混凝土坝小样本安全监控模型研究

李占超^{①②*}, 张慧莉^①, 刘兴阳^①, 秦栋^②^① 西北农林科技大学水利与建筑工程学院, 杨凌 712100;^② 河海大学水文水资源与水利工程科学国家重点实验室, 南京 210098

* E-mail: lizhanchao@gmail.com

收稿日期: 2014-05-16; 接受日期: 2014-06-28

国家自然科学基金(批准号: 51309195, 51139001)、水文水资源与水利工程科学国家重点实验室开放研究基金(批准号: 2012490211)、中央高校基本科研业务费专项资金(批准号: QN2013046)和西北农林科技大学博士点基金(批准号: 2012BSJJ007)资助项目

摘要 混凝土坝安全监控模型研究一直是大坝安全监控领域中的热点和难点, 本文将混凝土坝结构性态的演变作为一个动态过程, 针对混凝土坝安全监控模型中监测数据较少、自变量因子间相关性较强的问题, 结合混凝土坝结构性态转异诊断的 Bootstrap 法和偏最小二乘回归分析方法, 建立了混凝土坝小样本安全监控模型. 将其应用到某混凝土重力坝水平位移的分析当中, 结果表明, 混凝土坝小样本安全监控模型的拟合精度更高、预测效果更好、解释更加合理, 研究成果为混凝土坝安全监控分析提供了科学依据.

关键词混凝土坝
安全监控模型
小样本
转异诊断
偏最小二乘回归

1 引言

在混凝土坝安全监控模型研究中, 国内外普遍采用多元回归分析方法(如逐步回归法等), 研究监测效应量与环境因素等自变量之间的统计关系及其数学形式^[1]. 此时, 认为各自变量因子之间不存在密切的相关关系. 然而, 研究表明混凝土坝自变量因子间或多或少地存在一定的多重共线性, 基于普通最小二乘法的安全监控模型很难精确地分离水压、温度、时效等分量, 无法对监测效应量的变化做出准确的定量分析与合理的物理成因解释^[2~4]. 尤其是对于小型水库的混凝土坝而言, 在库水位变化不大时, 水深的一、二、三、四次项之间相关性强, 导致求回归系数的方程式出现病态. 克服自变量间多重共线性问题的主要方法有岭回归、主成分回归以及偏最小二乘回归(PLSR), 其中, 岭回归是直接减少设计矩阵的病

态, 而主成分回归和偏最小二乘回归是基于因子的提取, 使新的因子变量相互独立, 满足 Gauss-Markov 条件, 先建立因变量对因子的回归, 再还原成与自变量的回归模型^[5].

偏最小二乘回归在大坝安全监测资料分析中的应用是大坝安全监控中的热点之一, 邓念武^[6]首先采用偏最小二乘回归对大坝位移监测资料进行建模, 研究了单因变量偏最小二乘回归模型在土石坝沉降资料分析预测中的应用. 徐洪钟和吴中如^[7]针对统计回归计算中的水压因子难以入选以及入选以后计算结果不合理的问题, 应用偏最小二乘回归建立了坝顶水平位移的统计模型, 并与逐步回归和多元回归进行对比分析, 结果表明, 相对于逐步回归和多元回归, 基于偏最小二乘法的回归模型能够取得更加合理的回归系数. 仲琳^[8]将偏最小二乘法和模拟退火算法相结合, 对考虑碾压混凝土坝层面影响的坝体渗

引用格式: 李占超, 张慧莉, 刘兴阳, 等. 混凝土坝小样本安全监控模型研究. 中国科学: 技术科学, 2014, 44: 1043–1051

Li Z C, Zhang H L, Liu X Y, et al. Safety monitoring model with small samples of concrete dam (in Chinese). Sci Sin Tech, 2014, 44: 1043–1051, doi: 10.1360/N092014-00182

流量监控模型进行优化. 杨杰^[9]提出了以多重相关性干扰消除和数据内涵不确定性分析为核心的大坝安全监测“偏最小二乘回归——不确定性分析”综合分析方法, 并研究了其在土石坝绕坝渗流监测分析中的应用. 以上研究主要是利用偏最小二乘回归分析在克服自变量相关性上的优势, 而对于大坝监测效应量样本空间较小的情况, 已有研究还较少.

