

# 基于核主成分分析和粒子群优化 支持向量机的滑坡位移预测

彭 令<sup>1</sup> 牛瑞卿<sup>1</sup> 赵艳南<sup>1</sup> 邓清禄<sup>2</sup>

(1 中国地质大学(武汉)地球物理与空间信息学院,武汉市鲁磨路 388 号,430074)

(2 中国地质大学(武汉)工程学院,武汉市鲁磨路 388 号,430074)

**摘 要:**利用核主成分分析法对滑坡位移影响因子进行特征提取,以获得的主成分作为支持向量机的特征向量建立支持向量机模型,其中模型参数通过粒子群算法进行选择优化,构建出核主成分分析和粒子群优化支持向量机协同模型,对滑坡相对位移进行预测。预测结果的平均绝对误差和相对误差分别为 0.760 和 7.563%,与其他预测模型相比,其拟合和泛化能力最优,表明核主成分分析和粒子群优化支持向量机协同模型的预测结果与实际监测值具有很好的一致性。

**关键词:**滑坡;位移预测;核主成分分析;粒子群优化;支持向量机

**中图分类号:**P258; P642.22

滑坡是一个十分复杂的非线性动态系统,其变形演化过程受到外界影响因素的作用,当滑坡在变形演化的过程中,遭受外界影响因素的作用时,其变形位移多呈现出与之对应的变化特征,在实际的滑坡预测预报中,很容易对此类滑坡作出错误判断,因为变形过程的每一次突变都可能被误认为是斜坡演化已进入临滑阶段。因此,对滑坡位移与外界影响因素的综合响应分析和预测是处理滑坡变形曲线的有效手段<sup>[1]</sup>。文献[2-5]对滑坡位移预测进行了研究。但是在实践中存在各种缺陷,削弱了它的预测能力。

支持向量回归机(support vector regression, SVR)已被广泛应用于非线性时间序列预测<sup>[6-8]</sup>。但是在实际应用中,支持向量机的泛化性能对于模型参数的选择比较敏感,目前对其关键参数在理论上尚无直接选择法。粒子群算法(particle swarm optimization, PSO)是一种基于群体智能的全局寻优方法,适合于模型参数的选择和优化。同时,如果将所有滑坡影响因子作为 SVR 模型的输入特征,因其因子之间往往呈现较强的耦合性、非线性和信息冗余,不但使支持向量机模型过于复杂,而且会干扰模型预测策略,降低模型的拟合

和泛化能力。而核主成分分析(kernel principle component analysis, KPCA)能提取非线性特征信息,提高数据质量。因此,本文将核主成分分析、粒子群算法和支持向量机三者有机结合,建立 KPCA-PSO-SVR 协同优化模型,实现滑坡位移预测。

## 1 模型原理

### 1.1 核主成分分析

设原始空间  $R^d$  中有  $n$  个样本  $x_1, x_2, \dots, x_n$ , 由这  $n$  个样本构成的数据矩阵为  $X$ , 利用非线性映射函数将数据样本从原始空间映射到高维特征空间  $F$  中。首先,假设映射数据的均值为零,即:

$$\sum_{i=1}^n \varphi(x_i) = 0 \quad (1)$$

则特征空间  $F$  中映射数据的协方差矩阵表示为:

$$C^F = \frac{1}{n} \sum_{k=1}^n \varphi(x_k) \varphi(x_k)^T \quad (2)$$

对  $C^F$  进行特征矢量分析,其特征值为  $\lambda$ , 特征矢量为  $V$ , 则有  $\lambda V = C^F V$ 。定义矩阵  $K =$

收稿日期:2012-12-18。

项目来源:国家 973 计划资助项目(2011CB710601); 国家 863 计划资助项目(2012AA121303); 国家自然科学基金资助项目(40902099); 国土资源部三峡库区三期地质灾害防治重大科学研究资助项目(SXKY3-6-2)。

$[K_{i,j}]_{n \times n}, K_{i,j} = (\varphi(x_i) \cdot \varphi(x_j))$ ,使其可通过核函数来确定。设  $V^k$  是  $V$  的第  $k$  个特征矢量,对其进行归一化处理,即  $V^k V^k = 1$ ,则可得到原始空间中任意一样本  $x$  的映射数据  $\varphi(x)$  在特征矢量  $V^k$  上的投影为:

$$[V^k \cdot \varphi(x)] = \sum_{i=1}^n \alpha_i^k [\varphi(x_i) \cdot \varphi(x)] \quad (3)$$

