

基于统计模型与混沌理论的大坝安全监测混合预测模型

包腾飞,吴中如,顾冲时

(河海大学水利水电工程学院,江苏 南京 210098)

摘要 为解决传统统计模型在大坝安全监测领域应用中由于缺少对残差的分析而存在拟合精度高但预测效果欠佳的问题,在传统统计模型中增加残差预测项,并应用混沌理论对残差预测项的计算方法进行探讨,从而提出了一种新的混合预测模型,同时结合实测数据对此模型进行了检验.结果表明,新模型可提高预测精度并具有一定的实用性.

关键词 大坝安全监测 统计模型 混沌理论 混合预测模型

中图分类号 :TV698.1 文献标识码 :A 文章编号 :1000-1980(2003)05-0534-05

在大坝安全监控领域,统计模型得到了广泛应用.建立统计模型时,首先以水压、温度和时效为主要影响因素来建立大坝观测资料的回归模型,其形式为

$$\delta(t) = \delta_H(t) + \delta_T(t) + \delta_\theta(t) + x(t)$$
 (1)

式中 : $\delta(t)$ ——观测效应量 ; $\delta_H(t)$ ——水压分量 ; $\delta_T(t)$ ——温度分量 ; $\delta_\theta(t)$ ——时效分量 ; $x(t)$ ——残差.然后利用回归分析等方法确定 $\delta_H(t)$, $\delta_T(t)$ 和 $\delta_\theta(t)$ 的因子,并以

$$\hat{\delta}(t+T) = \delta_H(t+T) + \delta_T(t+T) + \delta_\theta(t+T)$$
 (2)

作为下一时刻($t+T$)的预测模型.然而,这种预测模型只考虑了 3 个分量因子而缺少对残差 $x(t)$ 的分析,因此存在历史资料拟合好而实际外推预测效果欠佳的问题.其原因是,预测模型将残差 $x(t)$ 看成是来自系统之外的随机因素作用的结果,在预测中不予考虑.实际上,残差 $x(t)$ 中不仅包含着外界干扰、测量误差等随机因素作用的结果,而且还包含着一些有用的信息.近年来的研究表明,残差序列 $\{x(t)\}$ 中存在着混沌现象^[1],它的演化不仅受外界随机因素的影响,而且受系统内在特性的影响.因此,为提高预测效果,预测模型中有必要考虑残差预测项 $x(t+T)$.于是预测模型可表示为

$$\hat{\delta}(t+T) = \delta_H(t+T) + \delta_T(t+T) + \delta_\theta(t+T) + x(t+T)$$
 (3)

本文着重论述 $x(t+T)$ 的计算方法.根据混沌学原理,预测 $x(t+T)$ 时必须将残差时间序列 $\{x(t)\}$ 嵌入到重构的高维相空间中,求出序列混沌吸引子的分维数 D_2 ,从而计算出用于预测的相空间维数 $d(d > D_2)$,然后重构用于预测的 d 维相空间,在此相空间中对 $x(t+T)$ 作出预测.

1 残差序列的相空间重构

随着混沌学的发展,人们对时间序列的复杂性有了新的认识.混沌学认为,即使是完全确定的非线性模型,在长时间演化后结果可能是类似随机的.因此,长时间预测是不准确或不可能的,但短期预测则是可行的.对于大坝监测效应量的预测来说,利用大量的历史观测数据对下一测值作出预测属短期预测行为,因此应用混沌理论是可以实现的.

残差序列为一时间序列,根据混沌理论,任一时间序列都包含着系统参与演化的大量信息,而系统某一状态变量随时间的变化,可以用一时间序列来表示^[1,2].虽然事先不知道所描述系统的相空间的真实维数,但总可以用时间序列的数据,通过下式所述形式支起一个 n 维空间 R^n .

$$X(t) = [x(t), x(t+\tau), \cdots, x(t+(n-1)\tau)]^T$$
 (4)

收稿日期 2002-09-16
基金项目 国家自然科学基金资助项目(50139030)
作者简介 包腾飞(1974—),男,湖北黄冈人,博士研究生,主要从事水工结构安全监控方面的研究.

