Least squares method is used usually to found regression equation between independent variables assembly and dependent variables. If independent variables have multiple correlation, deviation of least squares method is large and becomes unstable (Wang, 1999). So, a kind of new prediction method is very necessary.

Partial Least Squares regression analysis is a kind of new multivariate data analysis method. It is put forward in application fields (Fredl, 2009). It can combine model prediction analysis method with model-free data analysis method. Regression modeling, principal competent analysis and canonical correlation analysis are achieved under the same algorithm (Miles, 2006; Luo and Xing, 1988; Mata, 2011; Zhang *et al*., 2009). Moreover, the characteristics of multidimensional data may be observed on planar-draft. And great convenience is provided for multi-dimensional complex systems analysis, it makes things convenient for the application of engineering workers.

In this study ,based on the monitoring data o of RCCD, we uses partial least squares regression analysis method to analyze and predict temperature variation of RCCD, Unlike multiple linear regression analysis method, partial least squares regression analysis overcomes multiple correlation of independent variables, achieves organic combination on multiple linear regression, multiple linear regression and canonical correlation. The result showed that the performance of proposed method is better than that of the traditional calculating method.

RESREACH METHOD

The process of the model: there are a group of dependent variables $Y = (y_1, y_2 \ldots y_q)$ (q is the number of dependent variables) and independent variables X =

Corresponding Author: Yuqing Zhao, North China University of Water Source and Electric Power, Zhengzhou 450008, China

(x_1, x_2… x_q) (m is the number of independent variables) in multiple linear regression model. When overall data meets Gauss-Marov Theorem, by the Least Squares method:

$$B = X(X^T X)^{-1} X^T Y \quad (1)$$

B = Estimated regression coefficient.

When variable in X are severe multiple correlated, determinant (X^T, X) is almost equal to zero in formula (1). Serious rounding error will be contained when solving, $(X^T X)^{-1}$ sampling variability of regression coefficient estimates will become increased more significantly. Moreover, when variables in X are perfectly correlated, (X^T, X) is irreversible matrix, the regression coefficient cannot be solved .If regression model still be fitted by Least Squares model, regression result will appear many anomalisms, accuracy and reliability of regression will not be assured, partial least squares regression method can solve this kind of problem very well.

Partial Least Squares regression is integration and develop met of multiple linear regression, canonical correlation analysis and principal component analysis. It adopts following steps as: At first, the principal competent t_h (h = 1, 2,…) is extracted from the independent variables x, the components is independent; second, regression equation between these components and variables X is founded, the key is extraction of components, the components which is extracted from partial least squares regression not only can summarize the information in independent variables system, but also explain dependent variables very well. Therefore, regression modeling problem under the circumstances of the multiple correlations among variables can be solved effectively.

Partial least squares regression modeling: When q = 1, model is single variable, (PLS1); when q>1, model is multivariate. Now PLS1 process is given as:

Data standardization: The purpose of data standardization is to make collection centre of sample points coinciding with coordinate origin:

$$\begin{cases} F_0 = (F_{0y})_n \\ F_{0y} = \left[y - E(y) \right] / S_y \end{cases} \quad (2)$$

$$\begin{cases} E_0 = (E_{01,} E_{02,} \cdots E_{0m,})_{m \times n} \\ E_{0i} = \left[x_i - E(x_i) \right] / S_{xi} (i = 1, 2, \cdots m) \end{cases} \quad (3)$$

E_0 = Standardized matrix of X
F_0 = Standard matrix of y, $E(y)$
$E(xi)$ = Respectively the mean value of y and X
S_y and S_{xi} = Respectively the mean square deviation
n = The number of sample

The extraction of principal components: t_1, F_0 and E_0 have been known, so the first component t_1 is extracted from E_0, $t_1 = E_0 W_1$, W_1 is the first shaft of E_0, it is the combination coefficient. $\| w_1 \| = 1$, t_1 is linear combination of the standardized variables x^*_1, x^*_2, … x^*_m, this is the adjustment of original information.

Extracting the first component u_1 from F_0,$u_1 = F_0 C_1$, C_1 is the first shaft, $\| c_1 \| = 1$, so t_1 and u_1 should can represent the data variation information of X and y very well and t must possess the greatest explanatory ability for. According to the analysis principle of principal components and typical correlation analysis. In fact, covariance of t and u are required maximum. By derivation, formula (4) is got as:

$$\begin{cases} E_0^T F_0 F_0^T E_0 W_1 = \theta_1^2 W_1 \\ F_0^T F_0 E_0^T F_0 C_1 = \theta_1^2 C_1 \end{cases} \quad (4)$$

In formula (4):

