

稀疏样本下冬春季月平均气温空间插值研究

——以新疆玛纳斯河流域为例

杨耘^{1,2}, 李陇同¹, 刘艳³, 刘帅令¹, 王彬泽¹, 王丽霞¹, 程雪¹

(1. 长安大学地质工程与测绘学院, 陕西西安 710054; 2. 地理信息工程国家重点实验室长安大学合作部, 陕西西安 710054; 3. 中国气象局乌鲁木齐沙漠气象研究所, 新疆乌鲁木齐 830002)

摘要: 针对我国典型高寒山区——新疆天山中段玛纳斯河流域积雪-融雪过程模拟中气温空间数据的制备问题, 以气象站点稀少的玛纳斯河流域为研究区域, 利用最小二乘相关分析法开展了冬、春季(2015年11月-2016年4月)气温环境变量分析, 通过共线性检测确定了纬度、海拔、坡度、坡向、NDVI 5个环境变量组成了最优因子集, 构建了基于广义回归神经网络(GRNN)的月平均气温空间插值模型。采用区域内139个站点中的119个观测站点数据作为训练数据对GRNN模型进行训练, 确定了冬、春季6个月的区域气温空间插值模型。利用剩余的20个观测站点数据作为检验样本, 以均方根误差(RMSE)和平均相对误差(MRE)为评价指标, 对模型的回归误差进行分析。结果表明: 本模型6个月的平均RMSE值为1.46, 优于传统的地理加权回归克里金(GWRK)方法(其平均RMSE值为2.22)。此外, 从不同月份的气温空间插值分布图来看, 本文模型空间插值后的气温变化趋势与实际变化趋势一致。从气温的空间分布情况来看, 各空间点的气温与其海拔高程呈正相关, 且随地表覆盖类型变化。这也表明本文提出的插值策略并组合建立的GRNN模型对于稀疏气象站点条件下的气温空间插值精度更高, 一致性较好。

关键词: 广义回归神经网络(GRNN); 气温; 空间插值; 稀疏样本; 玛纳斯河流域

中图分类号: P468.0⁺21

文献标识码: A

文章编号: 1672-643X(2020)01-0248-06

Spatial interpolation of monthly average air temperature during winter – spring season using sparse samples: A case study in Manas River Basin in Xinjiang

YANG Yun^{1,2}, LI Longtong¹, LIU Yan³, LIU Shuailing¹, WANG Binze¹,
WANG Lixia¹, CHENG Xue¹

(1. College of Geological and Surveying & Mapping Engineering, Chang'an University, Xi'an 710054, China; 2. Department of Cooperation, Chang'an University, State Key Laboratory of Geo-information Engineering, Xi'an 710054, China; 3. Urumqi Institute of Desert Meteorology, China Meteorological Administration, Urumqi 830002, China)

Abstract: For preparation of air temperature spatial data used in the simulation of snow cover to snow melting process in the Manas River Basin with sparse meteorological stations in the middle part of Tianshan Mountains, Xinjiang, this paper carried out the analysis of environmental variables influencing air temperature in winter and spring (2015-11 – 2016-04) using least square correlation method so as to determine the optimal set of environmental variables through collinearity detection. Here the optimal factor set is composed of five environmental variables: latitude, elevation, NDVI, terrain slope and aspect. Then the Generalized Regression Neural Network (GRNN) model was constructed for spatial interpolation of monthly mean air temperature. Finally, the proposed GRNN model was trained with observation data of 119 stations among the total 139 in the region, and the spatial interpolation model of monthly air temperature for the six months in winter and spring were determined. Also the regression error of the proposed

收稿日期: 2019-08-01; 修回日期: 2019-10-11

基金项目: 长安大学中央高校基本科研业务费专项(300102269205、300102269201、300102269304); NSFC-新疆联合基金项目(U1703121); 国家自然科学基金项目(41301386)

