

# 基于数据挖掘的气象三要素空间一致性检验方法

张捷<sup>1</sup> 张小龙<sup>2</sup> 陈泊舟<sup>3</sup> 陈子凡<sup>4</sup>

(1 新余市气象局, 新余 338001; 2 延安市气象局, 延安 716100; 3 吴起县气象局, 吴起 716600;  
4 十堰市气象局, 十堰 442000)

**摘要:** 质量控制是确保高质量气象观测数据必不可少的重要环节, 空间一致性检验有助于更加精细化的数据质量控制。经典的空间一致性检验方法假设气象要素分布具有连续性、均匀性特征, 但在复杂天气系统中效果不佳。因此, 基于数据挖掘技术提出了一种新的空间一致性检验方法。将邻近5个台站的温度、湿度和2 min平均风速数据分别作为随机森林回归模型的输入, 模型输出测试台站的预测温度, 通过多次试验计算出预测值与真实值的均方误差。试验表明: 随机森林回归算法在温度、湿度和2 min平均风速上都优于空间回归检验算法。同时随机森林算法运行速度快, 并且不用人工手动设置阈值, 它可以有效地降低算法的时间复杂度, 运行时效可以满足实时业务需求, 这为将随机森林算法应用到气象数据质量控制中提供了有力的支撑。

**关键词:** 数据挖掘, 质量控制, 随机森林, 空间一致性检验

**DOI:** 10.3969/j.issn.2095-1973.2022.00.002

## The Spatial Consistency Test Method for the Three Elements of Meteorology Based on Data Mining

Zhang Jie<sup>1</sup>, Zhang Xiaolong<sup>2</sup>, Chen Bozhou<sup>3</sup>, Chen Zifan<sup>4</sup>

(1 Xinyu Meteorological Service, Xinyu 338001; 2 Yan'an Meteorological Service, Yan'an 716100; 3 Wuqi County Meteorological Service, Wuqi 716600; 4 Shiyan Meteorological Service, Shiyan 442000)

**Abstract:** Quality control is essential and indispensable for high-quality meteorological observation data. Data quality control with higher refinement can be achieved by spatial consistency test method. For the classical spatial consistency test method, meteorological elements have to be distributed in a continuous and uniform manner, hence the method is not effective in complex weather systems. This paper proposes a new spatial consistency test method based on data mining technology. The temperature, humidity and 2-minute average wind speed data of five adjacent stations are used as the input of the random forest regression model, and then the model outputs the predicted temperature of the test station. After multiple tests, the mean square error between the predicted value and the real value is calculated. The research results show that the random forest regression algorithm outperforms the spatial regression test algorithm in temperature, humidity and 2-minute average wind speed. Meanwhile, thanks to its high speed and the automatic threshold setting, the random forest method can effectively reduce the time complexity of the algorithm and meet the real-time operational requirements. These advantages are conducive to the application of random forest algorithm in meteorological data quality control.

**Keywords:** data mining, quality control, random forest, spatial consistency test

### 0 引言

质量控制是确保高质量气象观测数据必不可少的重要环节, 传统的地面气象观测数据质量控制方法包括: 气候学界限值检查、台站极值检查、内部一致性检查、空间一致性检查<sup>[1]</sup>。传统的质量控制方法主要根据天气学、气候学原理, 进行极值检查, 或是以气象要素的时间变化规律、或是各要素内部联系为线索, 进行内部一致性检查。这些传统方法只能发现明显异常的数据, 但对于与正确观测数据相差不大(如

温度相差3℃以内)的数据效果并不理想<sup>[2]</sup>。随着计算机技术的发展, 空间一致性检验技术应当被应用于地面自动站数据质量控制流程中<sup>[3]</sup>。

