

投影寻踪法在建立积雪深度模型上的应用

胡淑娟¹,雷晓云¹,白磊²,邢坤¹,刘艺¹,吕姣姣¹

(1. 新疆农业大学,新疆 乌鲁木齐 830052;2.中国科学院新疆生态与地理研究所,新疆 乌鲁木齐 830011)

摘要:针对融雪期积雪深度变化问题的高维非线性、非正态特点,利用投影寻踪法将积雪深度变化多指标问题转化为单一投影指标问题,采用遗传算法优化投影方向得出最佳投影值,最终得到对积雪深度变化影响较大的因子,实现在低维空间上反应积雪深度变化的目的。实例分析结果表明:影响积雪深度变化的各因素贡献力依次为:土壤地表温度>净辐射>空气温度>土壤体积含水量>风速>空气相对湿度>降水>水汽压。通过建立 MLP 神经网络积雪深度模型表明:利用投影寻踪法提取的积雪深度主要影响因子是可靠的,且提取的主要影响因子为简化积雪深度模型提供理论依据。

关键词:投影寻踪;遗传算法;雪深;模型;主因子

中图分类号:P333 **文献标识码:**A **文章编号:**1000-0852(2016)05-0017-05

积雪深度是估算雪水当量、流域水量平衡和春季融雪径流模拟的重要参数^[1],目前积雪深度的获取大都通过遥感或者建立模型。乔鹏^[2]使用遥感提取积雪信息,构造了以能量平衡为基础的融雪径流模型;卢新玉^[3]利用被动微波建立了基于 AMSR-E 的积雪深度反演模型,得到北疆地区积雪的分布情况;王澄海^[4]利用不同的气候模式,模拟预测新疆地区未来的积雪变化,预测表明新疆北部地区的积雪深度均呈减少趋势;董安青^[5]分析了新疆积雪与降水、气温、气压、相对湿度的相关关系,并且建立了积雪深度回归模型;娄伟平等^[6]建立了一种应用气象站观测数据进行积雪表面温度和深度预测的模型。积雪深度模型建立时需要全面考虑影响积雪深度的因素以保证模型的精确,根据前人研究可知积雪深度受到温度、降水、净辐射、水汽压等因素的影响^[6-7],要建立高精度的积雪深度模型就要把影响因素全都作为参数代入到积雪深度模型中,多输入参数使得积雪深度模型复杂。如果能从众多的影响因素中筛选出影响积雪深度的主要影响因素,只把主要影响因素作为积雪深度模型的输入参数,将很大程度上简化积雪深度

模型,并且自动气象站观测项目也可以相应的减少,只观测积雪深度的主要影响因素,达到降低观测成本的目的。

传统的多元分析法建模的前提条件多要求:总体服从正态分布、线性,但是在实际中搜集到的资料未必满足正态分布,并且当数据维数较高时,传统的多元分析法由于自身及外界的原因又有一些困难^[8],因此在处理高维的非正态、非线性的数据时传统的多元分析方法达不到期望的效果,难以找到我们所需数据规律和特征。投影寻踪法则是一种探索性数据分析方法,对数据本身没有过多要求和限制,可以把高维数据投影到低维子空间上,寻找出能反映高维数据结构或特征的投影方向,实现降低数据维数的目的。

目前,投影寻踪法已经广泛的应用到防洪调度方案优选、水资源评价、边坡稳定分析、环境监测等诸多领域^[9-13],但在积雪深度模型建立中的应用较少,由于积雪深度变化的影响因素构成的高维数据具有非线性、非正态特点,因此可以利用投影寻踪法把高维数据投影到低维子空间上,寻找出能反映高维数据结构或特征的投影方向,实现降低数据维数的目的,提取出影

收稿日期:2015-08-25

基金项目:国家自然科学基金(U1178301);天山中段玛纳斯河流域融雪机理及其径流过程模拟的研究;新疆水文水资源重点学科建设基金

作者简介:胡淑娟(1990-),女,河南平顶山人,硕士研究生,研究方向为水文学及水资源。E-mail:18299162995@163.com

通讯作者:雷晓云(1961-),女,教授,博士生导师,主要从事水资源高效利用理论技术教学与研究。E-mail:leixiaoyun888@163.com

响积雪深度变化的主因子,为简化积雪深度模型提供依据。

1 数据说明

文中实例计算所采用的数据来源于新疆巩乃斯河流域河源布设的自动气象站 2005~2009 年每半小时的观测数据,观测项目包括:雪深、空气温度、相对湿度、水汽压、净辐射、土壤地表温度、土壤体积含水率、风速、降水。由于在每年 3 月份积雪深度变化最大,因此该文在分析数据时选取 3 月份每日平均数值作为基础数据进行计算分析。