式中: τ ——延迟时间, $\tau = h\Delta t$ ($h = 1, 2, \dots$); Δt ——时间序列的时间间隔. 这种利用数据序列构造成的 n 维空间通常叫做“嵌入空间”. 设 $x(t_1), x(t_2), \dots, x(t_m + (n - 1)\tau)$ 为监测所得的残差时间序列, 取某一固定时间间隔 τ , 将原有序列拓展成 n 维相空间 R^n 的一个相型为

$$X(t_i) = [x(t_i), x(t_i + \tau), \dots, x(t_i + (n - 1)\tau)]^T \quad (i = 1, 2, \dots, m) \tag{5}$$

这里 $X(t_i)$ 为 n 维相空间的一个相点. $\{X(t_i)\}$ 则构成了 n 维相空间里的一个相型, 表征了该系统的某一状态, 而这 m 个相点的连线则构成了点在相空间的轨道, 表征了系统状态随时间的演化.

2 混沌吸引子分维数的计算

维数是空间和客体的重要特征量, 吸引子的维数说明刻画该吸引子所必需的信息量, 描述了该动力系统的复杂程度或系统的自由度. 一个非线性系统处于混沌状态, 其相空间的运动轨迹是十分复杂的, 而比较适合于用实验资料时间序列来计算的是关联维数 D_2 ^[3,4]. 其主要计算步骤为:

- a. 重构如上一节所示的 n 维空间.
- b. 计算关联函数

$$C(\epsilon) = \frac{1}{m^2} \sum_{\substack{i,j \\ i \neq j}}^m \theta(\epsilon - \|X(t_i) - X(t_j)\|) \tag{6}$$

式中: $\theta(\alpha)$ ——Heaviside 函数, $\alpha > 0$ 时 $\theta(\alpha) = 1$, $\alpha \leq 0$ 时 $\theta(\alpha) = 0$; $\|\cdot\|$ ——Euclidean 模(下同); $\|X(t_i) - X(t_j)\|$ ——相点 $X(t_i)$ 到 $X(t_j)$ 的距离; m ——相点数; $C(\epsilon)$ —— n 维相空间两点间距离小于 ϵ 的概率, 它刻画了相对于相空间某参考点 $X(t_j)$ 在 ϵ 内的聚集程度. 为保证 $0 \leq C(\epsilon) \leq 1$, 计算时应适当调整 ϵ 的取值范围.

- c. 对于 ϵ 的某个适当的范围, 吸引子的维数 D_2 应满足对数线性关系

$$D_2(n) = \frac{\ln C(\epsilon)}{\ln \epsilon} \tag{7}$$

- d. 增高“嵌入空间”的嵌入维数 n , 重复 a~c 3 个步骤, 直到 n 达到某一值 n_c 相应的维数估计值 $D_2(n)$ 不随 n 的增长而发生有意义的变化(即保持在给定误差范围内)为止, 即

$$D_2 = D_2(n_c) \approx D_2(n_c + 1) \approx D_2(n_c + 2) \approx \dots \tag{8}$$

这里 D_2 就是所求的混沌吸引子的分维数.

3 用于预测 $x(t + T)$ 的 d 维相空间重构及预测方法

预测时必须选择合适的相空间维数 d . 对于分维数为 D_2 的混沌吸引子, 如果 d 过小, 相空间无法容纳动力系统的吸引子, 因而无法全面展示系统的动力特性. 如果 d 过大, 除计算工作量大外, 还将减小可使用数据长度, 使所建相空间的相点变得过于稀疏, 相应地增大了预测误差. 根据混沌理论, 只要 d 在 $\text{int}(D_2 + 1) \leq d \leq \text{int}(2D_2 + 1)$ 中选取(int 为取整函数), 则会取到较好的预测效果.