经典的空间一致性检查方法假设气象要素分布具有连续性、均匀性的特点, 将某一观测站的观测资料与其周围其他站点的观测资料进行对比分析, 从而判断其数据是否正常。常见算法有: 反距离加权插值法、空间回归检验法、Madsen-Allerup方法。反距离加权插值法需要知道目标台站与其他台站的距离, 并且反距离加权插值法只考虑了两点之间的距离因素<sup>[4]</sup>, 忽略了地形对天气模式的影响。Madsen-Allerup方法, 假设目标台站与参考台站所处的区域内要素的空间分

收稿日期: 2022年4月15日; 修回日期: 2022年9月22日  
第一作者: 张捷(1997—), Email: 438676610@qq.com

布是均一的<sup>[5]</sup>, 不适用于山区地形。空间回归检验法考虑了邻近台站数据的相关性, 比反距离加权插值法具有更好的检验效果<sup>[6]</sup>, 但是计算过程比较复杂。近年来, 许多研究人员对空间一致性检验提出了诸多新的方法。何志军等<sup>[7]</sup>提出了一种四方位空间一致性检验方法, 对雷暴等小尺度天气系统具有良好的检测性能, 但是这种方法也需要使用到台站的空间位置信息。

随着大数据时代的来临, 越来越多的学者将数据挖掘技术应用到气象数据质量控制当中。数据挖掘是从庞大的数据集中提取有用的数据或信息, 总结这些数据和信息的内在联系, 提出一种简单易懂的数据模型方法<sup>[8]</sup>。周欢乐<sup>[9]</sup>利用关联分析算法对历史气象观测数据学习出一套关联规则, 不满足关联规则的新数据将被判定为异常数据。侯飙<sup>[10]</sup>提出了一种基于粒子滤波的神经网络气象数据质量控制方法, 相对于传统方法性能有所提高。李涛等<sup>[11]</sup>使用关联分析方法提出了一种气象观测设备一致性检测模型, 该方法能够挖掘出所有相关性很强的规则, 还可以利用规则匹配的方法对设备之间进行一致性检测。

数据挖掘技术专注于研究数据之间的相关性而不是因果性, 得到的结果是模糊的而不是精确的<sup>[12]</sup>。因此, 在现有的观测数据无法满足业务需求、布设新的探测设备代价太高或无法布设、尚未完全掌握大气运动规律、无法通过理论推演和方程计算准确预报天气或气候的情况下, 数据挖掘技术不失为一种退而求其次的选择<sup>[13]</sup>。

## 1 算法介绍

### 1.1 随机森林

随机森林算法是在决策树算法<sup>[14]</sup>的基础之上所发展起来的一种算法。CART算法<sup>[15]</sup>是一种经典的决策树算法, CART树既可用于解决分类问题, 也可用于解决回归问题。

2001年, Breiman提出了使用多个决策树来进行投票决策的随机森林算法<sup>[16]</sup>。随机森林算法以随机抽取的方式从整个训练数据集中抽取出一个子集用于生成一棵决策树, 多次重复上述步骤就可以获得多个形态不同的决策树, 最终获得一个决策树森林。决策树生成的随机性以及森林决策的组合性使得随机森林算法相比较于单棵决策树具有更好的泛化能力, 可以有效避免过拟合问题。

### 1.2 K-means 聚类算法

K-means 聚类算法是聚类算法中最经典的算法, 被广泛研究和应用于不同的学科领域, 并发展出大量不同的改进算法。虽然K-means 聚类算法被提出已经

超过50年了, 但目前仍然是应用最广泛的划分聚类算法之一<sup>[17]</sup>。K-means 算法的主要思路如下: 选择K个初始质心, 遍历数据列表, 将每个点指派到最近的质心, 从而形成K个簇。重新计算每个簇的质心, 并重复指派和更新步骤, 直至簇成员不再变化或质心不再变化<sup>[18]</sup>。