作者简介: 杨耘(1975-), 女, 新疆沙湾人, 博士, 副教授, 硕士生导师, 主要从事时空遥感变化分析。

通讯作者: 李陇同(1994-), 女, 硕士研究生, 主要从事气候环境时空变化分析。

model was analyzed with Root Mean Square Error (RMSE) and the Mean Relative Error (MRE) as measures using the rest observation data of 20 stations as test samples. Results show that the averagely RMSE value of this model is 1.46 in six months, which is superior to the traditional GWRK method with an average RMSE value of 2.22. In addition, from the spatial interpolation maps of air temperature of 6 months, the varying trend of the interpolated air temperature using our model is consistent with that of actual circumstances. The temperature of each spatial site is positively correlated with its elevation, and varies with the type of surface coverage. In summary, the proposed GRNN model with combination of interpolation strategy has shown an improved interpolation accuracy and better spatial consistency for spatial interpolation of air temperature even if few meteorological observation stations were provided.

Key words: generalized regression neural network (GRNN); temperature; spatial interpolation; sparse sample; Manas River Basin

1 研究背景

大气温度作为一个衡量地球表面能量平衡及温室效应的很重要的指标,也是影响融雪径流的重要环境变量之一。新疆北疆的玛纳斯河流域地处高寒山区,海拔高、积雪多,春季易引发融雪性洪水,因此,该地区的气温等气象空间数据的制备直接影响到积雪地区的融雪径流量估算的准确性,对防洪抗旱、合理开发利用水资源以及新疆农业经济的可持续发展也具有十分重要的意义^[1-2]。因此,如何获取该地区的气温空间分布,为积雪融雪径流分析提供更准确的基础数据是当前面临的科学问题。传统的气温空间插值方法主要采用克里金插值、反距离插值法等^[3-5]。诸多学者也基于地理加权回归与克里金插值组合算法(GWRK)开展了空间插值研究^[6]。但该研究区地处高寒山区,且气象观测站分布稀疏,传统的插值算法主要表现出以下问题:(1)利用常规的气象数据空间插值法精度偏低且不均匀^[7]。(2)前人对融雪期气温环境变量的选取比较单一,导致了插值精度不高^[8]。(3)传统的空间插值模型,同时需要大量的训练样本,才能达到较高的精度,不适合于稀疏气象观测数据条件下的地区。

众多研究表明,机器学习在复杂的非线性问题中表现出明显的优势^[9-10]。而与其他神经网络相比,广义回归神经网络(GRNN)具有更好的非线性映射的能力,并且学习速率快,对于不稳定数据的处理和小样本数据的回归分析具有自己独特的优势^[11-12]。但目前该模型在气温空间分布插值方面的应用研究较少。在此背景下,本文以新疆玛纳斯流域为研究区,开展了影响气温的环境变量分析及优化,以及基于多变量组合的 GRNN 月平均气温空间插值研究,以实现 2015 年 11 月-2016 年 4 月冬春季气温融雪因子的空间数据制备。

2 数据来源与研究方法

2.1 研究区概况

本文选取中国典型高寒山区的玛纳斯河及周围区域作为研究区。玛纳斯河发源于天山北麓,全长约 450 km,是准噶尔盆地最长的内陆河。流域地形环境复杂,南高北低,众多支流汇合经肯斯瓦特水文站流进平原,海拔从 5 000 m 以上降到 500 m 左右。日温差较大,年均气温 6.6℃,平原区年降水量为 110~200 mm,大于 10℃年积温 2400~3500℃。山地垂直地带性特征明显,流域上游水流湍急,下游水流平缓。高山区冰雪融水与上游降水成为玛纳斯河主要径流补给,季节性积雪对玛纳斯河春季径流量起着重要的调节作用,成为重要灌水资源,是新疆融雪性洪水高发区域之一。

2.2 数据来源

由于玛纳斯河流域内只有少量气象站点,因此,本文选取包含玛纳斯河及其周边面积约为 750 km²的区域为研究区,区域内有国家及地方设立的气象观测站点共 139 个,如图 1 所示。

图 1 中,蓝色点表示气象观测站点。红色曲线圈定区域为玛纳斯河流域,该流域内观测站点稀少,仅 2 个观测站,而该流域周边的区域内气象站点较多,研究区上部站点分布相对密集,占总观测站数量的五分之三,但分布不均匀。就土地覆盖类型来看,研究区西南部多为山地和森林,站点分布较少且分布不均匀,而东南部也多为山地,但气象站点分布较东南部稍多。

