

DOI:10.11705/j.issn.1672-643X.2017.05.39

# 基于独立分量分析的拱坝位移预测模型

刘甲奇<sup>1,2,3</sup>, 包腾飞<sup>1,2,3</sup>, 邓元倩<sup>4</sup>

(1. 河海大学 水利水电学院, 江苏 南京 210098; 2. 河海大学 水文水资源与水利工程科学  
国家重点实验室, 江苏 南京 210098; 3. 河海大学 水资源高效利用与工程安全国家工程  
研究中心, 江苏 南京 210098; 4. 河海大学 水文水资源学院, 江苏 南京 210098)

**摘要:** 在逐步回归模型中, 有些分量间存在比较强的相关性, 如果直接利用线性回归方法处理, 就会特别容易出现病态问题, 这样得到的回归模型不合理。为了解决这一问题, 引入独立分量分析方法(ICA), 并运用基于负熵极大化的独立性准则的快速独立分量分析方法(FastICA)分离出独立分量, 建立了某拱坝的安全监测位移预测模型; 同时为了表明结果准确性, 利用逐步回归方法建立了逐步回归模型。结果表明: 独立分量回归方法(ICR)相比逐步回归方法可以建立一个比较准确的回归模型, 由于独立分量回归分析方法没有剔除不显著因子, 考虑了各个因子对因变量的影响, 从而更加符合实际情况。

**关键词:** 拱坝位移; 独立分量分析; 安全监测; 逐步回归; 变形分析; 预测模型

中图分类号: TV698.1 文献标识码: A 文章编号: 1672-643X(2017)05-0237-05

## Arch dam displacement prediction model based on independent component analysis

LIU Jiaqi<sup>1,2,3</sup>, BAO Tengfei<sup>1,2,3</sup>, DENG Yuanqian<sup>4</sup>

(1. College of Water Conservancy & Hydropower Engineering, Hohai University, Nanjing 210098, China;  
2. State Key Laboratory of Hydrology - Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China;  
3. National Engineering Research Center of Water Resources Efficient Utilization and Engineering Safety, Hohai University, Nanjing 210098, China; 4. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China)

**Abstract:** In the stepwise regression model, there are strong correlations between some of the components. If the linear regression method is adopted, the regression ill-posed problem will be very likely to occur, so the regression model obtained will be unreasonable. In order to solve this problem, independent component analysis (ICA) is introduced and the displacement prediction model of a arch dam is established using the fast independent component analysis (FastICA) method based on the principle of independence of maximum entropy. In order to show the accuracy of the results, the stepwise regression model is established by stepwise regression method at the same time. The results show that the independent component regression method (ICR) can establish a more accurate regression model than the stepwise regression method. The reason is that the independent component regression analysis method does not eliminate the non-significant factors, which takes into account the effect of each factor on the dependent variable, so it is more in line with the actual situation.

**Key words:** arch dam displacement; independent component analysis; safety monitoring; stepwise regression analysis; deformation analysis; prediction model

## 1 研究背景

拱坝监测位移模型一般利用逐步回归分析方法

来建立, 在逐步回归的拱坝位移模型中, 由于其留下了显著因子, 一些对因变量有较小影响的因子被去除掉, 从某种程度上来讲, 其实是舍弃了这些因子所包

收稿日期: 2017-05-22; 修回日期: 2017-06-14

基金项目: 国家自然科学基金项目(51579086、51479054、51379068、51139001); 江苏省杰出青年基金项目(BK20140039); 江苏高校优势学科建设工程项目(水利工程)(YS11001)

作者简介: 刘甲奇(1991-), 男, 在读硕士研究生, 研究方向为水工建筑物安全监测。

通讯作者: 包腾飞(1974-), 男, 博士, 教授, 博士生导师, 研究方向为水工建筑物安全监控、评估及反馈分析。

含的一些数据信息<sup>[1]</sup>。此外,在利用逐步回归分析方法建模时,由于某些因子之间存在较强的相关性,回归方程中难免会出现一些病态问题,这就导致了模型预报的结果不准确。例如在位移预测模型中的水压分量因子中的  $H^1$ 、 $H^2$ 、 $H^3$ 、 $H^4$  之间存在很强的相关性,如果直接进行线性回归,回归模型中特别就会出现病态问题,得出不合理的回归模型。