θ_1 = Target function of the optimization problem
W_1 = Characteristic vector of $E^T_0 F_0 F^T_0 E_0$
θ^2_1 = Corresponding characteristic value
C_1 = Unit vector of maximum characteristic vector θ^2_1 which is correspond to matrix $F^T_0 F_0 E^T_0 F_0$

If θ_1 makes maximum, then W_1 become the Unit vector of maximum characteristic vector which is correspond to matrix $E^T_0 F_0 F^T_0 E_0$. In PLS1 process, C_1 =1, so, $u_1 = F_0$, E_0 and F_0 are unit vectors, formula (5) is obtained as:

$$w_1 = \frac{1}{\sqrt{\sum_{j=1}^{p} r^2(x_j, y)}} \begin{bmatrix} r(x_1, y) \\ \vdots \\ r(x_p, y) \end{bmatrix} \quad (5)$$

So, it can be obtained in Eq. (6):

$$t_1 = \frac{1}{\sqrt{\sum_{j=1}^{p} r^2(x_j, y)}} \left[r(x_1, y) E_{01} + \cdots + r(x_p, y) E_{0p} \right] \quad (6)$$

Then, the regression of for E_0, F_0 for t_1 is implemented, so:

$$E_0 = t_1 p'_1 + E_1 , \quad F_0 = t_1 r_1 + F_1 \quad (7)$$

p_1, r_1 is regression coefficient:

$$p_1 = \frac{E_0' t_1}{\| t_1 \|^2} \quad r_1 = \frac{F_0' t_1}{\| t_1 \|^2}$$

Residual matrix is:

$$E_1 = E_0 - t_1 p'_1 \quad F_1 = F_0 - t_1 r_1$$

The extraction of principal components t_2: Replacing E_0 with E_1, replacing F_0 with F_1, repeating the first step in the same way, so

$$r_1 = \frac{E_1' F_1}{\left\| E_1' F_1 \right\|} = \frac{1}{\sqrt{\sum_{j=1}^{p} Cov(E_{1j}, F_1)}} \begin{bmatrix} Cov(E_{11}, F_1) \\ \vdots \\ Cov(E_{1p}, F_1) \end{bmatrix} \tag{8}$$

$$t_2 = E_1 w_2 \tag{9}$$

The regression of E_1, F_1, for t_2 is implemented,

$$E_1 = t_2 p_2' + E_2, \quad F_1 = t_2 r_2 + F_2,$$

The coefficient of regression is following as:

$$p_2 = \frac{E_1' t_2}{\left\| t_2 \right\|^2} \quad r_2 = \frac{F_1' t_2}{\left\| t_2 \right\|^2}$$

The extraction of principal components: Rest components may be deduced by analogy. The third step, the fourth step…, until the component extraction number of partial least squares regression is determined using cross usefulness principle.

Reconstructing partial least squares regression model:

$$\hat{F}_0 = r_1 t_1 + r_2 t_2 + \cdots r_h t_h \tag{10}$$

According to property of Partial Least Squares regression, so:

$$t_i = E_{i-1} W_I = E_0 W^*_i \tag{11}$$

$$W^*_i = \prod_{k=1}^{i-1} (I - W_k P_k^T) W_i$$

Simultaneous Eq. (10) and (11), formula (12) is obtained as:

$$\hat{F}_0 = r_1 E_0 W_1^* + r_2 E_0 W_2^* + \cdots r_h E_0 W_h^* \tag{12}$$

Denoting:

$$y^* = F_0, x_i^* = E_{0i}, \quad \alpha_i = \sum_{k=1}^{h} r_k W_{ki}^*$$

Standardized regression equation is:

$$\hat{y}^* = \alpha_1 x_1^* + \alpha_2 x_2^* + \cdots \alpha_m x_m^* \tag{13}$$

Cross usefulness principle: y_i is monitoring data, t_i is the competent extracted from process of Partial Least Squares regression. y_{hi} is calculated values of sample points i, all sample points are used and components $t_1, t_2 \ldots t_h$ are calculated. $y_{h(-i)}$ is the calculated value of y_i ,when sample points are deleted, components t_1, $t_2 \ldots t_h$ are extracted to found regression model, then, calculating fitted values of y_i by this model:

$$\begin{cases} S_h = \sum_{i=1}^{n} (y_i - \hat{y}_{hi})^2 \\ PRESS_h = \sum_{i=1}^{n} (y_i - \hat{y}_{h(-i)})^2 \\ Q_h^2 = 1 - \frac{PRESS_h}{S_{h-1}} \end{cases} \tag{14}$$

When $Q_h^2 \geq 0.0975$, predicting ability of model will be improved obviously when new competent t_h is extracted, this is principle of cross usefulness.