通过对研究区内 139 个站点的日气温数据统计计算得到了每个站点的月平均气温,数据精度为 0.1℃。从网上下载了分辨率 1 km 的 MODIS 月植被指数数据产品 MOD13A3。从地理空间数据云下载了分辨率为 90 m 的 DEM 数据,对 DEM 数据进行

重采样到栅格大小为 $1\text{ km} \times 1\text{ km}$, 然后进行坡度、坡向因子的提取, 得到了分辨率为 1 km 的研究区域坡度坡向分布图。MODIS 月植被指数数据产品需要进行重投影、拼接、转换、裁剪等操作。

海拔)、地形(坡度、坡向)、地表覆盖类型等因子会对气温的空间分布有较大影响; 其次, 不同纬度的地区获得的太阳辐射能量及气温不同。气温与坡向关系表现在: 迎风坡空气湿润, 温度低; 背风坡, 空气干燥, 温度高。而气温与海拔高程呈明显的负相关。气温与下垫面也有关, 植被覆盖区的气温低, 反之气温高。但是不同的研究区域环境变量对该区域温度的影响略有不同, 因此, 本文开展了气温环境变量的分析, 见表 1。

研究区各空间点的海拔高程、坡度、坡向、归一化植被指数这 4 个环境变量图如图 2 所示。

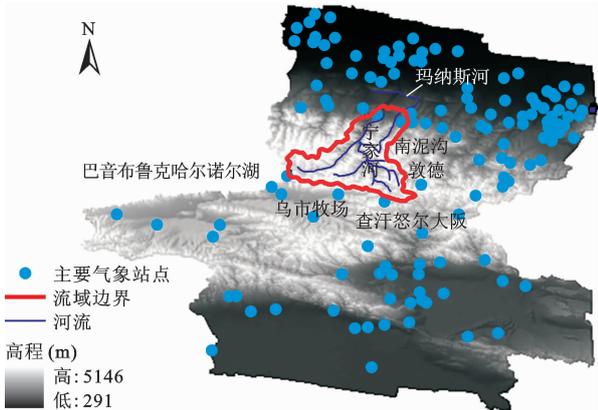


图 1 玛纳斯河流域及其周边区域气象站点分布

表 1 影响气温空间变化的环境变量

变量类型	变量名称
地理位置变量	经度
	纬度
	海拔高程
地形环境变量	坡向
	坡度
地表覆盖变量	归一化植被指数

3 气温环境变量分析及优化

3.1 气温的环境变量分析

已有研究表明, 研究区地理位置(经度、纬度及

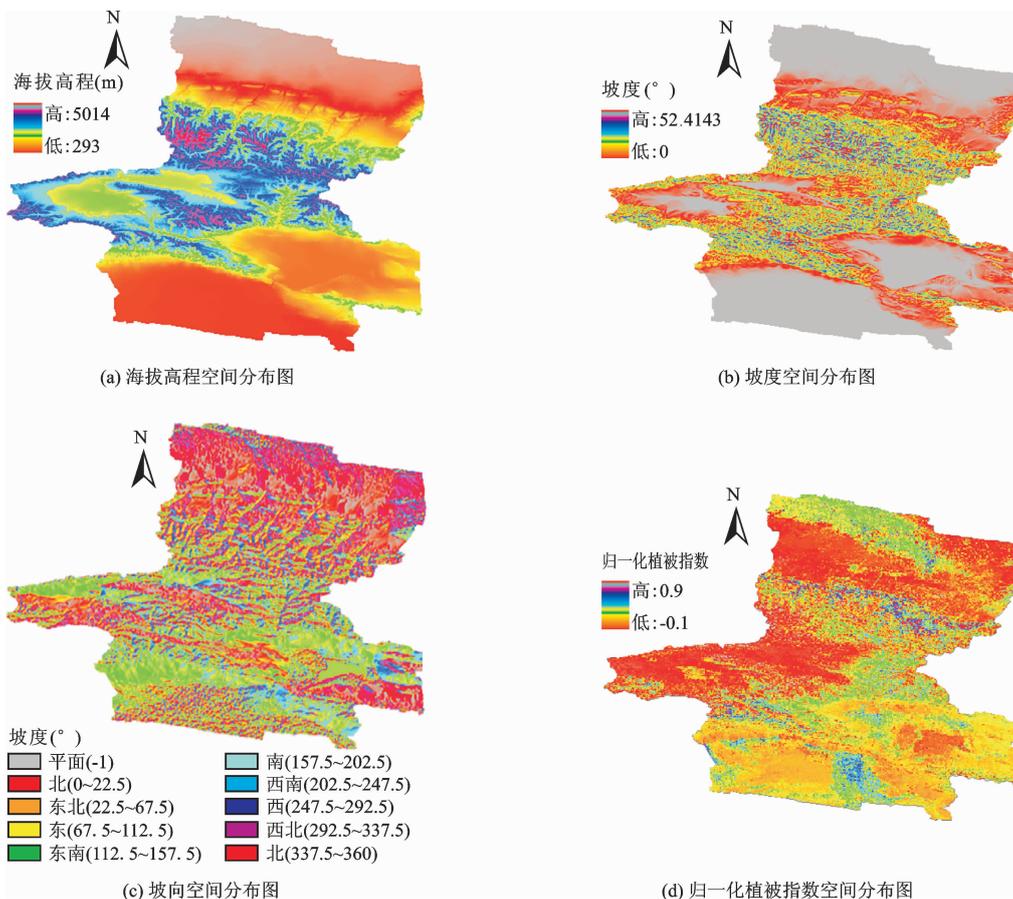


图 2 影响气温的环境变量空间分布图

从图 2(a)可以看出,该地区地形中部多为山地,海拔比较高,而东南部和西南部及该区域的北部海拔相对较低。图 2(b)表明,坡度变化区间为 0 ~ 52.41°,研究区域中部为山地所以坡度较大。分析图 2(d),该区域植被覆盖率较低,结合其他影像可知荒漠在这一区域占有很大的面积,植被覆盖类型多为草原,NDVI 值在 -0.1 ~ 0.9 范围内变化。通过影响气温的环境变量空间分布图,可以清晰地看到环境变量的分布变化趋势,从而直观地判断出其对温度的影响变化。