2 基于积雪深度变化下的遗传算法投影寻踪模型

影响积雪深度变化的水文气象因子包括大气温度、气压、水汽压比、降水、湿度、土壤湿度、土壤温度、地表反照率等^[14];王金叶等研究表明积雪深度在森林区受坡度、坡向、森林郁闭度、迎风面、被风面等影响很大^[15]。根据前人的研究成果可知,影响积雪深度变化的因素很多,但是由于建立积雪深度模型时的输入参数多为气象站获得的可测数据,且本文实例计算使用数据雪崩站所在地区为非森林区,故选取雪崩站观测到的净辐射、土壤地表温度、空气温度、空气相对湿度、降水、水汽压、土壤体积含水量、风速这 8 个影响因素来分析它们对积雪深度变化的贡献力。

2.1 投影寻踪

投影寻踪模型思路是:利用现代信息技术,通过某种组合把高维数据投影到低维子空间上,采用投影目标函数来衡量投影构形暴露某种数据结构的可能性大小,设定约束条件,寻找出投影目标函数最大化的投影方向向量^[16-17],根据最佳投影方向向量即评价指标权重,判定各评价指标对子系统和系统的影响程度。

建模主要过程如下:

(1) 评价指标原始值标准化处理

对越大越优(正向)的指标,一致无量纲化公式为:

$$X_{(i,j)}^* = \frac{X_{(i,j)} - X_{j\min}}{X_{j\max} - X_{j\min}} \quad (1)$$

对越小越优(逆向)的指标,一致无量纲化公式为:

$$X_{(i,j)}^* = 1 - \frac{X_{(i,j)} - X_{j\min}}{X_{j\max} - X_{j\min}} \quad (2)$$

式中: $X_{(i,j)}^*$ 为第 i 个样本的第 j 个指标值,其中 $i=1,2,\dots,n$, $j=1,2,\dots,p$ (n,p 分别为样本的个数和影响积雪深度

变化的指标数); $X_{(i,j)}$ 为指标值经过标准化处理后的数据序列; $X_{j\max}$ 、 $X_{j\min}$ 分别为第 j 个指标的最大值和最小值。

(2) 构造投影指标函数 $Q(a)$

投影寻踪法方法就是把 p 维数据 $\{X_{(i,j)}\} | j=1,2,\dots,p$ 综合成以 $a=\{a(1),a(2),\dots,a(p)\}$ 为投影方向的一维投影值 $Z(i)$:

$$Z(i) = \sum_{j=1}^p a(j)X_{(i,j)} \quad (i=1,2,\dots,n) \quad (3)$$

综合投影指标时,要求投影值 $Z(i)$ 的散布特征应为:局部投影点尽可能密集,最好凝聚成若干个点团,而在整体上投影点团之间尽可能散开。因此,投影指标函数可表达为:

$$Q(a) = S_z * D_z \quad (4)$$

式中: S_z 为投影值 $Z(i)$ 的标准差; D_z 为投影值 $Z(i)$ 的局部密度。

(3) 优化最佳投影方向

$Q(a)$ 值与投影方向 a 有关,当 $Q(a)$ 取最大值时 a 的方向最能反映数据结构特征的方向,所以可以通过求解投影指标函数优化问题来估计最佳投影方向,即求解:

$$\text{Max: } Q(a) = S_z \cdot D_z \quad (7)$$

$$\text{S.t: } \sum_{j=1}^p a^2(j) = 1 \quad (8)$$

2.2 遗传算法优化投影方向

上述以 $\{a(j) | j=1,2,\dots,p\}$ 为优化变量的复杂化非线性优化问题,难以用常规优化方法处理^[14]。遗传算法能够快速找到目标函数在约束条件下的全局最优解,通过遗传算法求解最优投影方向,实现积雪深度变化主要影响因子的提取,具体过程如下,图 1 为遗传算法优化投影方向的操作流程图。