RESULTS AND DISCUSSION

Jiangya hydraulic complex is full-face RCC gravity dam, the height of dam is 131 m. it is one of the highest gravity dam that have been build in the world.

The dam is divided into 13 parts. $5^{\#} \sim 7^{\#}$ is overflow dam blocks, $0^{\#} \sim 4^{\#}$ is retaining dam blocks on the right bank, $8^{\#} \sim 12^{\#}$ is retaining dam blocks on the right bank. In order to master the working condition of dam and provide reliable data for assuring the safety of dam. Dam sets up the safety detecting system. On the basis of 5# temperature monitoring data. Representative points of RCCD temperature are chosen to predict founding Partial Least Squares regression model. Monitoring data are shown in Table 1. In the same time, in order to reduce computation, pew software is adopted for calculate simulation.

Founding model: the calculated result of correlation coefficient between dependent variables and independent variables are shown in Table 2, it can be seen that dependent variables do not exists serious correlation, independent variable x2 is highly relevant (rock temperature) and predictor variable (concrete variable) y.

The number of principal components is calculated using software PEW based on principle of minimum prediction error. In most cases, the number of principal components by minimum principle of prediction error is consistent with the number by principle of cross usefulness. Calculating process of principal competent is shown in Table 3, the number of principal competent is three.

Table 1: Basic document of temperature and factors to concrete

Item number	Air temperature x_1 (°C)	Rock temperature x_2(°C)	Water temperature x_3(°C)	Concrete temperature y(°C)
1	7.96	27.56	14.94	26.3
2	10.00	27.53	12.60	26.12
3	13.00	27.50	12.49	25.93
4	20.60	27.50	13.06	25.8
5	25.35	27.30	12.60	25.68
6	26.54	27.15	13.28	25.49
7	30.18	27.28	16.37	25.29
8	28.95	27.28	18.47	25.10
9	26.93	27.32	19.19	24.95
10	17.20	27.26	18.21	24.81
11	13.64	27.18	17.09	24.64
12	11.02	27.17	15.96	24.50
13	8.06	27.15	14.36	24.38
14	9.55	27.14	13.30	24.25
15	14.49	27.08	12.78	24.11
16	17.51	26.93	12.79	23.97
17	22.88	26.91	13.00	23.82
18	25.52	26.88	14.40	23.65

Table 2: Correlation coefficient between x and y

r	x_1	x_2	x_3	y
x_1	1.000	0.214	0.289	0.051
x_2		1.000	0.080	0.932
x_3			1.000	0.054
y				1.000

Table 3: Process of extracting principal components

The number of principal components	Press error
1	2.402
2	1.475
3	1.368
4	1.442

Table 4: Cumulative explained ability value

Rd	t_1	t_2	t_3
x_1	0.0792	0.9199	1
x_2	0.9920	0.9946	1
x_3	0.0004	0.0001	1
X	0.3572	0.6382	1
y	0.8613	0.9234	0.9273

Partial Least Squares regression equation is obtained as: Original variables regression equation is:

$$\hat{y} = -82.5841 + 0.0235x_1 + 3.9715x_2 - 0.0717x_3$$

Standardized variables regression equation:

$$\hat{y}^* = 0.2189x_1 + 0.9951x_2 - 0.1972x_3$$

Model appraising:

The cumulative explained ability analysis: Cumulative explain ability value of t_1, t_2, t_3 are shown in Table 4. Cumulative explained ability of t_1, t_2, $t_3$for dependent variables and independent variables have reach 92%.so, t_1, t_2, t_3 can explain dependent variables and independent variables very well.

Fitting and examining analysis: in order to contrast, normal least square regression model is founded based on same data. Regression equation is following:

Table 5: Comparison of monitoring data and calculated results unit：C

		Partial least squares regression		Least squares regression	
Name	Monitoring data	Predicted value	Relative error (%)	Predicted value	Relative error (%)
Fitting stage	26.30	26.00	-1.15	26.12	-0.68
	26.12	26.07	-0.21	26.11	-0.04
	25.93	26.04	0.43	26.08	0.58
	25.80	26.18	1.47	26.26	1.79
	25.68	25.54	-0.53	25.52	-0.62
	25.49	24.92	-2.25	24.83	-2.58
	25.29	25.29	0.01	25.37	0.34
	25.10	25.12	0.08	25.26	0.63
	24.95	25.19	0.97	25.37	1.68
	24.81	24.78	-0.15	24.86	0.17
	24.64	24.44	-0.80	24.43	-0.84
	24.50	24.44	-0.25	24.39	-0.46
Fitting stage	24.38	24.39	0.07	24.27	-0.43
	24.25	24.46	0.84	24.3	0.21
	24.11	24.39	1.15	24.21	0.40
	23.97	23.88	-0.38	23.63	-1.42
	23.82	23.88	0.24	23.64	-0.78
	23.65	23.75	0.41	23.54	-0.46
Examining stage	23.47	23.55	0.32	23.43	-0.19
	23.27	23.21	-0.29	23.11	-0.69
	23.13	23.12	-0.02	23.04	-0.36
	22.99	22.95	-0.16	22.84	-0.62