K-means 算法中的K是事先给定的, 然而K值的选定有时候是难以估计的。很多时候, 研究人员事先并不知道数据集应该分成多少个类别才最合适。这是K-means 算法的一个不足, 许多学者针对这个问题提出了诸多的改进。Rousseeuw<sup>[19]</sup>提出使用轮廓系数来评价聚类效果的好坏, 平均轮廓系数越大, 表明聚类效果越好, 每个簇之间的分离性越好。虽然这种方法并不总是有效的, 但可以帮助研究者洞察数据中簇的个数。

### 1.3 Apriori 关联算法

Apriori 关联算法是一种最有影响的挖掘关联规则的频繁项集算法。该算法的基本思想如下: 首先找出所有的频繁项集, 这些项集出现的频繁度至少和预定义的最小支持度一样。然后由频集产生强关联规则, 这些规则必须满足最小支持度和最小可信度<sup>[20]</sup>。Apriori 关联算法可以用于发现隐藏在大型数据集中的有意义联系, 所发现的联系可以用关联规则或频繁项集的形式表示。

## 2 试验描述

本次试验使用四川省内145个国家气象站(图1)2020年11月的逐小时观测数据, 试验数据全部来源于中国气象数据网。四川省位于中国内陆西南部, 地貌以山地为主, 具有山地、丘陵、平原和高原四种地貌类型, 地形复杂, 难以对省内各个站点进行精细的天气系统划分。本次试验尝试使用聚类分析与关联分析去发现这145个台站数据之间具有怎样的联系。

### 2.1 四川省各地区地面观测气象数据的相关性研究

从试验数据中随机选取4个时刻的观测数据作聚类分析, 其中横坐标为聚类数量, 纵坐标为平均轮廓系数, 平均轮廓系数越接近1, 表明聚类效果越好。从图2中可以看出, 无论在哪个时刻, 当聚类数量为2时, 平均轮廓系数最大, 此时的聚类效果最好。这表明四川省的地面气象观测数据应当被划分为两个类别, 这正好对应于四川省内的川西高原和四川盆地两个区域。

将聚类数量设定为2, 对四川省2020年11月逐小时观测数据进行聚类分析, 得到A(对应四川盆地)、B(对应川西高原)两个聚类数据集。分别对两个聚类数据集做关联分析, 可以得到对应的频繁项数据集。频繁项数据集包含了经常一起出现在聚类数据集集中的