同时, 混凝土坝结构性态的演变是一个动态过程, 当坝体或坝基在外部荷载的作用下产生显著损伤时, 混凝土坝结构性态的演变规律将发生变化^[10]. 为评估结构当前的安全状态及预测结构性态的演变趋势, 需采用转异诊断方法对混凝土坝结构性态进行划分^[11]. 赖道平^[10]基于 Hurst 指数和相空间重构法, 提出了两种计算大坝分形维数诊断大坝结构性态的方法; 基于模糊推理和 D-S 证据推理, 提出了大坝结构性态转异诊断的数据融合方法. 而将大坝结构性态的转异诊断方法和安全监控模型相结合进行大坝效应量的建模和分析, 现有的研究还较少, 其中 Bao 等人^[12]将转异诊断方法应用到裂缝的安全监控模型中, 建立了混凝土坝裂缝转异监控模型, 分析结果较为合理.

综合以上分析, 自变量因子间的强相关性对混凝土坝安全监控模型的影响较大, 同时, 为分析混凝土坝结构性态的动态演变规律, 需要采用转异诊断方法划分监测效应量. 此时, 对应于当前结构性态的监测数据可能较少, 尤其是对于新建混凝土坝, 其监测序列较短且结构性态变化复杂. 因此, 本文针对自变量因子间的强相关性及监测数据较少问题展开研究, 将转异诊断的 Bootstrap 法和偏最小二乘回归法相结合, 建立混凝土坝小样本安全监控模型, 以期混凝土坝性态安全评估提供依据.

2 混凝土坝小样本安全监控模型的构建

2.1 混凝土坝结构性态转异诊断的 Bootstrap 法

现有的混凝土坝结构性态转异诊断方法的理论基础大多为中心极限定理^[13,14], 但是检验统计量收敛到其极限分布的速度通常非常慢, 并且在很多情况下统计量的极限分布形式是未知的. 而且, 对于时间序列模型, 极限分布并不能很好地考虑小样本情况, 而采用 Bootstrap 方法选取转异临界值能够充分考虑样本较少的情况^[15].

对于混凝土坝结构性态安全监控的线性回归模型

$$\mathbf{y}(i) = \mathbf{X}(i)^T \boldsymbol{\beta}_{i,m} + \mathbf{e}(i), \quad i \geq 1. \quad (1)$$

定义非污染假设及相应的转异检验假设^[16]

$$\boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_0, \quad 1 \leq i \leq m.$$

$H_0: \boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_0, \quad m < i < m + T_m + 1$, 混凝土坝结构性态未转异;

H_1 : 存在 $k_m^\circ \geq 0$ 满足 $\boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_0, m < i \leq m + k_m^\circ$, $\boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_m^\circ \neq \boldsymbol{\beta}_0, m + k_m^\circ < i < m + T_m + 1$, 混凝土坝结构性态发生转异.

定义统计量^[15]

$$\boldsymbol{\Gamma}(m, k, \gamma) = \frac{\sum_{m < i \leq m+k} (\mathbf{y}(i) - \mathbf{X}(i)^T \hat{\boldsymbol{\beta}}_m)}{g(m, k, \gamma)}, \quad (2)$$

其中, $\hat{\boldsymbol{\beta}}_m = C_m^{-1} \sum_{j=1}^m \mathbf{X}(j) \mathbf{y}(j)$ 为回归系数基于历史数据 $\mathbf{y}(1), \dots, \mathbf{y}(m)$ 的估计, 其中 $C_m = \sum_{i=1}^m \mathbf{X}(i) \mathbf{X}(i)^T$;

$$g(m, k, \gamma) = m^{1/2} \left(1 + \frac{k}{m} \right) \left(\frac{k}{m+k} \right)^\gamma, \quad 0 \leq \gamma < 1/2.$$

从而, 检验统计量为^[17]

$$\frac{1}{\hat{\sigma}_m} \sup_{1 \leq k < T_m+1} |\boldsymbol{\Gamma}(m, k, \gamma)|, \quad (3)$$

其中, 方差的相合估计可表示为

$$\hat{\sigma}_m^2 = \frac{1}{m-p} \sum_{i=1}^m (\mathbf{y}(i) - \mathbf{X}(i)^T \hat{\boldsymbol{\beta}}_m)^2.$$