3.2 气温与其环境变量的相关性分析

利用 SPSS 软件对上述环境变量与气温的相关性能(决定各环境变量与气温相关性关系大小的系数)进行分析,见表 2。根据这些环境变量与气温的相关性分析结果发现,气温与海拔高程、地形坡度、坡向、植被指数和纬度具有较强的相关性,与经度的相关性较弱。为了避免环境变量间的共线性对气温插值的不利影响,选取与气温相关性较强的 5 个变量进行共线性分析。

表 2 环境变量的相关性分析及共线性检验

序号	说明	相关性值	VIF
1	海拔高程	0.151	4.657
2	纬度	0.171	1.258
3	坡度	0.551	1.122
4	坡向	0.651	4.337
5	归一化植被指数	0.160	1.473

经相关性分析,经度与气温空间分布相关性值仅为 0.006。而其它 5 个环境变量与气温的相关性大于 0.15,且它们的方差膨胀因子均小于 10。考虑到上述 5 个环境变量间相关性相对较高,且不存在共线性。因此,本文选取表 1 中 5 个环境变量作为最优环境变量集。

4 基于 GRNN 模型的气温空间插值

4.1 GRNN 插值模型的建立

GRNN 具有较强的非线性映射能力和学习速度,样本数据少的情况下也能得到较好的插值结果。网络还可以处理不稳定数据,适于复杂的映射问题。其结构如下图 3 所示。

本文建立了包括输入层、模式层、加和层和输出层的 4 层 GRNN 网络拓扑结构进行研究区气温空间插值。输入层的神经元数目与学习样本中输入向量

的维数相同,各神经元直接将输入的变量传递到模式层。模式层的神经元个数为训练样本数 119,在求和层中对两类神经元进行求和并将模式层与各神经元的连接权值设定为 1。在 GRNN 网络模型的训练过程中,径向基核函数的分布密度值是影响插值结果的一个重要参数。通常,分布密度值越小,神经网络对样本数据的逼近性就越好;而其值越大,逼近过程就越光滑。为了保证网络对样本的逼近性误差最小,也兼顾网络对样本的逼近过程的平滑,本文选择循环式交叉训练法对径向基函数的分布密度值进行了选取。