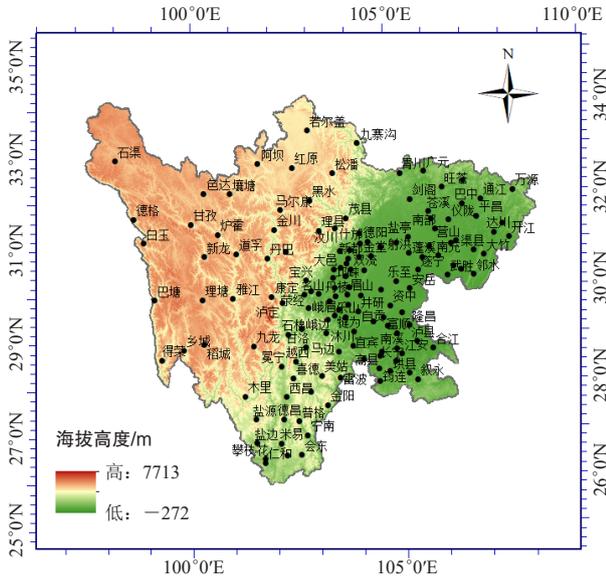


图1 四川省地形及观测点分布

Fig. 1 Topography and distribution of observation sites in Sichuan Province

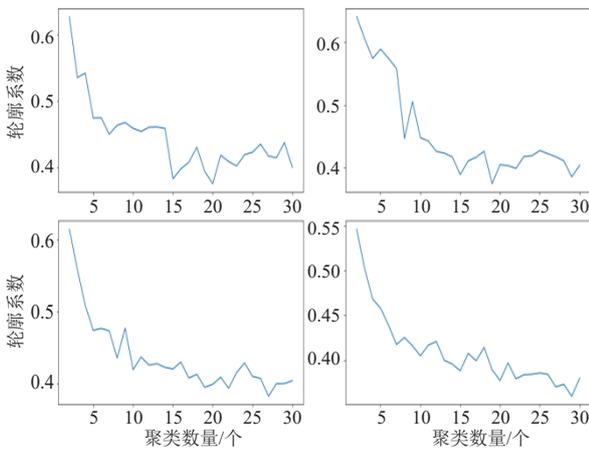


图2 轮廓系数与聚类数量的关系

Fig. 2 The relationship between the silhouette coefficient and the number of clusters

台站名，因此频繁项数据集可以反映出各个台站之间存在的某种相关性。将置信度设定为0.6，对A类数据集进行关联分析并得到频繁项集合 $\alpha$ 。在频繁项集合 $\alpha$ 中，随着维度的增加，该维度下的台站组合数量也随之减少， $\alpha$ 中得到的频繁项组合最高数据维度为6维。

根据站点分布及地形图可以发现，除沐川外，表1中的频繁项组合主要集中于成都周边。结合实际地形来看，这些站点主要位于成都平原。成都平原包括成都市各区县及德阳、绵阳、雅安、乐山、眉山等地的部分区域，地势平坦，四周有群山环抱。Onal等<sup>[21]</sup>指出，相同地形的邻近区域具有相同的气候模式。 $\alpha$ 频繁项集表明成都平原上的自动气象站在同一

表1  $\alpha$ 频繁项集最高维数据

Table 1 Highest dimensional data of  $\alpha$  frequent itemsets

序号	台站位置					
1	大邑	崇州	邛崃	郫县	双流	沐川
2	大邑	崇州	邛崃	郫县	新津	沐川
3	大邑	崇州	温江	邛崃	龙泉驿	郫县
4	崇州	温江	邛崃	龙泉驿	郫县	新津
5	崇州	温江	邛崃	龙泉驿	郫县	沐川
6	大邑	眉山	崇州	邛崃	双流	沐川
7	大邑	眉山	崇州	邛崃	郫县	双流
8	大邑	眉山	崇州	邛崃	郫县	新津
9	大邑	眉山	崇州	邛崃	郫县	沐川
10	大邑	眉山	崇州	邛崃	龙泉驿	郫县
11	眉山	崇州	温江	邛崃	龙泉驿	郫县
12	崇州	芦山	邛崃	天全	蒙经	郫县
13	眉山	崇州	邛崃	郫县	双流	沐川
14	眉山	崇州	邛崃	郫县	新津	沐川
15	大邑	崇州	邛崃	龙泉驿	郫县	新津
16	大邑	崇州	邛崃	龙泉驿	郫县	沐川
17	崇州	邛崃	龙泉驿	郫县	新津	沐川
18	眉山	崇州	邛崃	龙泉驿	郫县	新津
19	眉山	崇州	邛崃	龙泉驿	郫县	沐川

时刻的地面观测数据具有极强的相关性。除成都平原外，其他地区的地形以丘陵为主。受山坡阻隔的影响，其他地区各个台站的地面观测数据的相关性要明显弱于成都平原。值得一提的是，沐川地区的地面观测数据同成都平原地区的观测数据表现出极强的相关性。

对B数据集进行关联分析，得到频繁项集合 $\beta$ 。置信度为0.55时的集合 $\beta$ ，具有的最高维度为3维，且仅有一条数据：[理塘，得荣，巴塘]。川西高原以山地地形为主，地形复杂，地广人稀，观测站点间隔距离亦较大。因此，集合 $\beta$ 中各台站地面观测数据的相关性较弱。