上述推导是在假定映射数据的均值为零的前提下进行的,但通常这一假设条件难以成立,因此必须对映射数据中心化,引入  $\bar{K}$  来替代  $K$ :

$$\bar{K} = K - I_n K - K I_n + I_n K I_n \quad (4)$$

其中,  $I_n$  是  $n \times n$  的矩阵,且满足  $L_{i,j} = 1/n$  ( $n$  为样本数目)。由此得到数据样本第  $k$  维的非线性主成分为:

$$t_k = \bar{V}^F \cdot \varphi(x) = \sum_{i=1}^n \bar{\alpha}_i^k [\varphi(x_i) \cdot \varphi(x)] = \bar{\alpha}_i^k \cdot \sum_{i=1}^n \bar{\alpha}_i^k \bar{K}(x_i, x) \quad (5)$$

### 1.2 支持向量机

支持向量回归通过非线性变换将输入向量映射到高维特征空间,并构造最优决策函数,利用原空间的核函数取代高维特征空间中的点积运算,应用有限样本的学习训练,获得全局最优解<sup>[9]</sup>。SVR 估计函数为:

$$f(x) = W^T \varphi(x) + b \quad (6)$$

利用  $\epsilon$  不敏感损失函数将估计函数转化为优化问题:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} W^T W + C \sum_{i=1}^l \xi_i + C \sum_{i=1}^l \xi_i^* \quad (7)$$

约束条件为:

$$\begin{cases} W^T \varphi(x_i) + b - y_i \leq \epsilon + \xi_i \\ y_i - W^T \varphi(x_i) - b \leq \epsilon + \xi_i^*, \xi_i, \xi_i^* \geq 0, i = 1, \dots, l \end{cases} \quad (8)$$

式中,  $C$  为惩罚因子;  $\xi_i, \xi_i^*$  表示松弛因子;  $b$  为偏置量。

式(8)采用对偶理论转化成二次规划求解,则约束表达式的对偶式为:

$$\min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + \epsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \quad (9)$$

约束条件为:

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, l \quad (10)$$

其中,  $Q_{ij} = K(x_i, x_j) \equiv \varphi(x_i)^T \varphi(x_j)$ 。

通过二次规划算法可得 SVR 回归预测模

型为:

$$\sum_{i=1}^l (-\alpha_i + \alpha_i^*) K(x_i, x) + b \quad (11)$$

### 1.3 粒子群优化支持向量机参数

粒子在寻优过程中,每个优化问题的解均被认为是搜索空间中的一个粒子,每一个粒子都根据适应度函数计算出其适应值,且每个粒子还有一个速度,它们决定粒子飞翔的方向和距离,每一次迭代的过程中,粒子根据个体最优值和全局最优值更新。文献[10]详述了粒子群算法的寻优原理。粒子群优化支持向量机参数详细步骤见文献[11]。

## 2 滑坡位移特征分析

白家包滑坡位于秭归县归州镇向家店村,距香溪河入长江口 2.5 km。该滑坡面积约为  $2.2 \times 10^5 \text{ m}^2$ , 体积约为  $9.9 \times 10^6 \text{ m}^3$ , 为一深层土质滑坡。滑坡体上布设有 GPS 位移监测、地下水位监测和降雨量监测等(如图 1 所示)。GPS 监测点 ZG324 位于滑坡体中部,其位移是该滑坡整体变形运动的主要体现。因此,可以通过 ZG324 监测数据深入分析滑坡位移与影响因素之间的响应关系。2007 年监测点 ZG324 的相对位移量和同期降雨量、库水位、地下水位之间的相关性如图 2 所示。由图 2 可知,降雨量和地下水位与相对位移量均具有正相关性,而库水位与相对位移量之间具有负相关性。地下水位与相对位移量之间的相关性相对于降雨量和库水位更为复杂,表现出非线性的响应关系;而降雨量和库水位与相对位移量之间为线性响应关系,即降雨量越大,相对位移量越大,库水位下降,相对位移量增大;并且降雨量和库水位对位移量的影响具有一定滞后效应<sup>[11-12]</sup>。

降雨量与库水位之间具有负相关性,原因为每年 5 月开始随着降雨量的增多,即汛期的到来,三峡水库为防洪而放水,库水位大幅下降。由于降雨量和库水位下降与位移量之间具有正相关,所以当降雨量增多和库水位大幅下降两个影响因素同时出现时,往往造成相对位移量出现突变,如 2007 年 6 月。这表明在库水位下降过程中,若出现明显降雨过程,对滑坡的稳定性会造成更为不利的影