为解决建模过程中的上述问题,采用独立分量分析的方法来解决在回归模型建立过程中出现的这些问题。独立分量分析(ICA, Independent Component Analysis)是一种十分有效的盲源分离技术(BSS),Bell和Sejnowski用实验证明了ICA是解决BSS问题简单且高效的一个算法<sup>[2]</sup>,从而引起ICA研究的热潮;Hyvriinen提出基于负熵极大化的快速独立分量分析算法——FastICA算法,它是一种近似牛顿迭代的算法,不仅其收敛的速度快,而且精度高,使得ICA技术走向应用<sup>[3]</sup>。目前ICA技术已在语音识别、医学信号处理以及图像处理等领域取得成功。本文利用独立分量分析方法建立了拱坝变形的预测模型,探讨了其在监测位移数据处理中的应用。

## 2 ICA的基本原理

ICA方法的处理对象为一组具有相互统计独立特点的源信号经过一系列的线性组合而产生的混合信号,其主要目的就是在源信号未给出的条件下,从这些组合产生的混合信号当中提取出组合成混合信号的各个独立源信号。

### 2.1 ICA的基本模型

在不考虑噪声的情况下,假设有 $M$ 个观测向量: $X = [X_1, X_2, \dots, X_M]^T$ ,有 $N$ 个独立源信号 $S = [S_1, S_2, \dots, S_N]^T$ ,此 $M$ 个观测向量是由 $N$ 个独立源信号经过一定的线性组合而形成的。即:

$$X = AS \quad (1)$$

式中: $A$ 为混合系数 $a_{ij}$ 构成的 $M \times N$ 混合矩阵;独立源信号 $S$ 和混合矩阵 $A$ 均为未知,只有混合信号 $X$ 为已知<sup>[4-5]</sup>。ICA的主要目的是根据式(1)求出分离矩阵 $W$ ,然后利用分离矩阵 $W$ 从观测信号 $X$ 中恢复出源信号 $S$ 的一个近似估计 $Y$ 。

设输出的分离信号为 $Y = [Y_1, Y_2, \dots, Y_M]^T$ ,则分离过程为<sup>[4]</sup>:

$$Y = WX = WAS \quad (2)$$

当 $M = N$ 时,令(2)式中 $WA = B$ ,当 $B = I$ 时, $Y = S$ ,这是理想情况,在求出 $B$ 为广义置换矩阵(每一行和每一列都只有一个非零元素)时,可认为ICA

分离成功,这时输出信号 $Y$ 是独立源信号 $S$ 的一个估计,且 $Y$ 的各个分量间是相互独立的。

### 2.2 快速独立分量分析(FastICA)算法

FastICA是一种基于负熵最大化的快速ICA固定点算法,它的优点是计算简单、收敛的速度快,在实际当中得到了广泛的运用<sup>[6-7]</sup>。由于负熵是衡量信号非高斯性的一个依据,而根据中心极限定理<sup>[4]</sup>:多个独立的随机变量的混合信号逼近高斯分布,即多个独立源信号的混合信号的高斯性大于任意单个源信号的高斯性,在所有分离出来的源信号的非高斯性都达到最大时,每个分离出来的源信号也就越会接近于不同的单个源信号,从而达到分离信号中的各个分量逼近互相独立的目的。

FastICA算法<sup>[8-11]</sup>的实现步骤如下:

步骤1:先对监测数据 $X$ 进行中心化和白化预处理。中心化是指把观测信号 $X$ 减去其平均值,变换为均值为零的向量;白化是指对观测信号 $X$ 进行线性变换 $Q$ ,即 $Z = QX$ ,其中得到的 $Z$ 的各分量 $Z_i$ 之间互不相关,且具有单位方差,即 $E\{ZZ^T\} = I$ ;

步骤2:随机选取一个单位范数的权值向量 $W$ ;

步骤3:通过牛顿迭代法得到简化迭代公式 $w(k+1) = E\{xg(w^T x)\} - E\{g'(w^T x)\}w(k)$ 对 $W$ 进行迭代计算。其中 $g(\cdot)$ 为任意非线性函数,选用 $G_1 = 1/a_1 \log \cosh(a_1 u)$ , $G_1$ 适用于亚高斯和超高斯信号并存的一般情况<sup>[4]</sup>。一般取 $1 \leq a_1 \leq 2, a_2 = 1$ ;

步骤4:标准化 $W, w_i = w_i / \|w_i\|$ ;

步骤5:若第一阶段的提取算法结果不收敛,则需要返回步骤3;若 $\|w_i\|$ 没有变化或者变化非常小(小于一个定值时),则迭代计算结束,得到分离出来的独立分量 $C$ 。