## 2.2 随机森林算法在空间一致性检验上的测试

前一节的试验表明，在四川省的145个国家站中，成都平原内的地面观测数据具有极强的相关性，四川盆地的其余地区次之，川西高原再次之。

本次试验将会分别在三个地区选定一些站点，作为测试台站，将其与周围的地面观测数据用于训练随机森林模型，并做交叉检验。由于地面观测要素中，温度、湿度、风速是比较重要的气象要素，试验中用整点附近的2 min平均风速代替风速，选取邻近5个台站的温度、相对湿度和2 min平均风速数据分别作为回归模型的输入，模型输出测试台站的预测值，通过多次试验计算出预测值与真实值之间的均方误差，试验结果如表2所示。

表2 随机森林算法在空间一致性检验中的均方误差  
Table 2 The mean square error of random forest algorithm in spatial consistency test

序号	测试台站	训练台站	观测要素	均方误差
1	新津	大邑, 崇州, 温江, 邛崃, 龙泉驿	温度	0.147
			相对湿度	0.230
			2 min平均风速	0.172
2	龙泉驿	眉山, 崇州, 邛崃, 郫县, 新津	温度	0.108
			相对湿度	0.219
			2 min平均风速	0.112
3	崇州	大邑, 邛崃, 新津, 温江, 双流	温度	0.201
			相对湿度	0.285
			2 min平均风速	0.189
4	夹江	峨眉山, 洪雅, 青神, 乐山, 峨眉	温度	0.219
			相对湿度	0.375
			2 min平均风速	0.208
5	富顺	荣县, 威远, 自贡, 隆昌, 泸县	温度	0.333
			相对湿度	0.427
			2 min平均风速	0.302
6	南部	苍溪, 阆中, 仪陇, 蓬安, 营山	温度	0.190
			相对湿度	0.317
			2 min平均风速	0.301
7	理塘	稻城, 乡城, 旺格, 道孚, 巴塘	温度	1.159
			相对湿度	1.750
			2 min平均风速	1.233
8	红原	若尔盖, 松潘, 黑水, 阿坝, 马尔康	温度	0.725
			相对湿度	0.961
			2 min平均风速	1.080
9	米易	仁和, 盐边, 攀枝花, 会理, 会东	温度	0.155
			相对湿度	0.533
			2 min平均风速	0.470

结合图1来看, 表2中1、2、3号试验中台站位于成都平原, 4、5、6号试验中台站位于四川盆地的其他地区, 7、8、9号试验中台站位于川西高原。从上表可以看出, 无论是温度、相对湿度还是2 min平均风速, 1~3号试验的均方误差都相对更小, 7~9号试验的均方误差相对较大, 这也印证了上一节试验得到的结论。成都平原地区地形相似, 各个台站大致处于同一个天气系统中, 因此随机森林算法得到的回归误差也较小。其他地区的地形要更为复杂, 相比较而言, 随机森林算法在这些地区得到的回归误差普遍要大一些。

总的来说, 将随机森林算法用于回归分析, 可以获得较好的预测数值。同时, 随机森林算法运行速度快, 并且不用人工手动设置阈值, 它可以有效地降低算法的时间复杂度, 运行时效可以满足实时业务需求, 这为将随机森林算法应用到气象数据质量控制中提供了支撑。

### 2.3 随机森林与空间回归检验的比较

空间回归检验算法是地面气象观测数据在空间一

致性检验中常用的方法。首先选择一批距离目标站最近的台站, 计算它们与目标站一元回归方程的均方根误差。选取5个具有最小均方根误差的台站作为参考站, 给均方根误差小的参考站以大的权重来计算目标站观测值的加权估计值和估计值的加权标准差。再使用加权估计值和加权标准差确定目标站的观测值应属于的范围, 如果观测值不在这个区间内, 则认为观测值异常。