$$y = -98.158 + 0.027x_1 + 4.525x_2 - 0.044x_3$$

Further comparative analysis is done by practical fitting and examining .the calculated results of two kind of model are shown in Table 5. It is known that Partial Least Squares regression model is better than least squares regression model according to the relative error, partial least squares regression model has more explanatory in practical system .

CONCLUSION

The temperature of RCCD is affected by rock temperature, air temperature and water temperature. These factors are not serious multiple correlated; large error will be brought using general least square regression model. Partial least square regression achieves integration of multiple linear regression, principal component and canonical correlation analysis. Compared with multiple linear regressions, it is more advanced, calculating result is more reliable, by calculate simulation, rock temperature is the most important factor influences concrete temperature of RCCD, temperature of dam body is decreasing slowly. Later, the emphasis of monitoring work should focus on rock temperature. Partial least square regression may be applied on analysis and prediction of the temperature of RCCD, some scientific basis is provided for temperature controlling and preventing crack of RCCD

REFERENCES

Bayagoob, K.H., A.A. Awad and A.A. Aeid, 2010. Coupled thermal and structural analysis of roller compacted concrete arch dam by three-dimensional finite element method. Struct. Eng. Mech., 36(4): 401-419.

Chen, S.H., P.F. Su and I. Shahrour, 2011. Composite element algorithm for the thermal analysis of mass concrete: Simulation of lift joint. Finite Elem. Anal. Des., 47(5): 536-542.

Fredl, B., 2009. Discussion of Different PLS Estimation Approaches for the Functional Logit Model. Electric Industry Press, Beijing.

Jin, F., Z. Chen and J. Wang, 2010. Practical procedure for predicting non-uniform temperature on the exposed face of arch dams. Appl. Thermal Eng., 30: 2146-2156.

Kuzmanovic, V., L. Savicand J. Stefanakos, 2010. Long-term thermal two and three dimensional analysis of roller compacted concrete dams supported by monitoring verification. Canadian J. Civil Eng., 37(4): 600-610.

Lahmer and Tom, 2011. Optimal experimental design for nonlinear ill-posed problems applied to gravity dams. Inverse Probl., 7(12): 66-83.

Luo, J. and Y. Xing, 1988. Economic Statistics Analysis Method and Prediction. Qinghua University Press, Beijing.

Mata, J., 2011. Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models. Inverse Probl., 33(3): 903-910.

Malm, R. and A. Ansell, 2010. Cracking of concrete buttress dam due to seasonal temperature variation. Aci. Struct. J., 108(1): 13-22.

Miles, 2006. Classical and Modern Regression with Applications. Academic Press, Beijing.

Peng, L., 2005. Temperature Control and Crack Prevention of Mass Concrete. The Yellow River Water Conservancy Press, China.

Popescu and D. Theodor, 2011. A new approach for dam monitoring and surveillance using blind source separation. Int. J. Innov. Comput. Inform. Control, 7(7): 3811-3824.

Wang, H., 1999. Partial Least Square Regression Method and Application. National Defense Industry Press, Beijing.

Wang, W.M., J.X. Ding and G.J. Wang, 2011. Stability analysis of the temperature cracks in Xiaowan arch dam. Technol. Sci., 54(3): 547-555.

Wu, C., 2000. Roller compacted concrete temperature control and crack prevention, advances in water science. Hehai University, 2(8): 16-25.

Wu, Z., 2002. The Theory and Application of Hydraulic Structure Security Monitoring and Controlling. Academic Press, Beijing.

Xu, H. and X. Li, 2012. Inferring rules for adverse load combinations to crack in concrete dam from monitoring data using adaptive neuro-fuzzy inference system. Sci. China-Tech. Sci., 55: 136-141.

Yang, J., Y. Hu and Z. Zuo, 2012. Thermal analysis of mass concrete embedded with double-layer staggered heterogeneous cooling water pipes. Appl. Thermal Eng., 35: 145-156.

Yu, H., S. Li and Y. Liu, 2012. Study on temperature distribution due to freezing and thawing at the fengman concrete gravity dam. Thermal Sci., 15: S27-S32.

Zhang, H., Z. Dang and F. Yu, 2009. Application of grey linear regression model on calculating resistance coefficient of pipe. J. North China Instit. Water Conserv. Hydroelect. Power, 1: 12-14.