响。地下水位与降雨量和库水位之间呈现出非线性的响应关系,由于地下水位受降雨和库水位的双重影响而变化。库水位上涨未引起该滑坡变形

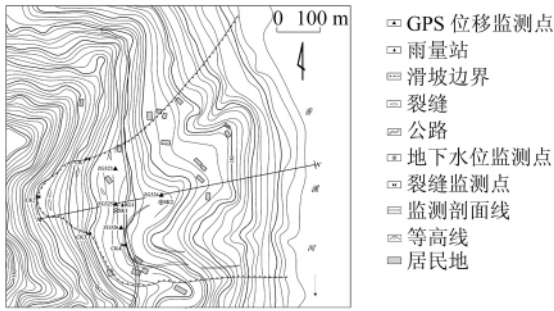


图1 白家包滑坡专业监测布置示意图

Fig. 1 Schematic Diagram of Monitoring Arrangement in Baijiabao Landslide

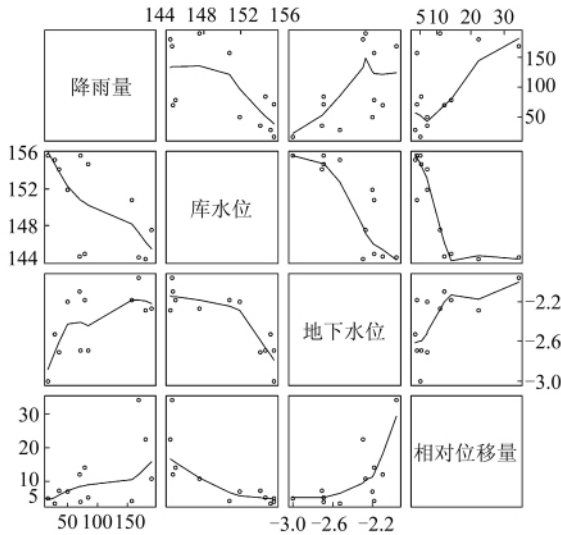


图2 ZG324的相对位移、降雨量、库水位和地下水水位之间的相关性

Fig. 2 Correlation Between Precipitation, Reservoir Water Level, Ground Water Level and Relative Displacement of ZG324

的明显变化,而库水位下降对该滑坡变形产生了很大影响。究其原因主要是该滑坡滑体物质为含砾石黏土,渗透性差,水库蓄水时,水向坡体内渗透缓慢,库水位与滑坡体地下水水位形成负落差,反压坡体,有利于滑坡体的稳定;而水库退水时,地下水向水库排水,由于坡体渗透性差,地下水排除缓慢,地下水与库水位形成正落差,动水压力指向坡体外侧,不利于滑坡的稳定,特别是在水位下降速度较大时,表现出弱透水滞后型滑坡的变形特征。

### 3 滑坡位移预测

#### 3.1 位移预测流程

以2007年GPS位移监测点ZG324的相对位移为预测对象,选取降雨量、库水位和地下水水位

三类因素为影响因素。由于降雨量和库水位对滑坡位移影响具有一定的滞后效应,所以选取当月降雨量、前两个月累计降雨量、当月日最大降雨量、当月一次连续最大降雨量、当月库水位、当月库水位变化幅度、当月库水位变化速率、当月地下水水位和当月地下水变化幅度9个因子作为滑坡相对位移的初始影响因子,建立滑坡相对位移与影响因素之间的响应模型,对滑坡相对位移进行预测,其详细流程如下。

1) 对初始影响因子和位移监测数据进行异常点剔除、滤波等预处理后,并把它们统一归一化到-1~1。

2) 设置KPCA核函数为高斯径向基核函数,对滑坡相对位移的9个初始影响因子进行核主成分分析。主成分方差累计贡献率如图3所示,其中第一主成分方差贡献率为0.622,而前三主成分累计贡献率达0.917,累计贡献率为0.9的主成分基本上包含了全部特征所具有的信息。因此,选取前三个主成分作为建立支持向量机模型的新影响因子。