为了方便对比, 本次试验将空间回归检验的加权估计值与随机森林的回归预测值进行比较。以新津作为目标台站, 大邑、崇州、温江、邛崃、龙泉驿作为训练台站。分别以温度、相对湿度和2 min平均风速作为预测对象进行三次试验, 分别计算三种算法的预测值与观测值的平均均方误差, 试验结果如表3所示。

表3 关于温度、相对湿度、平均风速的预测误差比较  
Table 3 Comparison of prediction errors on temperature, humidity, average wind speed

项目	随机森林	空间回归检验
温度	0.17	5.67
相对湿度	0.54	3.02
2 min平均风速	0.28	0.60

从表3可以看出, 无论是在对温度、相对湿度、还是平均风速的预测上, 随机森林算法的平均均方误差要小于空间回归检验算法。这表明随机森林算法在空间一致性检验中能够比较好的预测目标数据, 这为将随机森林算法应用到气象数据质量控制中提供了支撑。同时也验证了空间回归检验算法对如风速空间变化大的要素具有良好的检验效果。

由于空间回归检验方法在很大程度上受到季节、地理环境、站台密集程度等的影响<sup>[22]</sup>, 所以空间回归检验方法更适用于同一季节、地理环境相似且站台密集的地方, 而随机森林检验方法使用范围更广。

### 2.4 基于空间一致性检验的异常检测

周边台站的地面观测数据可以预测出目标台站在当前时刻的数据, 若目标台站实际采集的数据与预测数据出入较大, 则可以认为目标台站出现异常。借用雷达信号的术语来表示, 当一个正常的的数据被判定为异常时, 则称这种情况为虚警; 当一个异常数据被判定为正常时, 则称这种情况为漏报。

试验以查准率和查全率作为本次试验的评价参数, 查准率表示判定为正常的的数据中有大比例是真实正常的, 查全率表示真实正常的的数据有多大比例被判定为正常。

以新津作为目标台站, 大邑、崇州、温江、邛崃、龙泉驿作为训练台站。以温度、相对湿度、2 min

平均风速为预测数据分别进行三次试验。当真实数据与预测数据差值的绝对值小于阈值时，此数据将被判定为正常，反之将被判定为异常。由于历史数据均是经过质量控制之后的正常数据，本次试验将对历史数据进行采样并为其加上随机噪声，以此来制造异常数据。

表4~表6均表明随着阈值的增大，系统对各项数据的查全率均逐渐增加，查准率均逐渐减少。阈值增大之后，将有更多的正常数据被判定为“正常”数据，所以查全率会逐渐上升。同时，也有更多的异常数据被判定为“正常”，因此查准率将会随着阈值的增大而降低。当温度阈值设置为0.5℃，目标台站新津可以获得92.6%的查准率以及87.6%的查全率。相对湿度阈值设置为5%时，目标台站新津可以获得88%的查准率以及92%的查全率。需要指出的是，对2 min平均风速进行异常检测所得到的查准率较低。这表明随机森林算法在对风速进行回归预测时准确度并不理想，事实上风速这一要素的空间相关性较低<sup>[23]</sup>，因此随机森林算法对风速数据进行空间一致性检验时效果差强人意。

表4 对温度异常的检测

Table 4 Anomaly detection of temperature

阈值/℃	查准率	查全率
0.2	0.964	0.532
0.3	0.968	0.690
0.4	0.936	0.816
0.5	0.926	0.876
0.6	0.890	0.880

表5 对相对湿度异常的检测

Table 5 Anomaly detection of relative humidity

阈值/%	查准率	查全率
2	0.960	0.490
3	0.958	0.690
4	0.934	0.770
5	0.880	0.920
6	0.734	0.926

表6 对2 min平均风速异常的检测

Table 6 Anomaly detection of 2-minute average wind speed

阈值/( m/s )	查准率	查全率
0.4	0.631	0.56
0.5	0.627	0.68
0.6	0.603	0.733
0.7	0.605	0.773
0.8	0.580	0.840