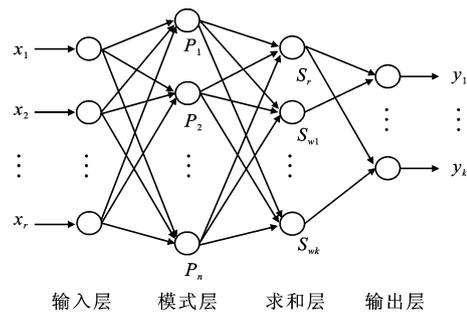


图 3 GRNN 结构示意图

4.2 月平均气温空间插值结果与对比分析

考虑到该研究区观测站点少且分布不均匀的情况,本文从玛纳斯河流域及其外围区域分布的 139 个气象站点中选取 119 个站点的气温数据作为训练样本,对 GRNN 模型进行训练。最后,以各空间点的最优环境变量观测值为输入参数,利用训练好的 GRNN 神经网络模型对 2015 年 11 月 - 2016 年 4 月期间每个月进行了气温空间分布制图,见图 4。

从图 4 中可以看出,11 月平均气温变化范围为 -7.415 ~ 6.295°C,12 月平均气温变化范围为 -19.868 ~ -3.803°C,1 月平均气温变化范围为 -20.439 ~ -6.940°C,2 月的气温变化范围为 -22.020 ~ 0.636°C,3 月平均气温变化范围为 -16.773 ~ -0.261°C,4 月平均气温变化范围为 2.147 ~ 19.727°C。2015 年 11 月到 12 月平均气温呈现显著下降趋势,到 2016 年 1 月和 2 月最低气温持续下降但是平均气温回暖,3 月和 4 月的月平均气温明显上升,其时间变化趋势与实际的气温变化一致。从空间变化来看,处于不同地理位置的气温与海拔高度呈明显的负相关。

此外,与 GWRK 方法的空间插值结果相比发现,GRNN 模型插值后的空间分布图更为平滑,避免了 GWRK 方法导致的突变现象。

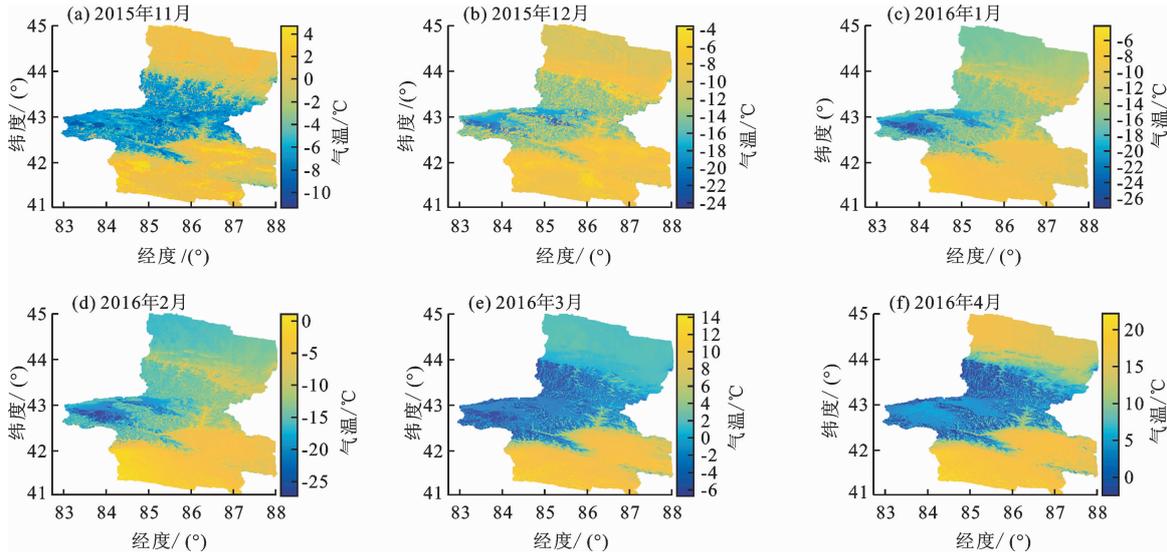


图4 基于 GRNN 模型的 2015 年 11 月 - 2016 年 4 月期间月平均气温空间分布图

用选取的剩余 20 个观测站点的数据作为检验样本,采用均方根误差 (*RMSE*) 和平均相对误差 (*MRE*) 指标对插值结果进行精度评价,并与 GWRK 插值结果进行对比,见表 3。