在原假设及在 $l \leq k^\circ$ 的备择假设下, 存在

$$\begin{aligned} & \mathbf{y}(l) - \mathbf{X}(l)^T C_m^{-1} \sum_{j=1}^m \mathbf{X}(j) \mathbf{y}(j) \\ &= \mathbf{e}(l) - \mathbf{x}(l) C_m^{-1} \sum_{j=1}^m \mathbf{e}(j) \mathbf{X}(j). \end{aligned} \quad (4)$$

因此,

$$\boldsymbol{\Gamma}(m, l, \gamma) = \frac{\left(\sum_{i=m+1}^{m+l} \mathbf{e}(i) - \sum_{i=m+1}^{m+l} \mathbf{X}(i)^T C_m^{-1} \sum_{j=1}^m \mathbf{X}(j) \mathbf{e}(j) \right)}{g(m, l, \gamma)}. \quad (5)$$

假定已经获得 $m+k$ 个观测, 并且直到目前原假设没有被拒绝. 基于所得到的 $m+k$ 个观测 $(\mathbf{y}(i), \mathbf{X}(i))$, $i=1, \dots, m+k$ 的混凝土坝性态转异诊断的 Bootstrap 法如下.

首先, 利用 Bootstrap 估计 $\mathbf{e}_{m,k}^*(i)$ 代替式中的 $\mathbf{e}(i)$,

并且保持 $X(i), i=1, \dots, m+k$ 不变. 其中, Bootstrap 误差定义为

$$e_{m,k}^*(i) = \hat{e}_{m,k}(U_{m,k}(i)), i=1, \dots, m+T_m, \quad (6)$$

其中, $\hat{e}_{m,k} = y(j) - X(j)^T \hat{\beta}_{m,k}, j=1, \dots, m+k; \{U_{m,k}(i): 1 \leq i \leq m+T_m\}$ 为独立于因变量 $\{y(i): 1 \leq i \leq m+T_m\}$ 和回归变量 $\{X(i): 1 \leq i \leq m+T_m\}$ 的独立同分布随机变量, 满足 $P(U_{m,k}(1) = j) = 1/(m+k), j=1, \dots, m+k$.

其次, 检验统计量 $\sup_{1 \leq l < T_m+1} |\Gamma(m, l, \gamma)|$ 依赖于还未观测的回归变量 $X(l), l \geq k$, 更确切地说是该统计量依赖于 $\sum_{i=m+1}^{m+l} X(i)^T$, 而当 $l > k$ 时该式包含未知的回归变量. 为了尽可能多地利用观测信息, 并且尽可能地接近原统计量, 采用 $c_1(m, k, l)$ 替代式 $\sum_{i=m+1}^{m+l} X(i)^T$, 其中 $c_1(m, k, l)$ 应满足

$$c_1(m, k, l)^T (1, 0, \dots, 0) = 1. \quad (7)$$

从而, 基于回归 Bootstrap 法的统计量为

$$\sup_{1 \leq l < T_m+1} |\tilde{F}(m, l, \gamma)|, \quad (8)$$

其中,

$$\begin{aligned} & \tilde{F}(m, l, \gamma)(e(1), \dots, e(m+l)) \\ &= \left(\sum_{i=m+1}^{m+l} e(i) - c_1(m, k, l)^T C_m^{-1} \sum_{j=1}^m X(j) e(j) \right) / g(m, l, \gamma). \end{aligned}$$

$m+k$ 时刻的临界值可以从如下分布函数的分位点中得到

$$\begin{aligned} & F_{m,k}^{(R^*)}(\alpha) = \\ & P_{m,k}^* \left[\frac{1}{\hat{\sigma}_{m,k}^{(R^*)}} \sup_{1 \leq l < T_m+1} |\hat{F}(m, l, \gamma)(e_{m,k}^*(1), \dots, e_{m,k}^*(m+l))| \leq \alpha \right], \end{aligned} \quad (9)$$

其中,

$$\left(\hat{\sigma}_{m,k}^{(R^*)} \right)^2 = \frac{1}{m-p} \sum_{i=1}^m \left(\sum_{j=1}^m X(j) e_{m,k}^*(j) \right) \sum_{j=1}^m X(j) e_{m,k}^*(j).$$

当检测到混凝土坝结构性态转异时, 由于诊断判断中检测的转异点与真实转异点之间存在检验延迟 τ , 从而接下来的主要任务就是确定转异的真实时刻, 该结果也为寻找转异原因及其影响奠定基础. 此时, 检验的样本为 $\{(y(i), X(i)): m < i \leq m+n\}$, 且在该诊断样本中仅且存在一个转异点, 从而简化了确

定转异时刻的问题^[18].