## 3 利用ICA建立拱坝位移回归模型

由于影响拱坝位移量的主要因子为水压、温度、时效3类因子<sup>[12]</sup>,这3类因子在统计上是相互独立的,因此考虑将独立分量分析方法引入到拱坝位移回归模型的建立中,对拱坝的效应量进行处理,分离得到3个主要的独立分量,然后利用最小二乘法,将分离得到的独立分量和影响拱坝变形的各个因子建立多元回归方程,最终得到基于独立分量回归的拱坝位移回归方程。

利用独立分量分析方法 $C = WX$ 分析出的独立分量 $C$ 中包括观测数据 $X$ 中的大多数变化和信  
息<sup>[13]</sup>,其中 $W$ 为分离矩阵,用观测数据 $X$ 的独立分量 $C$ 来替换输入变量,然后再进行回归分析,这样就

建立了提取出的独立分量与因变量间的多元回归的模型——独立分量回归模型:

$$Y = d_0 + d_1 C_1 + d_2 C_2 + \dots + d_l C_l \quad (3)$$

独立分量回归模型的建模<sup>[8-9,11]</sup>步骤:

(1) 采集位移数据  $X = [X_1, X_2, \dots, X_M]$ , 对数据进行误差分析, 去除误差, 并对数据去中心化和白化预处理;

(2) 利用 FastICA 对预处理后的数据进行独立分量分析, 以目标函数为  $G(y) = -(1/a_1) \log \cosh(a_1 y)$ , 求出影响位移数据的独立分量  $C$  和分离矩阵  $W$ , 一般  $a_1$  取 1;

(3) 在第(2)步的基础上, 建立独立分量  $C$  与因变量  $Y$  的多元回归模型, 然后利用最小二乘法求解方程:  $Y = d_0 + d_1 C_1 + d_2 C_2 + \dots + d_l C_l$ , 得到回归系数  $d_0, d_1, \dots, d_l$ ;

(4) 利用提取出的独立分量  $C$  和分离矩阵  $W$ , 求得各个观测自变量所对应的系数。

### 4 实例分析

由坝工理论可知: 拱坝的位移  $\delta$  按其形成的原因可分为 3 个部分<sup>[12]</sup>: 水压分量  $\delta_H$ 、温度分量  $\delta_T$  和时效分量  $\delta_\theta$ , 即:

$$\delta = \delta_H + \delta_T + \delta_\theta \quad (4)$$

#### 4.1 水压分量

由工程力学及坝工理论可知, 重力拱坝水平位移的水压分量  $\delta_H$  与水深  $H^1, H^2, H^3, H^4$  有关, 即:

$$\delta_H = \sum_{i=1}^4 a_i H^i \quad (5)$$

式中:  $H$  为坝前水深;  $a_i$  为水压因子回归系数。

#### 4.2 温度分量

温度分量  $\delta_T$  的表达方式常见有以下两种:

(1) 考虑坝体内部温度变化滞后于气温, 同时考虑混凝土温度位移量同温度变化呈线性关系, 故选取多组前期气温平均值的线性组合来建立温度分量  $\delta_T$  的数学表达式, 即:

$$\delta_T = \sum_{i=1}^{m_1} b_i \bar{T}_i \quad (i = i_1, i_2, \dots, m_l) \quad (6)$$

式中:  $b_i$  为回归系数;  $T_i$  为观测日前  $i$  日气温平均值。 $i$  的取值根据实际情况来定, 重力坝可取  $i = 5, 20, 60, 90$  d。

(2) 考虑坝体内部温度随环境气温近似做简谐波动, 且滞后于环境气温某一相位角, 同时考虑混凝土的热膨胀变形量与其温度成线性关系, 故选用多周期谐波函数的线性组合来建立温度分量  $\delta_T$  的表

达式, 即:

$$\delta_T = \sum_{i=1}^n \left( b_{1i} \sin \frac{2\pi it}{365} + b_{2i} \cos \frac{2\pi it}{365} \right) \quad (7)$$

式中:  $n$  根据工程的实际情况取 2;  $t$  为测值当天至起测日累计天数;  $b_{1i}, b_{2i}$  为温度因子回归系数,  $i = 1 \sim 2$ 。