对于时间序列 $\{x(t_i)\}$ ($i = 1, 2, \dots, N$), 将其分为 2 段: $x(t_1) \sim x(t_m)$ 和 $x(t_{m+1}) \sim x(t_N)$. 后者很短, 仅供检验用(正式预测时则不分段). 前者按如下排列构成相型.

$$X(t_i) = [x(t_i), x(t_i - \tau), x(t_i - 2\tau), \dots, x(t_i - (d - 1)\tau)]^T \quad (i = m, m - 1, \dots, 1 + (d - 1)\tau) \tag{9}$$

式中: τ ——延迟时间; $X(t_i)$ —— d 维相空间里的一个相点.

由于提前预测的时间 T 一般与重构相空间里所用的延迟时间 τ 和时间序列间隔 Δt 系同量级, 它们相对于时间序列的长度是很小的, 故有如下基本假设: 给定时间的结构, 在预测区域内不变^[5].

设参考态为 $X(t_m)$, 其最近邻态为 $X_{nbt}(t_i)$. 最近邻态定义为: $\{X(t_j)\}$ ($j = 1, 2, \dots, (m - 1)$) 中, 使得 $\|X(t_m) - X(t_j)\|$ 达到最小的那个相点为 $X_{nbt}(t_i)$.

参考态 $X(t_m)$ 经预测时间 T 后, 演化为 $X(t_m + T)$. 当提前预测时间 $T \leq \tau$ 时, 显然 $X(t_m + T)$ 的 d 个分量中只有第一个分量 $x(t_m + T)$ 是未知的, 而其余 $(d - 1)$ 个分量都是已知的, 即 $x(t_m + T)$ 为预测对象.

考虑到 $X(t_m)$ 的最近邻态 $X_{nbt}(t_i)$ 经预测时间 T 后, 演化为 $X_{nbt}(t_i + T)$, 由最近邻态式可知, 态 $X_{nbt}(t_i)$ 与 $X(t_m)$ 相差很小, 它们经演化后由于演化时间 T 很短, 所以态 $X_{nbt}(t_i + T)$ 与 $X(t_m + T)$ 仍然相差很小, 可认为演化后的距离仍然等于演化前的距离. 可设演化前和演化后的距离分别为 L_0 和 L_T , 其形式分别为

$$L_0 = \|X(t_m) - X_{nbt}(t_i)\| \tag{10}$$

$$L_T = \|X(t_m + T) - X_{nbt}(t_i + T)\| \tag{11}$$

基于如上假设,有

$$L_0 = L_T \tag{12}$$

即

$$\|X(t_m) - X_{nbt}(t_i)\| = \|X(t_m + T) - X_{nbt}(t_i + T)\| \tag{13}$$

式(12)或式(13)为相空间近邻等距假设.

根据式(9)所示相点的状态分布,式(13)便可直接用于预测.例如,假定已找到 $X(t_m)$ 的最近邻的相点 $X(t_b)$,那么在 t_m 时刻, $X(t_m)$ 与 $X(t_b)$ 之间的距离为

$$L_0 = \{[x(t_m) - x(t_b)]^2 + [x(t_m - \tau) - x(t_b - \tau)]^2 + \cdots + [x(t_m - (d - 1)\tau) - x(t_b - (d - 1)\tau)]^2\}^{1/2} \tag{14}$$

在 $t_m + T$ 时刻, $X(t_m + T)$ 与 $X(t_b + T)$ 之间的距离为

$$L_T = \{[x(t_m + T) - x(t_b + T)]^2 + [x(t_m + T - \tau) - x(t_b + T - \tau)]^2 + \cdots + [x(t_m + T - (d - 1)\tau) - x(t_b + T - (d - 1)\tau)]^2\}^{1/2} \tag{15}$$

在式(14)中, $X(t_m)$ 与 $X(t_b)$ 中各分量等均为已知量.在式(15)中,由于 τ 不会小于 T ,因此除 $x(t_m + T)$ 为预测值外,其余各值均是已知量.