对于混凝土坝线性回归模型, 此时, 确定裂缝转异时刻的问题可以表示为: 对于某一时刻 $t, m+1 \leq t \leq m+n, \beta_t \neq \beta_0$, 其中 β_0 针对于在线诊断样本 $\{(y(i), X(i)): m < i < t\}$, β_t 针对于在线诊断样本 $\{(y(i), X(i)): t \leq i \leq m+n\}$.

定义如下统计量

$$\begin{aligned} & F_\varepsilon(m, n, q) \\ &= \sum_{i=m+1}^{q-1} (y(i) - X(i)^T \hat{\beta}_0)^2 + \sum_{j=q}^{m+n} (y(i) - X(i)^T \hat{\beta}_t)^2, \end{aligned} \quad (10)$$

其中, $\hat{\beta}_0$ 和 $\hat{\beta}_t$ 分别为模型回归系数 β 基于裂缝在线诊断样本 $\{(y(i), X(i)): m < i \leq q-1\}$ 和 $\{(y(i), X(i)): q \leq i \leq m+n\}$ 的估计.

从而, 转异时刻 t 满足

$$F_\varepsilon(m, n, t) = \min_{m < q \leq m+n} F_\varepsilon(m, n, q). \quad (11)$$

2.2 基于偏最小二乘回归的混凝土坝安全监控模型

设有一个因变量 y , p 个自变量 $\{x_1, x_2, \dots, x_p\}$, 观测样本点数为 n , 由此便构成了 2 张自变量与因变量的数据表 $X = [x_1, x_2, \dots, x_p]_{n \times p}$ 和 $Y = [y]_{n \times 1}$. 偏最小二乘回归分别在 X 与 Y 中提取成分 t_1 和 u_1 , 并要求 t_1 和 u_1 尽可能大地携带各自数据表中的变异信息, 并且 t_1 和 u_1 的相关程度达到最大, 同时自变量成分 t_1 对因变量成分 u_1 有着最强的解释能力^[19]. 提取第一个自变量成分 t_1 和因变量成分 u_1 后, 偏最小二乘回归分别实施 X 对 t_1 的回归以及 Y 对 t_1 的回归, 如果回归方程的精度已达到满意程度, 则算法停止, 否则利用 X 被 t_1 解释后的残余信息以及 Y 被 t_1 解释后的残余信息进行第二轮的成分提取, 如此往复, 根据交叉有效性检验, 直到能达到一个较满意的精度为止. 若最终对 X 共提取了 m 个成分 $t_1, t_2, \dots, t_m$, 则实施 y 对 $t_1, t_2, \dots, t_m$ 的回归, 然后再转化成 y 关于原自变量 $x_1, x_2, \dots, x_p$ 的回归方程.

通常偏最小二乘回归方程并不需要提取全部的成分, 而只需选取前 m 个成分 ($m < A, A = \text{秩}(X)$), 即可得到一个精度较高与拟合效果较好的模型, 对于所提取的成分个数, 一般可以通过交叉有效性检验来

确定. 对于实测变量 $y_i (i=1, 2, \dots, n)$, 计 $\hat{y}_{hi}$ 是从全部样本点提取成分 $t_1, \dots, t_h$ 建模后, 第 i 个样本点的拟合值, 则 y_i 的误差平方和为 SS_h

$$SS_h = \sum_{i=1}^n \left(y_i - \hat{y}_{hi} \right)^2. \quad (12)$$

把除去某个样本点 i 的所有样本集合作为一个样本, 用这部分样本并使用 h 个成分拟合一个回归方程; 然后把排除的样本点 i 代入前面拟合的回归方程, 得到 y_i 在样本点 i 上的拟合值 $\hat{y}_{h(-i)}$, 则定义 y_i 的预测误差平方和为 $PRESS_h$

$$PRESS_h = \sum_{i=1}^n \left(y_i - \hat{y}_{h(-i)} \right)^2. \quad (13)$$

若回归方程的稳健性不好, 误差很大, 就会加大 $PRESS_h$ 值. 通常情况下, $PRESS_h$ 大于 SS_h , 而 SS_h 小于 SS_{h-1} , SS_{h-1} 是用全部样本点拟合的具有 $h-1$ 个成分的方程的拟合误差, 如果 h 个成分回归方程的含扰动误差能够在一定程度上小于 $h-1$ 个成分回归方程的拟合误差, 则认为增加一个成分 t_h 后, 预测的精度能够明显提高. 因此希望 $PRESS_h/SS_{h-1}$ 的值越小越好, 但是并没有一个确定的值, 通常情况下指定当 $PRESS_h/SS_{h-1} \leq 0.95^2$ 时, 则认为增加成分 t_h 是有益的.