#### 4.3 时效分量

时效分量  $\delta_\theta$  一般用下式表示:

$$\delta_\theta = c_1 \theta + c_2 \ln \theta \quad (8)$$

式中:  $\theta$  为监测日至始测日的累计天数除以 100;  $c_1, c_2$  为时效因子回归系数。

由以上分析, 可得该拱坝的径向位移的混合模型的表达式为:

$$\begin{aligned} \delta &= \delta_H + \delta_T + \delta_\theta \\ &= a_0 + \sum_{i=1}^n a_i H^i + \sum_{i=1}^n \left( b_{1i} \sin \frac{2\pi it}{365} + b_{2i} \cos \frac{2\pi it}{365} \right) + c_1 \theta + c_2 \ln \theta \end{aligned} \quad (9)$$

本文选取对某拱坝 3 条倒垂线的径向位移监测数据进行分析。算例中采用该拱坝 5 年的监测数据为分析对象进行独立分量回归建模。

图 1 为该拱坝典型监测点 PL-7、PL-8、PL-9 的径向位移的过程线图。该拱坝距今已建成多年, 时效分量对拱坝位移的影响已经稳定, 而且该拱坝测点的变化幅值在一定的范围内有规律地进行波动, 选取几个典型监测点的历年测值数据进行研究分析。

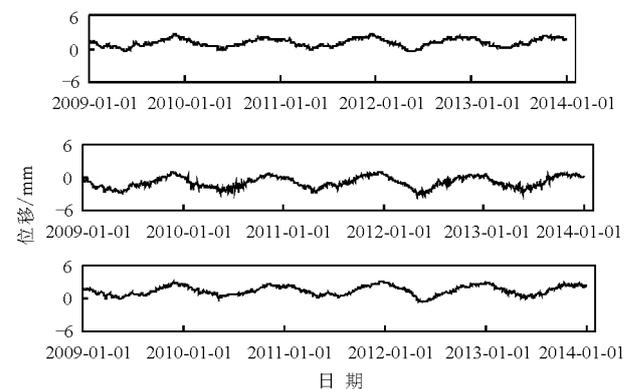


图 1 典型监测点顺河向水平位移历时曲线图

图 2 与图 3 是利用该拱坝 2009 年 1 月到 2013 年 12 月监测的水位和温度数据绘制的历时曲线图, 由图可知, 水位与温度都呈现一定的周期性, 有利于解混时恢复出源信号。

为了求解方便, 首先对观测数据  $X$  进行中心化与白化预处理, 中心化就是将数据去除均值, 白化一般采用的方法是标准主分量分析 (PCA) 的方法去除观测数据之间的相关性, 然后根据 FastICA 一般的

求解过程,对预处理后的数据进行 FastICA 处理,求出分离矩阵  $W$ ,进而得到分离出来的各个独立分量数据  $C$ ,然后对分离出来的独立分量数据运用最小二乘法计算得到模型的回归系数,最后利用分离矩阵和求解到的回归系数得到各个变量的回归系数,最终回归结果可以反映各个自变量对位移的影响。

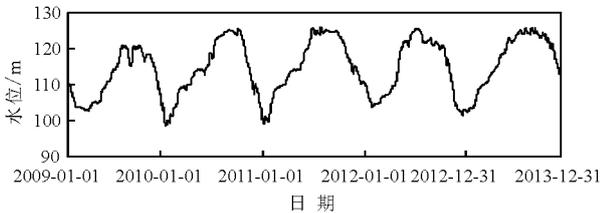


图2 2009-01-01 - 2013-12-31 水位历时曲线

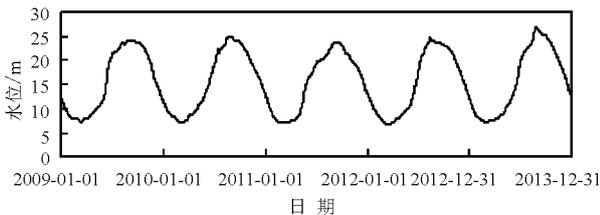


图3 2009-01-01 - 2013-12-31 温度历时曲线

在本例中,把独立分量分析理论应用到拱坝变形分析中,利用独立分量回归的方法建立拱坝变形回归模型。剔除噪声信号的影响,拱坝变形的位移是水位分量、温度分量与时效分量 3 个信号的叠加<sup>[14]</sup>,这 3 个信号都服从统计独立,选择 3 个典型垂线位移的监测点:PL-7、PL-8、PL-9,把这 3 个监测点看为 3 个信号接收器,而源信号也有 3 个,这就满足了盲源分离中接收器个数大于或者等于源信号个数的要求<sup>[15]</sup>,这时就可把监测数据看成是这 3 个源信号的线性混合来研究分析。