(1) 随机生成初始群体,其由 n 个个体构成,每个个体由 $\{a_1, a_2, \dots, a_8\}$ 组成, a_i 为 $[0, 1]$ 上的 8 个均匀分布的随机数, a_i 即为投影方向。

(2) 按照(7)式计算群体中每个个体的适应度。

(3) 保留群体中适应度最高的个体,它不参与交叉和变异运算,而直接将其复制到下一代。对群体中的其它个体,采用轮盘赌选择法进行选择。

(4) 根据选择出来的个体进行交叉运算、变异运算。

(5) 生成新一代群体,反复进行(2)~(4),每进行一次,群体就进化一代,直到适应度满足要求。

(6) 适应度达到最优所对应的投影方向即为最优投影方向 a^* 。

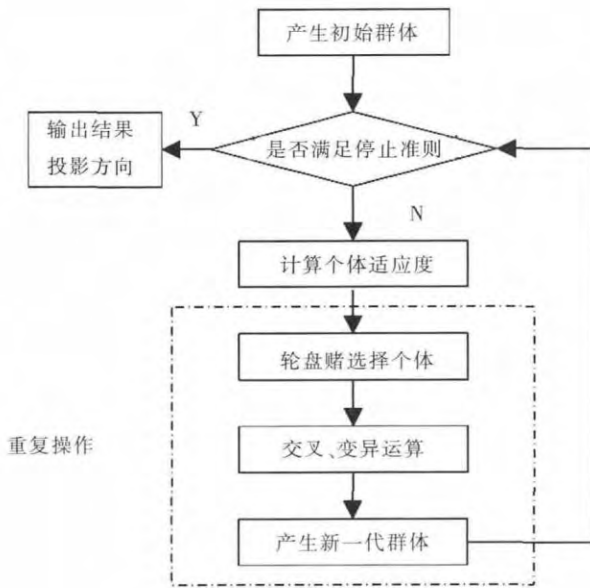


图1 遗传算法优化投影方向运算流程图
Fig.1 The flow of the optimized projection direction by genetic algorithm

3 积雪深度模型应用及模拟结果分析

3.1 基于投影寻踪法的关键影响因子筛选

雪崩站获得的观测项目: 雪深 Y 、空气温度 $X1$ 、空气相对湿度 $X2$ 、水汽压 $X3$ 、净辐射 $X4$ 、土壤地表温度 $X5$ 、土壤体积含水量 $X6$ 、风速 $X7$ 、降水 $X8$, 结合投影寻踪模型的步骤对数据进行标准化处理, 由于指标的不同属性在标准化处理时有两种不同的处理方式, 所以要判断各指标(影响因素)对于积雪深度而言是正指标还是负指标, 因此借助 spss 软件分析 8 个影响因素与积雪深度的相关关系, 确定各影响因素标准化处理的方式, 相关分析结果见表 1。

根据表 1 的结果显示, 可知空气温度、空气相对湿度、水汽压、净辐射、土壤地表温度、土壤体积含水量、降水均与雪深呈现负相关, 故使用(2)式进行标准化处理, 只有风速平均值与雪深呈现出正相关, 故使用(1)式进行标准化处理, 标准化处理结果见表 2。

表1 融雪期积雪深度与各影响因子的相关系数

Table1 The correlation coefficient of the snow depth and influence factors in snowmelt period

Pearson 相关性	雪深	空气温度	空气相对湿度	水汽压	净辐射	土壤地表温度	土壤体积含水量	风速平均值	降水
雪深	1	-0.69**	-0.21	-0.58*	-0.78**	-0.85**	-0.93**	0.5*	-0.06