另一种等价的定义称为交叉有效性, 成分 t_h 的交叉有效性定义为

$$Q_h^2 = 1 - \frac{PRESS_h}{SS_{h-1}}. \quad (14)$$

当 $Q_h^2 \geq 0.0975$ 时, t_h 成分的贡献是显著的, 因此在偏最小二乘建模的每一步计算结束之后, 均需要进行交叉有效性检验, 若在第 h 步有 $Q_h^2 < 0.0975$, 则认为没有必要再提取下一个成分, 可以停止提取成分的计算.

偏最小二乘回归通过引入变量投影重要性指标 (VIP) 来测度自变量对因变量的解释能力. x_j (第 j 个自变量) 对 y 的解释是通过 t_h 传递的, 若 t_h 对 y 的解释能力很强, 而 x_j 在构造 t_h 时具有重要作用, 则 x_j 对 y 的解释能力被视为很大, 可定义为

$$VIP_j = \sqrt{\frac{p}{Rd(y; t_1, t_2, \dots, t_m)} \sum_{h=1}^m Rd(y; t_h) w_{hj}^2}. \quad (15)$$

其中, w_{hj} 是轴 w_h 的第 j 个分量, 用于衡量 x_j 对构造 t_h

主成分的贡献大小, $Rd(y; t_h)$ 和 $Rd(y; t_1, t_2, \dots, t_m)$ 分别是 y 由 t_h 和 $t_1, t_2, \dots, t_m$ 所解释的变异精度, 分别代表了 t_h 对 y 的解释能力和 $t_1, t_2, \dots, t_m$ 对 y 的累计解释能力, 且存在

$$Rd(y; t_h) = r^2(y, t_h), \quad (16)$$

$$Rd(y; t_1, t_2, \dots, t_m) = \sum_{h=1}^m Rd(y, t_h), \quad (17)$$

其中, $r(y, t_h)$ 是因变量 y 和主成分 t_h 的相关系数.

3 工程实例

某低水头河床式水电站, 水库正常蓄水位 310.00 m, 水库总库容 14530 万 m^3 , 具有日调节性能. 电站总装机容量 150 MW, 最大水头 16.8 m, 引用流量 1148 m^3/s , 2008 年 8 月 4 台机组全部投产发电. 水电站混凝土重力坝, 最大坝高为 30 m, 坝轴线长为 580 m. 现以坝顶 A06 水平位移监测点 2012 年 7 月至 2013 年 12 月间的水平位移监测数据 (图 1) 为例, 对本文所建的混凝土坝小样本安全监控模型进行分析.

为评价混凝土坝结构性能转异诊断对其安全监控模型的影响, 现将前 30 组监测数据建立整体的偏最小二乘回归模型, 并对后 5 组监测数据进行预测值分析. 所采用的统计模型为^[1]

$$\hat{\delta} = a_0 + \sum_{i=1}^3 a_i H^i + b_1 T + c_1 \theta + c_2 \ln \theta,$$

其中, H 为上下游水位差; T 为观测当天观测时段的平均气温; $\theta = t/100$, t 为观测日至观测基准日的累计天数; $a_0, a_i (i=1, 2, 3), b_1, c_1, c_2$ 为水位、温度和时效对应的回归系数.

该模型共选用 6 项影响因子, 其中水压因子 3 项, 温度因子 1 项, 时效因子 2 项, 6 个自变量的相关系数见表 1.

根据表 1 可见, 自变量 H^1, H^2, H^3 两两之间的相关系数均大于 0.99, 自变量 θ 与 $\ln \theta$ 之间的相关系数为 0.998, 相关程度均非常高; H^3 与 T 负相关, 相关系数为 -0.771, 相关程度较高; 此外, H^2 与 T, H^1 与 T 之间的相关系数分别为 -0.749 和 -0.725, 相关程度也比较高. 由此可见, 水压因子之间存在很强的相关性, 时效因子之间也均存在很强的相关性, 水压因子与温度因子之间存在较强的相关性.