由假设(12),利用式(14)和式(15),可得

$$\begin{aligned} & \{[x_p(t_m + T) - x(t_b + T)]^2 + [x(t_m + T - \tau) - x(t_b + T - \tau)]^2 + \cdots + \\ & [x(t_m + T - (d - 1)\tau) - x(t_b + T - (d - 1)\tau)]^2\}^{1/2} = \\ & \{[x(t_m) - x(t_b)]^2 + [x(t_m - \tau) - x(t_b - \tau)]^2 + \cdots + \\ & [x(t_m - (d - 1)\tau) - x(t_b - (d - 1)\tau)]^2\}^{1/2} \end{aligned} \tag{16}$$

式中 $x_p(t_m + T)$ 即为 $x(t_m + T)$ 的预测值.

4 计算实例

采用浙江某水电站大坝视准线实测资料进行试验.该坝为混凝土宽缝重力坝,坝顶高程 115.0 m,最大坝高 105.0 m,坝顶全长 466.5 m,分为 26 个坝段,编号为 0 号 ~ 25 号.该电站大坝采用视准线观测坝顶水平位移,从右至左 2 号 ~ 20 号,22 号,23 号坝段共 21 个测点.本文以 7 号坝段视准线测点的观测数据为例.该测点的观测数据序列起止时间分别为 1974 年 1 月和 2001 年 11 月.将观测资料分为 1974 年 1 月 ~ 1998 年 12 月、1999 年 1 月 ~ 2001 年 11 月 2 段.对前一段观测资料建立统计模型的回归方程,确定水平位移影响因子,分离出残差并用于计算分维数 D_2 和残差预测空间的维数 d ;后一段观测数据用于检验预测模型式(3).统计模型的回归方程形式为^[6]

$$\delta = \delta_H + \delta_T + \delta_\theta = a_0 + \sum_{i=1}^3 a_i (H^i - H_0^i) + \sum_{i=1}^6 \left[b_{1i} \left(\sin \frac{2\pi it}{365} - \sin \frac{2\pi i t_0}{365} \right) + b_{2i} \left(\cos \frac{2\pi it}{365} - \cos \frac{2\pi i t_0}{365} \right) \right] + c_1 (\theta - \theta_0) + c_2 (\ln \theta - \ln \theta_0) \tag{17}$$

式中: $a_i, b_{1i}, b_{2i}, c_1, c_2$ ——回归系数; H ——观测位移时坝前水深,即库水位减去各坝段坝底高程; H_0 ——各时段初始水深; i ——从起始日至观测日的累计天数; i_0 ——从起始日至所计算时段起始日的天数; θ ——从测点起测日开始的天数乘以 0.01; θ_0 ——时段初始日至起测日的累计天数乘以 0.01.采用逐步回归分析法,逐步回归方程的回归系数、复相关系数(R)和显著性水平(F)见表 1,水平位移拟合过程线见图 1,残差序列见图 2.

表 1 回归成果

Table 1 Results of regression

a_1	a_2	a_3	b_{11}	b_{21}	b_{12}	b_{22}	b_{13}	b_{23}	b_{14}	b_{24}
5.80	-0.063 2	0.000 238	-0.741	-4.69	0.00	0.00	0.00	-0.288	0.125	0.00
b_{15}	b_{25}	b_{16}	b_{26}	c_1	c_2	a_0	R	S	F	
0.00	0.00	0.00	0.00	0.043 8	-0.142	2.01	0.966	1.05	694	