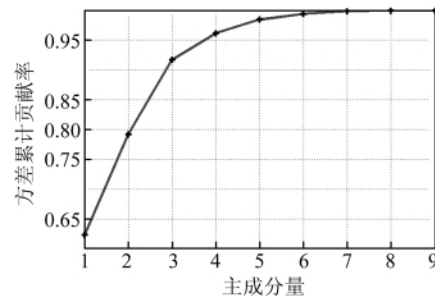


图3 主成分方差累计贡献率

Fig. 3 Cumulative Percentage of Principal Component

3) 将新影响因子和位移数据分成两部分,其中2007年1月至10月的数据作为模型训练样本,11月和12月的数据为模型检验数据。

4) 设置PSO算法的初始参数,学习因子 $c_1$ 、 $c_2$ 均设为2,惯性权重 $w$ 设为0.6,种群规模设为20,最大迭代步数设为100。然后搜索得到支持向量机的参数惩罚因子 $C=8.825$ 和RBF核函数参数 $\gamma=0.241$ 。利用获得的最佳模型参数,对样本数据进行学习,建立滑坡相对位移与影响因素之间的非线性响应模型。

5) 利用所建立的模型对检验样本进行预测,检验模型的有效性和预测能力。

6) 选择模型评价指标并评价滑坡位移预测模型。

#### 3.2 预测结果与分析

利用上述方法所建立的模型对白家包滑坡

2007 年 GPS 监测点 ZG324 的相对位移量进行预测,其整体预测结果如表 1 所示。由表 1 可知,KPCA-PSO-SVR 模型对样本数据拟合的最小绝对误差为 0.087,最大绝对误差为 1.576,相对误差分别为 1.621%和 13.253%,说明该模型具有较好的拟合能力。对于 2007 年 11 月和 12 月的检验数据而言,模型预测的绝对误差和相对误差分别为 0.047、0.732、1.187%和 14.366%,预测精度满足要求,表明该模型具有较强的泛化能力。但随着预测时间间隔的增大,其模型的预测能力有所降低。KPCA-PSO-SVR 模型对整个时序数

据预测的平均绝对误差是 0.760,平均相对误差为 7.563%,综合表明该模型对滑坡位移具有较好的拟合和泛化能力。

在经过同样 KPCA 处理后的情况下,分别采用不同的优化算法搜索 SVR 的模型参数,然后建立相应的响应模型,对滑坡位移时序数据进行预测,其结果见表 1。由表中的平均绝对误差和平均相对误差指标可以看出,采用 PSO 进行参数寻优后所建立的模型的预测能力明显强于基于传统网格搜索法(grid search,GS)和遗传算法(genetic algorithm,GA)优化的模型。

表 1 2007 年 12 期监测点 ZG324 的相对位移预测值及其误差  
Tab. 1 Predicted Relative Displacement of ZG324 and Error Value

时间 年/月/日	实测值 /mm	KPCA-GS-SVR			KPCA-GA-SVR			KPCA-PSO-SVR		
		预测值 /mm	绝对误差/mm	相对误差/%	预测值 /mm	绝对误差/mm	相对误差/%	预测值 /mm	绝对误差/mm	相对误差/%
2007/1/15	3.606	3.121	0.484	13.437	3.130	0.475	13.182	3.128	0.478	13.253
2007/2/02	5.385	5.977	0.592	10.996	5.876	0.491	9.117	5.298	0.087	1.621
2007/3/11	7.071	6.225	0.846	11.964	6.598	0.473	6.694	7.199	0.128	1.814
2007/4/12	4.243	4.717	0.474	11.183	4.729	0.486	11.466	4.726	0.483	11.393
2007/5/13	10.770	12.326	1.556	14.446	12.318	1.547	14.365	11.321	0.550	5.110
2007/6/18	34.059	28.339	5.720	16.795	32.578	1.481	4.349	32.580	1.479	4.341
2007/7/18	22.361	20.879	1.482	6.627	20.876	1.485	6.639	20.880	1.481	6.622
2007/8/18	12.166	13.722	1.557	12.795	13.721	1.556	12.789	13.728	1.563	12.845
2007/9/20	14.142	12.581	1.561	11.038	12.570	1.572	11.116	12.567	1.576	11.141
2007/10/20	7.211	7.826	0.615	8.530	7.727	0.516	7.151	7.721	0.510	7.067
2007/11/10	4.000	3.877	0.123	3.064	3.729	0.271	6.776	3.953	0.047	1.187
2007/12/13	5.099	6.189	1.090	21.384	5.948	0.849	16.647	5.832	0.732	14.366
最小值			0.123	3.064		0.271	4.349		0.047	1.187
最大值			5.720	21.384		1.572	16.647		1.576	14.366
平均值			1.342	11.855		0.934	10.024		0.760	7.563