根据交叉有效性检验指标 $Q_h^2 \geq 0.0975$ 进行成分

表 1 自变量间的相关系数表(30 组数据)

自变量	H^1	H^2	H^3	T	θ	$\ln\theta$
H^1	1.000	0.998	0.994	-0.725	-0.438	-0.422
H^2	0.998	1.000	0.998	-0.749	-0.445	-0.429
H^3	0.994	0.998	1.000	-0.771	-0.451	-0.434
T	-0.725	-0.749	-0.771	1.000	0.315	0.303
θ	-0.438	-0.445	-0.451	0.315	1.000	0.998
$\ln\theta$	-0.422	-0.429	-0.434	0.303	0.998	1.000

提取, 交叉有效性检验结果见表 2; 采用偏最小二乘回归模型所拟合的 A06 监测点水平位移实测值、拟合值过程线见图 1; 所得回归方程的回归系数、复相关系数 R 以及剩余标准差 S 见表 3; 最后 5 组监测数据的预测值见表 4; 各个自变量因子的 VIP_j 分布图见图 2.

由表 2 可知, 当 $h=3$ 时, $Q_3^2 < 0.0975$, 因此, 只需要提取前 2 个主成分就能较好地解释因变量. 由图 1 可见, 拟合效果较好, 但是只拟合出了整体的变化趋势, 过程线突变处未得到较好的拟合, 部分样本点的残差相对较大, 如第 5~9 个样本点. 由表 3 可知该

回归方程的复相关系数虽然达到了 0.876, 但是剩余标准差也较大, 其值为 2.040 mm, 主要是由于部分样本点的残差较大导致的, 这与由图 1 得出的结论相符. 由表 4 可知, 对于后 5 组监测数据, 预测值的相对误差分别为 23.6%, 33.8%, 45.5%, 52.3%和 59.3%,

表 2 交叉有效性检验表(30 组数据)

成分个数	Q_h^2	临界值
1	0.608	0.0975
2	0.341	0.0975
3	-0.018	0.0975

图 1 监测点水平位移实测值、拟合值过程线(30 组数据)

表 3 偏最小二乘回归模型回归结果汇总表(30 组数据)

参数	a_0	a_1	a_2	a_3	b_1	c_1	c_2	R	S (mm)
结果	118.4513	0.1115	0.0038	0.0002	0.0257	-1.1682	-30.5641	0.876	2.040

表 4 监测点水平位移预测值与实测值对比表(30 组数据)

监测序号	实测值(mm)	预测值(mm)	相对误差
31	-10.28	-12.71	23.6%
32	-9.55	-12.77	33.8%
33	-8.81	-12.81	45.5%
34	-8.64	-13.15	52.3%
35	-8.46	-13.48	59.3%

图2 各因子重要性指标分布图(30组数据)

1~6 分别代表 $H, H^2, H^3, T, \theta, \ln \theta$

预测效果并不理想. 由图2可见, 时效因子对监测点水平位移的影响最大, 其 VIP_j 值均大于1, 分别为1.3477和1.3545; 其次是水压因子, 其 VIP_j 值比较接近, 分别为0.7903, 0.7975和0.8022, 而温度因子的影响最小, 其 VIP_j 值为0.6672.

根据以上结果可以看出, 利用30组监测数据所得出的回归方程复相关系数较高, 剩余标准差较大, 预测效果并不理想. 该模型的 VIP_j 值表明: 时效是影响该点水平位移的主要因素.

采用本文所建立的混凝土坝结构性态转异诊断的 Bootstrap 法, 依据监测数据进行转异诊断分析, 诊断结果如图3所示. 根据实际转异点的确定方法所得到的转异点对应于监测序号12.

因此, 根据2013年1月至2013年10月的18组监测数据, 进行偏最小二乘回归分析, 并对后5组监测数据进行预测值分析, 分析结果见表5~7, 图4和5.