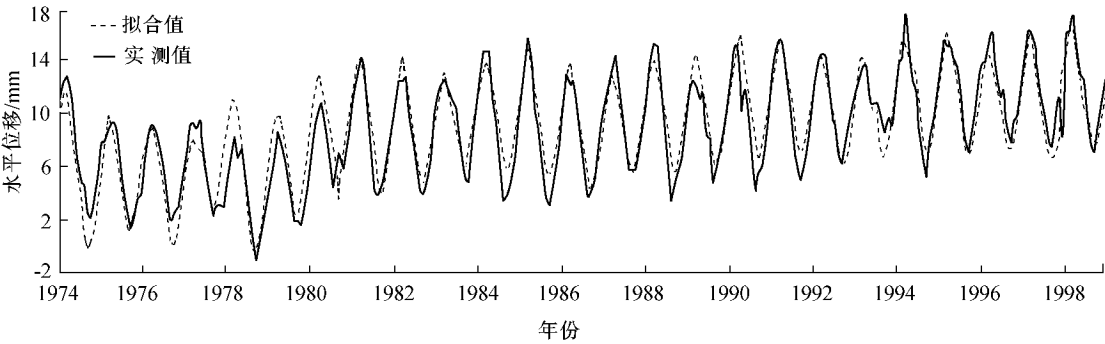


图 1 水平位移拟合过程线
Fig.1 Fitting graph of horizontal displacement

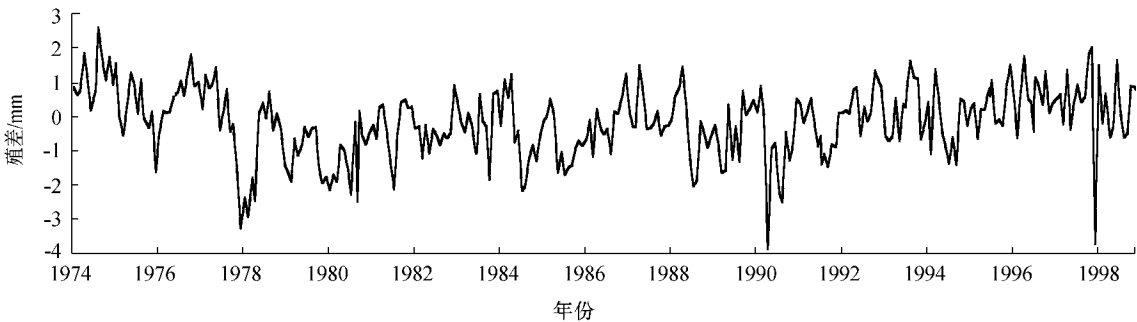


图 2 水平位移残差序列过程线
Fig.2 Residual error series of horizontal displacement

利用残差序列计算分维数 D_2 和预测空间的维数 d .
取延迟时间 $\tau = 1$, 提前预测的时间 $T = \tau$ (即预测下一测值). 由于事先不知‘嵌入空间’的维数 n 取多大才能容纳吸引子, 因此假定 $n = 1, 2, \dots, 10$. 将残差序列 $x(t_1), x(t_2), \dots, x(t_i), \dots$ 按式(5)构成相型, 根据式(6)计算 $C(\epsilon)$, 绘制 $\ln C(\epsilon) \sim \ln \epsilon$ 曲线, 计算该曲线下部 (ϵ 较小时) 直线段的斜率, 考察斜率 D 随 n 的变化趋势. 当 n 增大到某一值后, 斜率 D 不再发生显著变化, 此时的斜率 D 就是所求的分维数 D_2 . D 随 n 的变化曲线见图 3.

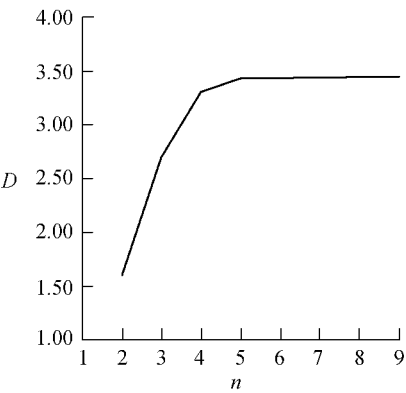


图 3 $D \sim n$ 关系曲线

当 $n \geq 4$ 时, 图 3 中的曲线趋于水平, 因此由图 3 可知 $D_2 = 3.44$. 由于 $\text{int}(D_2 + 1) \leq d \leq \text{int}(2D_2 + 1)$, 取 $d = 6$, 即用于残差预测的相空间(嵌入空间)的维数为 6, 按照式(9)方法就可以构造残差序列的预测空间.