利用不同的优化算法搜索 SVR 模型参数,并以滑坡位移的 9 个初始影响因子直接作为 SVR 的特征向量,即不经过 KPCA 处理,其相应模型的预测结果误差如表 2 所示。由表 2 可知,基于三种优化算法所建立的模型的预测误差相对于通过 KPCA 处理后所建立的模型都大,其预测能力均有下降。究其原因,主要是 KPCA 处理在保留原数据主要信息的前提下,提取出数据中的非线性特征信息,去除数据间的相关性,减少冗余信息及数据噪音,降低数据维数,使原来的多维问题得以简化,基于此建立的预测模型更加优化有效。PSO-SVR 模型预测的平均绝对误差和平均相对误差值均最小,这也佐证了对于滑坡位移数据集模型参数优化 PSO 算法优于传统网格搜索法和遗传算法。

### 3.3 预测模型评价

通过计算上述 6 种模型的均方差(MSE)、相关系数的平方( $r^2$ )和 AIC 值如表 3 所示,其中,  $N$  表示不经过核立成份(KPCA)分析。从表中可

表 2 不同模型预测误差

Tab. 2 Error of Predicted Vale Based on Different Models

时间/ 年/月/日	GS-SVR		GA-SVR		PSO-SVR	
	绝对误差/mm	相对误差/%	绝对误差/mm	相对误差/%	绝对误差/mm	相对误差/%
2007/1/15	0.576	15.975	0.579	16.063	0.572	15.876
2007/2/02	1.507	27.978	1.117	20.737	0.510	9.468
2007/3/11	1.500	21.206	1.503	21.259	1.495	21.139
2007/4/12	1.298	30.597	0.911	21.477	0.684	16.115
2007/5/13	1.558	14.467	1.556	14.448	1.550	14.390
2007/6/18	9.384	27.552	1.845	5.417	1.487	4.366
2007/7/18	1.488	6.656	1.485	6.643	1.487	6.648
2007/8/18	1.558	12.804	1.565	12.867	1.561	12.832
2007/9/20	1.573	11.119	1.567	11.080	1.571	11.105
2007/10/20	1.516	21.018	1.518	21.057	1.514	20.989
2007/11/10	0.428	10.707	0.530	13.253	0.307	7.685
2007/12/13	1.730	33.936	1.571	30.816	1.484	29.105
最小值	0.428	6.656	0.530	5.417	0.307	4.366
最大值	9.384	33.936	1.845	30.816	1.571	29.105
平均值	2.010	19.501	1.312	16.260	1.185	14.143

知,KPCA-PSO-SVR 模型的 MSE 和 AIC 值均最小,而  $r^2$  值最大。这综合说明了经过 KPCA 处

理后,利用 PSO 搜索算法进行模型参数优化而建立的 SVR 模型为最优稳健模型,对滑坡变形位移量与影响因素之间的复杂响应关系具有最佳的拟合及预测能力。而 GS-SVR 模型的 MSE 和 AIC 值相对最大而  $r^2$  值最小,表明其拟合及预测能力相对有限。

表 3 不同模型评价指标值

Tab. 3 Value of Performance of Different Models

指标	(N/KPCA)-GS -SVR	(N/KPCA)-GA -SVR	(N/KPCA)-PSO -SVR
MSE	18.268/7.551	3.778/2.278	3.267/1.815
$r^2$	0.928/0.975	0.979/0.988	0.981/0.992
AIC	74.363/51.761	55.451/37.383	53.708/34.655

## 4 结 语

通过滑坡位移与影响因素之间的综合响应分析,利用降雨、库水位和地下水等影响因素预测滑坡位移,结果表明该方法是一种行之有效的位移预测方法,对滑坡预测预报具有关键作用。将核主成分分析、粒子群算法和支持向量机进行有机结合,构建出 KPCA-PSO-SVR 协同优化模型,对滑坡相对位移进行预测,其预测结果的平均绝对误差和相对误差分别为 0.760 和 7.563%。通过与其他预测模型对比,发现该模型的拟合和泛化能力最优。