由表5可知, 当 $h=3$ 时, $Q_3^2 < 0.0975$, 只需提取前2个主成分即可. 由图4可见, 与之前相比, 总体的拟合效果更好, 不存在部分样本点残差过大的情况. 由表6可知该回归方程的复相关系数较高, 其值为0.946, 剩余标准差较小, 其值为0.299 mm, 这与由图4得出的结论是相符的. 由表7可知, 对于后5

组监测数据, 预测值的相对误差分别为6.1%, 4.8%, 3.7%, 2.5%和1.3%, 预测效果较为理想. 由图4可见, 水压因子对观测点水平位移的影响最大, 其 VIP_j 值均大于1, 分别为1.1527, 1.1446和1.1352; 其次是温度因子, 其 VIP_j 值为0.8461; 而时效因子的影响相对较小, 其 VIP_j 值比较接近, 分别为0.8297和0.8175.

根据以上结果可以看出, 利用18组监测数据所得出的回归方程复相关系数较高, 剩余标准差较小, 且预测效果较为理想. 该模型的 VIP_j 值表明水位是影响该点水平位移的主要因素. 为了进一步评价该模型的精度, 还可以通过主成分对自变量 X 以及因变量 Y 的解释能力来判断, 主成分对各个自变量及因变量的累计解释能力见表8.

从表8可以看出, 提取第一个主成分 t_1 后, t_1 对自变量中水压因子的解释能力已经较高; 提取第二个主成分 t_2 后, 主成分对时效因子的解释能力得到了明显的提高. 主成分 t_1, t_2 对自变量 X 的累计解释能力为95.49%, 主成分 t_1, t_2 对因变量 Y 的累计解释能力为89.52%. 因此, 提取的两个主成分可以解释原自变量系统中95.49%的变异信息、可以解释原因变量系统中89.52%的变异信息, 由此可见, 所提取的主成分对自变量与因变量的解释能力均较高.

图3 依据监测数据的转异诊断结果

表 5 交叉有效性检验表(18 组数据)

成分个数	Q_n^2	临界值
1	0.740	0.0975
2	0.460	0.0975
3	-0.238	0.0975

根据以上结果可以看出, 通过转异诊断, 在新的时间段内利用 18 组监测数据所得出的回归方程的复相关系数从之前的 0.876 升高至 0.946, 剩余标准差从之前的 2.040 减小至 0.299 mm, 后 5 组监测数据预测值的相对误差明显降低. 因此, 回归方程的精度与预测能力均得到了较为明显的提高, 可以认为, 该模

型的可信度更高. 根据该分析结果可以发现: 水压因子是影响该点水平位移的主要因素, 温度因子对该点的水平位移也有一定程度的影响, 而时效因子对该点水平位移的影响程度则较小, 这与现有的研究成果^[1]是比较符合的.

4 结论

1) 在样本点数较少且自变量相关性较强的情况下, 偏最小二乘回归可以通过成分提取的方式, 充分地挖掘自变量与因变量所携带的信息并消除了自变量之间的多重相关性从而得出精度较好的回归方程.

图 4 监测点水平位移实测值、拟合值过程线(18 组数据)

表 6 偏最小二乘回归模型回归结果汇总表(18 组数据)

参数	a_0	a_1	a_2	a_3	b_1	c_1	c_2	R	S (mm)
结果	-21.0480	0.1779	0.0057	0.0002	-0.0107	0.0620	1.4836	0.946	0.299

表 7 监测点水平位移预测值与实测值对比表(18 组数据)

监测序号	实测值(mm)	预测值(mm)	相对误差
31	-10.28	-9.66	6.1%
32	-9.55	-9.09	4.8%
33	-8.81	-8.49	3.7%
34	-8.64	-8.42	2.5%
35	-8.46	-8.35	1.3%

图 5 各因子重要性指标分布图(18 组数据)

1~6 分别代表 $H, H^2, H^3, T, \theta, \ln \theta$

表 8 主成分累计解释能力表

成分	t_1	t_2
$Rd(H; t_h)$	0.8920	0.9937
$Rd(H^2; t_h)$	0.9164	0.9955
$Rd(H^3; t_h)$	0.9367	0.9956
$Rd(T; t_h)$	0.7146	0.7804
$Rd(\theta; t_h)$	0.6871	0.9814
$Rd(\ln \theta; t_h)$	0.6715	0.9830
$Rd(X; t_h)$	0.8030	0.9549
$Rd(Y; t_h)$	0.7966	0.8952

2) 工程实例也说明在考虑混凝土结构性态转异时, 采用偏最小二乘回归在处理大坝安全监控领域中的小样本数据问题具有较明显的优势. 在自变量间存在较强相关性时, 综合转异诊断方法和偏最小二乘回归分析的混凝土坝小样本安全监控模型的精度更好、预测能力更强、解释更加合理.