接下来运用 FastICA 方法对该拱坝的位移数据进行分析处理,得到 3 个独立分量:水位分量、时效分量和温度分量,如图 4 所示:

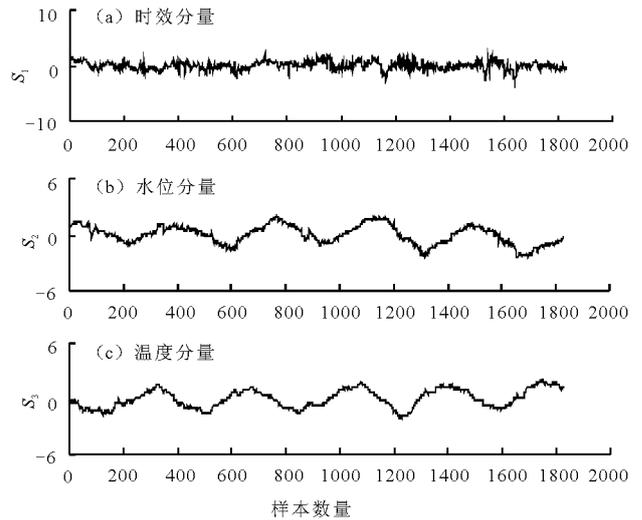


图4 解混后的时效、温度和气压分量

根据解混后的独立成分的数据信号,因变量位移只与水位、温度、时效这 3 个变量相关,故可以假设回归模型<sup>[14]</sup>为:

$$Y = b_0 + b_1S_1 + b_2S_2 + b_3S_3 \quad (10)$$

应用 ICA 方法在公式(10)中包括了水位、温度和时效效应。由于分离出来的变量是相互独立的,所以就减少了各个变量之间的多重共线性,利用这些分离出来的独立分量数据建立回归模型。

利用图 4 中的各个分量的数据,按照 FastICA 回归的步骤进行回归模型计算,得到的回归模型为:

$$Y = 0.005 + 0.043S_1 + 0.531S_2 + 0.646S_3 \quad (11)$$

式中: $S_1$  为时效分量;  $S_2$  为水位分量;  $S_3$  为温度分量。

为了验证该方法的合理性,对观测数据进行了逐步回归模型的建立,将逐步回归模型与独立分量回归模型两种回归方法比较得到的回归系数结果如表 1 所示。

表1 两种方法得到的各个分量系数表

方法	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$b_1$	$b_2$	$b_3$	$b_4$	$c_1$	$c_2$
ICR	2.432	-0.004	$5.09 \times 10^{-3}$	$-8.22 \times 10^{-4}$	$9.18 \times 10^{-4}$	0.736	-0.829	0.632	0.552	0.009	0.096
逐步回归	2.102	0	125.426	0	-138.669	18.658	-12.594	17.325	15.142	-0.953	1.131

表 1 中常数项  $a_0$  数值较大,这是由于没有考虑初始状态的影响,但并不影响该研究的结论。利用建立好的独立分量回归模型对拱坝 PL-7 测点 2014 年 1 月 1 日 - 2014 年 4 月 1 日的水平位移进行预测,将预测值和实测值对比,对比结果如图 5 所示。

从图 5 可以看出,采用独立分量回归的方法的预

测结果比逐步回归的好。采用 MAPE(平均绝对百分误差)、RMSE(均方根误差)、MAE(平均绝对误差)三项指标评价预报结果<sup>[16-17]</sup>,其计算公式如下:

$$MAPE = \frac{\sum_{i=1}^n \frac{\varepsilon_i}{y_i}}{n} \times 100\% \quad (12)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n \varepsilon_i^2}{n}} \quad (13)$$

$$MAE = \frac{\sum_{i=1}^n \text{abs}(\varepsilon_i)}{n} \quad (14)$$

式中: $\varepsilon_i = |y_i - \hat{y}_i|$ ,  $\hat{y}_i$  为回归模型预报(拟合)值, 为预报(拟合)期数。

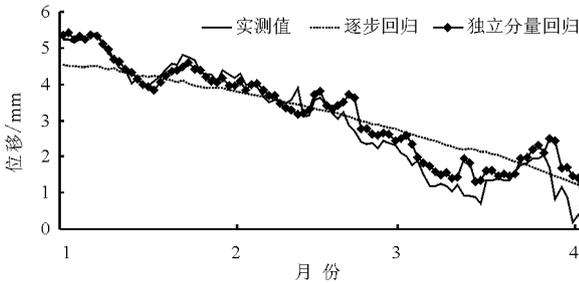


图 5 2014-01-01 - 04-01 各模型拟合曲线图

两种方法的回归参数评价统计表如表 2 所示, 独立分量回归模型的复相关系数在 0.9863, 说明该回归模型的拟合程度较好。与逐步回归得到的模型结果对比如表 2 所示。