### 参 考 文 献

- [1] 许强,汤明高,徐开祥,等. 滑坡时空演化规律及预警预报研究[J]. 岩石力学与工程学报,2008,27(6):1 104-1 112
- [2] 徐冲,刘保国,刘开云,等. 基于粒子群—高斯过程回归耦合算法的滑坡位移时序分析预测智能模型[J]. 岩土力学,2011,32(6):1 669-1 675
- [3] 赵洪波,冯夏庭. 非线性位移时间序列预测的进化—支持向量机方法及应用[J]. 岩土工程学报,2003,25(4):468-471
- [4] 林大超,安凤平,郭章林,等. 滑坡位移的多模态支持向量机模型预测[J]. 岩土力学,2011,32(S1):451-458
- [5] 杜娟,殷坤龙,柴波. 基于诱发因素响应分析的滑坡位移预测模型研究[J]. 岩石力学与工程学报,2009,28(9):1 783-1 789
- [6] 范千,花向红. 基于相空间重构与支持向量机预测滑坡位移的一种新方法[J]. 武汉大学学报·信息科学版,2009,34(2):248-251
- [7] 董辉,傅鹤林,冷伍明,等. Boosting 集成支持向量回归机的滑坡位移预测[J]. 湖南大学学报(自然科学版),2007,34(9):6-10
- [8] 王新洲,范千,许承权,等. 基于小波变换和支持向量机的大坝变形预测[J]. 武汉大学学报·信息科学版,2008,33(5):469-471
- [9] Vapnik V N. 统计学习理论的本质[M]. 张学工,译. 北京:清华大学出版社,2000
- [10] Kennedy J, Eberhart R C. Particle Swarm Optimization[C]. IEEE International Conference on Neural Networks, Perth, WA, 1995
- [11] Guzzetti F, Peruccacci S, Rossi M, et al. The Rainfall Intensity-duration Control of Shallow Landslides and Debris Flows: an Update[J]. Landslides, 2008, 5(1):3-17
- [12] 刘新喜,夏元友,张显书,等. 库水位下降对滑坡稳定性的影响[J]. 岩石力学与工程学报,2005,24(8):1 439-1 444

第一作者简介:彭令,博士生,研究方向为地质灾害预测预报。  
E-mail:wuhanpengling@163.com

## Prediction of Landslide Displacement Based on KPCA and PSO-SVR

PENG Ling<sup>1</sup> NIU Ruiqing<sup>1</sup> ZHAO Yannan<sup>1</sup> DENG Qinglu<sup>2</sup>

(1 Institute of Geophysics and Geomatics, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China)

(2 Engineering Faculty, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China)

**Abstract:** Taking Baijiabao landslide in the Three Gorges reservoir area for example, the kernel principal component analysis (KPCA) was employed to extract main features from influential factors data in order to obtain the principal component, which was used as the feature of support vector machine (SVM), and then building the displacement prediction model, the

(下转第 161 页)

## Automatic Preprocessing of Tidal Gravity Observation Data

XU Chuang<sup>1</sup> LUO Zhicai<sup>1,2</sup> LIN Xu<sup>1</sup> ZHOU Boyang<sup>1</sup>

(1 School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 Key Laboratory of Geospace Environment and Geodesy, Ministry of Education,  
Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

**Abstract:** The preprocessing of tidal gravity observation data is very important to obtain high-quality tidal harmonic analysis results. The preprocessing methods of tidal gravity observation data are studied systematically, and average filtering method and wavelet filtering method for downsampling original tidal gravity observation data are given in the paper, as well as the linear interpolation method and the cubic spline interpolation method for processing interrupt data. The automatic preprocessing software of the tidal gravity observation data (APTsoft) is developed, which can calibrate and correct automatically abnormal data such as spikes, steps and interrupts. Finally, the experimental results show that the preprocessing methods and APTsoft are very effective, and APTsoft can be applied to the automatic preprocessing of tidal gravity observation data.

**Key words:** tidal gravity; automatic preprocessing; wavelet filtering; APTsoft

---

**About the first author:** XU Chuang, Ph. D candidate, majors in Earth's rotation and tide and Earth's free oscillation.

E-mail: xuchuang526610@163.com

(上接第 152 页)

SVM model parameter of which was optimized by the particle swarm optimization (PSO) algorithm. Finally, the cooperative optimization model based on KPCA and PSO-SVR was proposed. And it was applied to predict displacement of landslide based on influential factors. The average absolute error and relative error of the prediction results were 0.760 and 7.563% respectively. By comparison with other forecast models, it was found that the fitting and generalization of this model are the best. The results indicate that the predicted value of the model is consistent with the monitoring data. It plays a key role in landslide hazard prediction and warning.

**Key words:** landslide; displacement prediction; kernel principal component analysis; particle swarm optimization; support vector machine

---

**About the first author:** PENG Ling, Ph. D candidate, majors in deformation monitoring and geological hazard prediction.

E-mail: wuhanpengling@163.com