注:** 在 0.01 水平(双侧)上显著相关;* 在 0.05 水平(双侧)上显著相关。

表2 影响积雪深度的各因素标准化处理

Table2 The standardized treatment of the influence factor of snow depth

时间	空气温度	空气相对湿度	水汽压平均值	净辐射平均值	土壤地表温度	土壤体积含水量	风速	降水
3月10日	1	0.74	1	0.97	1	1	0.59	1
3月11日	0.71	1	0.93	1	0.93	0.97	0.948	1
3月12日	0.31	1	0.73	0.81	0.82	0.93	0.67	0.88
3月13日	0.23	0.57	0.45	0.58	0.82	0.9	0.4	0.93
3月14日	0.16	0.95	0.57	0.78	0.71	0.88	0.74	0.96
3月15日	0.03	0.8	0.35	0.62	0.68	0.85	0.19	0.93
3月16日	0.2	0.03	0.05	0.66	0.73	0.84	0.22	0
3月17日	0.03	0.37	0.02	0.38	0.71	0.83	0	0.7
3月18日	0.23	0.07	0.11	0.67	0.53	0.82	1	0.31
3月19日	0.45	0.54	0.63	0.82	0.51	0.78	0.65	0.54
3月20日	0.28	0.92	0.64	0.81	0.4	0.76	0.62	0.96
3月21日	0.12	0.82	0.46	0.46	0.33	0.51	0.3	0.95
3月22日	0	0.81	0.33	0.59	0.32	0.29	0.52	0.93
3月23日	0.13	0.45	0.19	0.76	0.48	0.3	0.46	0.88
3月24日	0.18	0	0.01	0.61	0.42	0.33	0.25	0.53
3月25日	0.13	0.72	0.46	0.16	0.34	0.25	0.21	0.95
3月26日	0.19	0.88	0.55	0.56	0.25	0.17	0.37	0.96
3月27日	0.04	0.86	0.45	0.15	0.16	0.11	0.31	0.95
3月28日	0.03	0.95	0.45	0.19	0.23	0.03	0.49	0.94
3月29日	0.03	0.75	0.36	0	0.29	0	0.22	0.92
3月30日	0.15	0.07	0	0.56	0	0.06	0.26	0.55

根据前述理论,使用 MATLAB 软件编写遗传算法优化投影方向的投影寻踪程序,将通过标准化处理的样本数据作为训练数据,代入到 MATLAB 编写源代码中,得到各指标的的最佳投影方向为:

$$\alpha^*=(0.27, 0.03, 0.14, 0.47, 0.76, 0.26, 0.15, 0.11)$$

最佳投影方向向量实质上就是对应各影响因子的权重,它可以反映出各影响因子对积雪深度变化的影响程度,其值越大说明相应的影响因子对融雪期积雪深度变化的影响越大^[18]。得到各影响因子对积雪深度变化的贡献力排序:土壤地表温度>净辐射>空气温度>土壤体积含水量>风速>水汽压>降水>空气相对湿度。

3.2 基于 MLP 神经网络的积雪深度模拟模型

为了验证投影寻踪法提取出的主要影响因素在积雪深度预测模型应用中的实用性,利用 MLP 神经网络建立积雪深度预测模型,将土壤地表温度、净辐射、空气温度作为输入层的 3 个输入因子进行计算。MLP 神经网络模型的建立借助 spss 软件完成,数据选用积雪雪崩站 2005 年 4 月份每间隔 30min 的监测数据,共 744 组数据。

将 2005 年 4 月份 744 组数据作为模拟数据进行计算,将土壤地表温度、净辐射、空气温度作为神经网络输入层的 3 个输入因子,对其进行 MLP 神经网络的计算。表 3 给出了经过神经网络计算后得到的预测雪深,预测值与实测值的相对误差均值最大值为 22.72%,可知投影寻踪法得到的主要影响因子土壤地表温度、净辐射、空气温度在建立积雪深度模型时具有较强实用性。在建立积雪深度模型时对于积雪深度变化贡献力小的影响因子,可以忽略不考虑,从而达到简化积雪深度模型的目的。

表3 不同雪深的模拟结果与实测结果对比表
Table3 Comparison between the simulated and measured snow depths

雪深范围 /cm	样本数	深度误差 均值/cm	最大深度 误差值/cm	相对误差 均值/%	最大相对 误差值/%
20~25	36	3.22	3.97	14.76	33.91
15~20	44	2.17	2.63	13.19	27.87
10~15	27	1.82	2.83	13.59	33.33
5~10	205	1.32	2.93	13.89	31.24
0~5	432	1.56	-3.42	22.72	38.83