表 2 回归模型的复相关系数 R 及预报精度指标

模型	复相关系数 R	MAPE / %	RMSE / mm	MAE / mm
独立分量回归	0.9863	3.26	0.68	0.70
逐步回归	0.9636	3.53	0.76	0.79

从表 2 可以看出, 通过计算得到独立分量回归模型的 MAPE 为 3.26%, RMSE 为 0.68 mm, 将独立分量分析回归模型得到的拟合值与实测值相比较, 其差值基本都在 1 mm 之内, 预报的精度比较高, 其精度预报符合拱坝变形监测的精度要求。

## 5 结 论

本文将独立分量分析方法引入到拱坝的位移变形预测模型中, 有效地解决了拱坝在建模过程中存在的线性相关问题。由工程实例的结果得出:

(1) 独立分量回归的优点在于它分离出来的各个分量之间是相互独立的, 因此避免了回归方程中存在的病态问题, 并且提高了模型的预测精度, 取得了较为理想的结果;

(2) 利用独立分量回归方法计算出的独立分量中包含了各个自变量对因变量的影响系数, 所以更加符合实际情况, 这样就可以对建立的模型各个自

变量系数大小的比较来确定各个因素对拱坝位移的影响的大小, 从而确定位移的成因来指导拱坝安全的监测工作。

## 参考文献:

- [1] 李红祥, 岳东杰, 李立瑞. 基于主成分回归的大坝位移模型[J]. 水电与抽水蓄能, 2008, 32(5):61-64.
- [2] 胡楠. 独立分量分析在滑坡预报中的应用[D]. 成都: 成都理工大学, 2008.
- [3] 杨俊美, 余华, 韦岗. 独立分量分析及其在信号处理中的应用[J]. 华南理工大学学报(自然科学版), 2012, 40(11):1-12.
- [4] 王明祥. 独立分量分析方法及在图像处理中的应用研究[D]. 上海: 上海大学, 2005.
- [5] 杨福生, 洪波. 独立分量分析的原理与方法[M]. 北京: 清华大学出版社, 2006.
- [6] 张杰, 刘辉, 欧伦伟. 改进的 FastICA 算法研究[J]. 计算机工程与应用, 2014, 50(6):210-212.
- [7] 张明远, 宋华珠, 李彬, 等. 基于 ICA-SVM 的超长直径钢管桩承载力预测[J]. 岩土力学, 2012, 33(9):2759-2764.
- [8] 刘斌, 戴吾蛟, 曾凡河, 等. 利用独立分量回归建立大坝多测点位移模型[J]. 大地测量与地球动力学, 2016, 36(2):124-128.
- [9] Dai Wujiao, Liu Bin, Ding Xiaoli, et al. Modeling dam deformation using independent component regression method[J]. Transactions of Nonferrous Metals Society of China, 2013, 23(7):2194-2200.
- [10] 王建雄, 张立民, 钟兆根. 基于 FastICA 算法的盲源分离[J]. 计算机技术与发展, 2011, 21(12):93-96.
- [11] 刘斌, 戴吾蛟, 黄大伟, 等. 独立分量回归及其在变形分析中的应用研究[J]. 大地测量与地球动力学, 2012, 32(6):90-93.
- [12] 吴中如. 水工建筑物安全监控理论及其应用[M]. 南京: 河海大学出版社, 2003.
- [13] 杨英华, 杨劭伟, 刘晓志, 等. 基于独立分量回归的加热炉钢温预报模型[J]. 系统仿真学报, 2008, 20(10):2523-2525.
- [14] 詹新武. 独立分量分析理论及其在变形监测数据处理与分析中的应用研究[D]. 长沙: 中南大学, 2014.
- [15] 孟宗, 王晓燕, 马钊. 融合小波分解与时频分析的单通道振动信号盲分离方法[J]. 中国机械工程, 2015, 26(20):2751-2756.
- [16] 李潇. 组合动态变形预测模型及其方法研究[D]. 武汉: 武汉大学, 2008.
- [17] 李双平, 张斌. 基于小波与谱分析的大坝变形预报模型[J]. 岩土工程学报, 2015, 37(2):374-378.