利用式(3)对 1999 年 1 月 ~ 2001 年 11 月的水平位移进行预测. 具体做法是用式(17)回归的结果计算 $\delta_H(t+T) + \delta_T(t+T) + \delta_\theta(t+T)$, 用式(16)计算 $x(t+T)$. 水平位移预测结果见图 4.

Fig.3 Relation between D and n

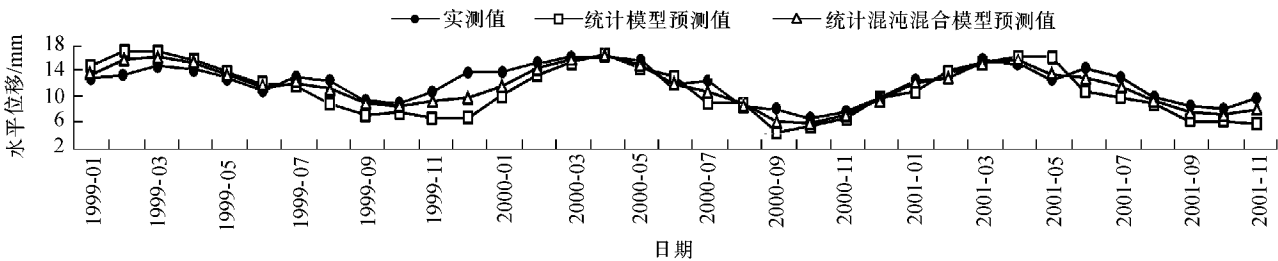


图 4 水平位移预测结果

Fig.4 Predicted results of horizontal displacement

由图 4 可以看出, 统计混沌混合新模型预测过程线比统计模型预测过程线更接近于实测过程线. 这说

明,统计混沌混合新模型减小了预测误差,提高了预测精度,其预测效果比统计模型预测效果更佳.同时也应该指出,新模型并没有完全消除误差,而建立完全消除误差的模型是不现实的.

5 结 论

- a. 通过对大坝水平位移统计模型分离出的残差时间序列所进行的相空间的重构,得出了分维数为一分数的结果,印证了残差序列存在着的混沌现象.
- b. 统计和混沌混合模型对水平位移的预测值比统计模型的预测值的精度有一定的提高,说明混合模型具有一定的实用价值.

参考文献:

[1] 徐洪钟.大坝动力系统的安全监控非线性分析模型研究[D].南京:河海大学,2001.
[2] 黄润生,周家斌,张仁涛,等.武汉地区雨强分维数随时间变化初探[J].武汉大学学报(自然科学版),1995,41(1):109—115.
[3] 徐平.大坝观测数据的吸引子分析[J].大坝观测与土工测试,1992,16(4):44—47.
[4] GRASSBERGER P,PROCACCIA J.Characterization of strange attractors[J].Phys Rev Lett,1983,50:346—349.
[5] 林振山.长期预报的相空间理论和模式[M].北京:气象出版社,1993.21—46.
[6] 吴中如,沈长松,阮焕祥.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,1990.36—49.

Statistic model and chaos theory-based hybrid forecasting model for dam safety monitoring

BAO Teng-fei, WU Zhong-ru, GU Chong-shi

(College of Water Conservancy and Hydropower Engineering, Hohai Univ., Nanjing 210098, China)

Abstract :The conventional statistic model is widely used in the field of dam safety monitoring. Owing to the lack of analysis of residual errors, the forecast results are not satisfactory even though the fitting precision of the model is high. Aiming at the above problem, a hybrid forecasting model is developed based on the chaos theory by introduction of the residual error predicting factor into the conventional model and analysis of its calculating method. The model is tested by field data, and the accuracy and applicability of the model are verified.

Key words :dam safety monitoring ; statistical model ; chaos theory ; hybrid forecasting model