表3 GWRK 与 GRNN 模型的插值误差对比

时间/ (年-月)	GWRK 方法		GRNN 模型	
	<i>RMSE</i>	<i>MRE</i>	<i>RMSE</i>	<i>MRE</i>
2015-11	1.765	1.302	1.3032	1.0799
2015-12	1.117	1.161	1.1109	1.0799
2016-01	3.622	2.070	1.5841	1.2963
2016-02	1.615	1.349	1.5297	1.1239
2016-03	3.281	1.923	1.9821	1.6917
2016-04	1.918	1.430	1.2715	1.1132

由表 3 可以看出,本模型 6 个月的平均 *RMSE* 值为 1.46,优于传统的 GWRK 方法(其平均 *RMSE* 值为 2.22),而本文 GRNN 模型的平均 *MRE* 值也比 GWRK 方法低 0.31。为了更直观地反映 GRNN 模型插值误差的总体趋势,图 5 给出 GRNN 模型残差分布图,并与 GWRK 方法进行了对比。

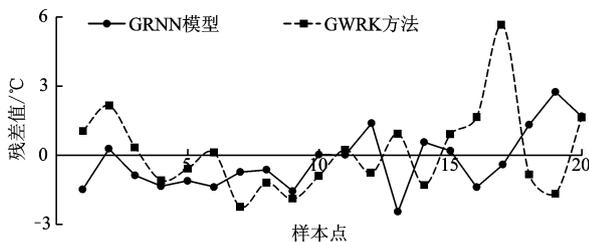


图5 两种气温空间插值结果的残差对比

由图 5 可以看出,本文建立的 GRNN 模型残差更趋于 0,总体小于 GWRK 方法的残差。其次,在湖泊或者水库附近 GWRK 方法的插值误差高于本文模型,这表明地表覆盖类型对气温空间插值结果影响较大。

综合以上分析,本文建立的 GRNN 模型可实现更高精度、更可信的空间插值结果。

5 结论

为了实现新疆天山中段玛纳斯河流域积雪-融雪过程模拟中气温融雪环境变量空间数据制备,本文引入 GRNN 模型,开展了研究区气象站点稀疏、观测数据较少且分布不均匀的条件下的冬、春季节 6 个月的 GRNN 气温空间插值模型的研究。通过对研究区时空插值结果的分析,得出如下结论:与传统的 GWRK 插值方法相比,在观测站点少且分布不均匀的情况下,本文模型的气温空间插值精度高于 GWRK 方法,且气温空间插值结果更平滑,缓解了传统插值法出现的不平滑现象。其次,从模型的 *RMSE* 分布可以看出,GRNN 模型的插值误差曲线波动较小,表明该模型更稳定,且本文的 GRNN 模型空间插值速度比 GWRK 方法更快。今后还需要引入更多的环境变量并考虑不同时期观测数据的相关性以进一步提高空间插值精度。

参考文献:

[1] 付强,颜培儒,李天霄,等. 冻融期不同覆盖和气象因子对土壤导热率和热通量的影响[J]. 农业工程学报, 2017,33(20):98-105.