3) 本文由于所得的气温资料较少(只有观测时段的平均气温)且监测序列较短, 所以在建模时未考虑水位与温度的滞后效应, 因此, 考虑滞后效应所建立的混凝土坝小样本安全监控模型还有待于进一步研究.

参考文献

- 1 顾冲时, 吴中如. 大坝与坝基安全监控理论和方法及其应用. 南京: 河海大学出版社, 2006
- 2 杨杰, 吴中如. 观测数据拟合分析中的多重共线性问题. 四川大学学报(工程科学版), 2005, 37: 19–24
- 3 吴道闻, 韩大建. 因子相关性对大坝监测资料回归分析的影响. 大坝观测与土工测试, 1996, 20: 23–25
- 4 李智录. 大坝安全监控统计模型研究. 博士学位论文. 西安: 西安理工大学, 2006
- 5 王惠文. 偏最小二乘回归方法及其应用. 北京: 国防工业出版社, 1999
- 6 邓念武. 偏最小二乘回归在大坝位移资料分析中的应用. 大坝观测与土工测试, 2001, 25: 16–18
- 7 徐洪钟, 吴中如. 偏最小二乘回归在大坝安全监控中的应用. 大坝观测与土工测试, 2001, 25: 22–24
- 8 仲琳. 碾压混凝土坝安全监控模型及其应用研究. 硕士学位论文. 南京: 河海大学, 2005
- 9 杨杰. 大坝安全监控中若干不确定性问题的分析方法与应用研究. 博士学位论文. 南京: 河海大学, 2006
- 10 赖道平. 地质缺陷对混凝土坝结构性态演变和转异的影响研究. 博士学位论文. 南京: 河海大学, 2005
- 11 景继. 混凝土坝裂缝开度预测及警戒值拟定方法研究. 博士学位论文. 南京: 河海大学, 2010
- 12 Bao T F, Qin D, Zhou X W, et al. Abnormality monitoring model of cracks in concrete dams. Sci China Tech Sci, 2011, 54: 1914–1922
- 13 Gu C S, Li Z C, Xu B. Abnormality diagnosis of cracks in the concrete dam based on dynamical structure mutation. Sci China Tech Sci, 2011, 54: 1930–1939
- 14 Li Z, Gu C, Wu Z. Nonparametric change point diagnosis method of concrete dam crack behavior abnormality. Math Prob Eng, 2013, 2013: 1–13
- 15 Kirch C. Bootstrapping sequential change-point tests. Sequential Anal, 2008, 27: 330–349
- 16 Horvath L, Huskova M, Kokoszka P, et al. Monitoring changes in linear models. J Stat Plan Infer, 2004, 126: 225–251
- 17 Hušková M, Kirch C. Bootstrapping sequential change-point tests for linear regression. Metrika, 2012, 75: 673–708
- 18 Aue A, Horváth L. Structural breaks in time series. J Time Ser Anal, 2013, 34: 1–16
- 19 张恒喜. 小样本多元数据分析方法及应用. 西安: 西北工业大学出版社, 2002

Safety monitoring model with small samples of concrete dam

LI ZhanChao^{1,2}, ZHANG HuiLi¹, LIU XingYang¹ & QIN Dong²

¹ College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling 712100, China;

² State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

Study of concrete dam safety monitoring model is always a hot and difficult problem in the field of dam safety monitoring. In this paper, the evolution of concrete dam is regarded as a dynamic process. In view of less monitoring data and strong correlation among variable factors, a safety monitoring model with small samples of concrete dam is established through a combination of Bootstrap method of concrete dam structural abnormality diagnosis and partial least squares regression analysis method. And this model is applied to the analysis of horizontal displacement of a concrete gravity dam. Results show that the fitting precision of this model is higher, the prediction effect is better and the explanation is more reasonable. This contribution provides a scientific basis for concrete dam safety monitoring analysis.

concrete dam, safety monitoring model, small sample, abnormality diagnosis, partial least squares regression

doi: 10.1360/N092014-00182