4 结论与建议

本文建立了遗传算法优化投影方向的积雪深度

投影寻踪模型,通过对雪崩站自动气象实测数据的训练计算分析,结果表明:融雪期影响积雪深度变化的各因素贡献力排序为土壤地表温度>净辐射>空气温度>土壤体积含水量>风速>水汽压>降水>空气相对湿度。根据提取出来的主要影响因子建立 MLP 神经网络积雪深度预测模型,将土壤地表温度、净辐射、空气温度作为神经网络的输入因子得到预测雪深的最大相对误差平均值为 22.72%,积雪深度的模拟精度较高,因此以后在利用自动气象站监测数据建立积雪深度模型时,可以只监测贡献力大的影响因子作为积雪深度模型的输入参数,达到降低监测成本和简化积雪深度模型的目的。

参考文献:

- [1] 冯学智,柏延臣,史正涛,等. 新疆地区积雪深度的克里格内插估计[J]. 冰川冻土, 2000,22(4):358-361. (FENG Xuezhi, BAI Yanchen, SHI Zhengtao, et al. Snow depth in north Xinjiang region estimated by Kriging interpolation[J]. Journal of Glaciology and Geocryology, 2000,22(4):358-361. (in Chinese))
- [2] 房世峰,裴欢,刘志辉,等. 遥感和 GIS 支持下的分布式融雪径流过程模拟研究[J]. 遥感学报, 2008,12(4):655-662. (FANG Shifeng, PEI Huan, LIU Zhihui, et al. Study on the distributed snowmelt run off process based on RS and GIS[J]. Journal of Remote Sensing, 2008,12(4):655-662. (in Chinese))
- [3] 赵晓萌,李栋梁,陈光宇. 基于 GIS 的东北及邻近地区积雪深度空间化方法[J]. 干旱区研究, 2012,29(6):927-933. (ZHAO Xiaomeng, LI Dongliang, CHEN Guangyu. GIS-based spatializing method for estimating snow cover depth in northeast China and its neighbours[J]. Arid Zone Research, 2012,29(6):927-933. (in Chinese))
- [4] Rao P K, 许健民译. 气象卫星——系统、资料及其在环境中的应用[M]. 北京:气象出版社, 1994:29-31,354-367. (Rao P K, Translated by XU Jianmin. Meteorological Satellite: Application of System and Date in Environmental [M]. Beijing: China Meteorological Press, 1994:29-31,354-367. (in Chinese))
- [5] 赵金顺,李立军,魏庆朝. 积雪灾害防治中的积雪深度与风速的计算[J]. 中国安全科学报, 2005,15(9):3-6. (ZHAO Jinshun, LI Lijun, WEI Qingchao. Calculation of snow depth and wind velocity in snow calamity control[J]. China Safety Science Journal, 2005,15(9):3-6. (in Chinese))
- [6] 董安青. 基于遥感方法的新疆积雪覆盖与气象因子的相关性研究[D]. 乌鲁木齐: 新疆师范大学, 2009. (DONG Anqing. Study on the Relevance of Snow Cover and Meteorological Factors of Xinjiang Based on Remote Sensing Methods [D]. Urumqi: Xinjiang Normal University, 2009. (in Chinese))
- [7] 娄伟平,吴利红,赵慧娟. 积雪温度和深度特征变化分析[J]. 科技导报, 2006,24(9):32-35. (LOU Weiping, WU Lihong, ZHAO Huijuan. Analysis of snow cover temperature and depth with data from weather stations[J]. Science & Technology Review, 2006,24(9):32-35.