- [2] 杨 笛,熊 伟,许吟隆,等. 气候变化背景下中国玉米单产增速减缓的原因分析[J]. 农业工程学报,2017,33(S1):231-238.
- [3] 程 朋,任鸿瑞. 山西高原气温空间插值分析[J]. 河北师范大学学报(自然科学版),2016,40(1):73-79.
- [4] 马秀霞,黄领梅,沈 冰. 陕西省月平均气温空间插值方法研究[J]. 水资源与水工程学报,2017,28(5):100-105.
- [5] 占龙飞,张传江,胡菊芳,等. 基于 DEM 的复杂地形气温空间插值研究[J]. 气象与减灾研究,2018,41(1):57-62.
- [6] 聂 磊,舒 红,刘 艳. 复杂地形地区月平均气温(混合)地理加权回归克里格插值[J]. 武汉大学学报(信息科学版),2018,43(10):1553-1559.
- [7] 崔晓临,程 贇,张 露,等. 基于 DEM 修正的 MODIS 地表温度产品空间插值[J]. 地球信息科学学报,2018,20(12):1768-1776.
- [8] 毕 旭. 湖北省气温和降水的变化特征及其与地理因子的相关性分析[D]. 武汉:华中师范大学,2013.
- [9] ARULMOZHI V, REGHUNADHAN R. Predicting the protein localization sites using artificial neural networks [J]. Journal of Cheminformatics,2013,5(1):46.
- [10] 花晓蕾,唐慧强,张红燕,等. 基于 RBF 神经网络的空间插值法在空气质量监测中的应用[J]. 传感器与微系统,2015,34(1):157-160.
- [11] 王尔申,李兴凯,张芝贤,等. 基于广义回归神经网络
- 的粒子滤波算法研究[J]. 沈阳航空航天大学学报,2014,31(6):54-58.
- [12] 何报寅,张 文,乔晓景,等. 基于 GRNN 模型的悬浮物遥感反演研究——以长江城陵矶段为例[J]. 华中师范大学学报(自然科学版),2014,48(5):737-742.
- [13] 白 琳,徐永明,何 苗,等. 基于随机森林算法的近地表气温遥感反演研究[J]. 地球信息科学学报,2017,19(3):390-397.
- [14] 徐永明,覃志豪,沈 艳. 基于 MODIS 数据的长江三角洲地区近地表气温遥感反演[J]. 农业工程学报,2011,27(9):63-68.
- [15] VANCUTSEM C, CECCATO P, DINKU T, et al. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa [J]. Remote Sensing of Environment,2009,114(2):449-465.
- [16] 赵颖慧,吕泓辰,甄 贞,等. 黑龙江省气象因子插值优化及与落叶松 NPP 相关性分析[J]. 南京林业大学学报(自然科学版),2018,42(3):1-9.
- [17] 何咏琪. 基于遥感及 GIS 技术的寒区积雪水文模拟研究[D]. 兰州:兰州大学,2014.
- [18] 赵万里. 混凝土抗渗性能预测的广义回归神经网络模型及应用[J]. 混凝土,2011(2):46-48+127.
- [19] 吴耀华. 基于 GM-GRNN 的电力系统长期负荷预测[J]. 继电器,2007,35(6):45-48+53.

(上接第 239 页)

- [9] 董启朋,姚海林,卢 正,等. 粗粒土三维细观组构与本构关系研究[J]. 岩石力学与工程学报,2014,33(S2):4334-4342.
- [10] ROTHENBURG L, KRUYT N P. Micromechanical definition of an entropy for quasi-static deformation of granular materials [J]. Journal of the Mechanics and Physics of Solids,2009,57(3):634-655.
- [11] JANG E R, JUNG Y H, CHUNG C K. Stress ratio-fabric relationships of granular soils under axisymmetric stress and plane-strain loading [J]. Computers and Geotechnics,2010,37(7-8):913-929.
- [12] NICOT F, DARVE F, RNVO Group, et al. A multi-scale approach to granular materials [J]. Mechanics of Materials, 2005, 37(9):980-1006.
- [13] LI Xia, YU Haisui. Tensorial characterisation of directional data in micromechanics [J]. International Journal of Solids and Structures,2011, 48(14-15):2167-2176.
- [14] 孙其诚,金 峰,王光谦,等. 二维颗粒体系单轴压缩形成的力链结构[J]. 物理学报,2010,59(1):30-37.
- [15] ODA M. Co-ordination number and its relation to shear strength of granular materials [J]. Soils and Foundations, 1977,17(2):29-42.
- [16] 杨 涵,徐文杰,张启斌. 散体颗粒介质变形局部化宏-细观机制研究[J]. 岩石力学与工程学报,2015,34(8):1692-1701.