- (in Chinese))
- [8] 王芹芹, 雷晓云. 新疆内陆河流域融雪期积雪深度变化影响因素研究[J]. 水资源与水工程学报, 2012,23(6):27-31. (WANG Qinqin, LEI Xiaoyun. Study on influence factors of snow depth change during period of snow melting in inland river basin of Xinjiang[J]. Journal of Water Resources and Water Engineering, 2012,23 (6):27-31. (in Chinese))
- [9] 付强, 赵小勇. 投影寻踪模型原理及其应用[M]. 北京: 科学出版社, 2006:46-51. (FU Qiang, ZHAO Xiaoyong. Projection Pursuit Model: Principle and Applications [M]. Beijing: Science Press, 2006:46-51. (in Chinese))
- [10] 张灵, 陈晓宏. 基于投影寻踪的防洪系统脆弱性评价模型[J]. 水电能源科学, 2010,28(7):37-39+56. (ZHANG Ling, CHEN Xiaohong. Assessing model of vulnerability for flood control system based on projection pursuit[J]. Water Resources and Power, 2010,28(7):37-39+56. (in Chinese))
- [11] 付强, 付红, 王立坤. 基于加速遗传算法的投影寻踪模型在水质评价模型中应用[J]. 地理科学, 2003,23(2):236-239. (FU Qiang, FU Hong, WANG Likun. Study on the PPE model based on RAGA to evaluating the water quality [J]. Geographical Research, 2003,23(2): 236-239. (in Chinese))
- [12] 金永强, 苏怀智. 基于和声搜索的边坡稳定性投影寻踪聚类分析[J]. 水利学报, 2007,10(S1):682-686. (JIN Yongqiang, SU Huaizhi. Slope stability projection on pursuit clustering analysis based on harmony search[J]. Journal of Hydraulic Engineering, 2007,10(S1): 682-686. (in Chinese))
- [13] 张欣莉, 丁晶, 金菊良. 基于遗传算法的参数投影寻踪回归及其在洪水预报中应用[J]. 水利学报, 2010,6(6):45-48. (ZHANG Xinli, DING Jing, JIN Juliang. Application of parametric projection pursuit regression based on genetic algorithm in flood forecasting [J]. Journal of Hydraulic Engineering, 2010,6(6):45-48.(in Chinese))
- [14] 毛德华, 邹君, 李杰. 基于遗传算法的投影寻踪方法在洞庭湖区洪涝灾害易损性评价中的应用 [J]. 冰川冻土, 2010,32 (2):389-396. (MAO Dehua, ZOU Jun, LI Jie. Application of projection pursuit method based on genetic algorithm to vulnerability evaluation of flood disasters [J]. Journal of Glaciology and Geocryology, 2010,32 (2):389-396. (in Chinese))
- [15] Baker D G, Ruschy D L, Skaggs R H et al. Air temperature and radiation depressions associated with snow cover[J]. Appl Meteor, 1992,3(1):247-254.
- [16] 王金叶, 常宗强, 金博文, 等. 祁连山林区积雪分布规律调查[J]. 西北林学院学报, 2001,16(S1):14-16. (WANG Jinye, CHANG Zongqiang, JIN Bowen, et al. Investigation on distribution rule of snow pack in Qilian mountains forest region[J]. Journal of Northwest Forestry University, 2001,16(S1):14-16. (in Chinese))
- [17] 陈曜, 丁晶, 赵永红. 基于投影寻踪原理的四川省洪灾评估[J]. 水利学报, 2010,41(2):220-225. (CHEN Yao, DING Jing, ZHAO Yonghong. Assessment on flood disaster in Sichuan Province based on principle of projection pursuit method [J]. Journal of Hydraulic Engineering, 2010,41(2):220-225. (in Chinese))
- [18] Friedman J H, Tukey J W. A projection pursuit algorithm for exploratory data analysis[J]. IEEE Trans. on Computer, 1974,23(9): 881-890.

Application of Projection Pursuit Method in Building Snow Depth Model

HU Shujuan¹, LEI Xiaoyun¹, BAI Lei², XING Kun¹, LIU Yi¹, LV Jiaojiao¹

(1. Xinjiang Agricultural University, Urumqi 830052, China;

2. Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China)

Abstract: In the snow melt period, snow depth variation has the features of high dimension, nonlinear and non-normal distribution. In this study, the multi index problem on snow depth change was transformed into the problem of single projection index by the projection pursuit method with optimized the projection direction by genetic algorithm. So the snow depth change dimension has been reduced with the method. The result shows that the factors effect sequence on snow depth dynamic change follows: the ground temperature > net radiation > air temperature > soil moisture > wind speed > relative humidity > total precipitation > vapor pressure. MLP neural network was used to build the snow depth model, which can simulate snow depth. The result shows that using projection pursuit method to get the snow depth of major impact factor is reliable, and major impact factor can simplify the snow depth model.

Key words: projection pursuit; genetic algorithm; snow depth; model; key factor