### 3 总结与展望

本次试验所研究的基于数据挖掘的气象三要素空

间一致性检验方法相较于现有的数据质量空间一致性检查方法来说，具有更高的精确度，且运行速度快，不用人工手动设置阈值，有效地降低算法的时间复杂度，运行时效可以满足实时业务需求，对数据挖掘技术应用在气象数据质量控制上的研究具有深远意义。

1) 通过聚类分析和关联分析发现可将观测数据粗略划分为两个大类，大致对应于川西高原与四川盆地这两大地形板块。而后通过比较这两种地形的三要素预测误差，验证了四川盆地的地形划分方法。

2) 由于成都平原地形平坦开阔，气象观测数据的空间相关性高于其他地区，预测的误差较小；而川西高原以山地、高原为主，地形复杂，此地区内地面气象观测数据的空间相关性低于其他地区，预测的误差较大。因此，地形划分对空间一致性判断结果起很大作用。

3) 将空间回归检验的加权估计值与随机森林的回归预测值进行比较，无论是在预测温度、相对湿度还是平均风速上，随机森林算法的均方误差远小于空间回归检验算法，这表明随机森林算法比空间回归检验算法具有更高的预测准确度。同时随机森林算法运行速度快，并且不用人工手动设置阈值，它可以有效地降低算法的时间复杂度，运行时效可以满足实时业务需求，这为将随机森林算法应用到气象数据质量控制中提供了有力的支撑。

4) 本文提出的异常检测方法在温度和相对湿度两个要素上具有较好的检测性能。但对于风速这一类空间相关性较低的气象数据，检测性能则不尽如人意。

虽然本研究取得了一定的进展与成果，但并未使用完整一年或者一年中不同季节、月份的数据来进行试验，后续将继续对基于数据挖掘的质量控制方法的季节、极端天气过程的适应性进行研究。

#### 参考文献

- [1] 韩江益, 王立俊, 羊清雯, 等. 南海区域站气象资料质量控制方法改进. 气象科技, 2018, 46(3): 474-478, 496.
- [2] 王海军, 刘莹. 综合一致性质量控制方法及其在气温中的应用. 应用气象学报, 2012, 23(1): 69-76.
- [3] 刘小宁, 任芝花. 地面气象资料质量控制方法研究概述. 气象科技, 2005, 33(3): 199-203.
- [4] 范文波. 地面气象观测数据综合质量控制方法研究与实现. 南京: 南京信息工程大学, 2016.
- [5] 熊安元. 北欧气象观测资料的质量控制. 气象科技, 2003, 31(5): 314-320.
- [6] 刘小宁, 鞠晓慧, 范邵华. 空间回归检验方法在气象资料质量检验中的应用. 应用气象学报, 2006, 17(1): 37-42.
- [7] 何志军, 封秀燕, 何利德, 等. 气象观测资料的四方空间一致性检验. 气象, 2010, 36(5): 118-122.
- [8] Tan P-N, Steinbach M, Kumar V. 数据挖掘导论: 完整版. 北京: 人民邮电出版社, 2011.
- [9] 周欢乐. 基于关联规则的地面气象观测数据综合质量控制系统设计

(下转57页)

- (2): 25-30.
- [21] 陈兢, 李润, 关佐恒. 跳台滑雪运动雪板与雪面间摩擦系数研究. 冰雪运动, 1992, (4): 43-45.
- [22] 张才成. 对影响高山滑雪滑行速度因素的分析. 冰雪运动, 2000, (4): 25-26.
- [23] Kennedy F E, Schulson E M, Jones D E. The friction of ice on ice at low sliding velocities. Philosophical Magazine A, 2000, 80(5): 1093-1110.
- [24] Rist M A. High-stress ice fracture and friction. Journal of Physical Chemistry B, 1997, 101(32): 6263-6266
- [25] 路璐, 刘春玲, 刘琳. 滑雪游客感知价值、满意度与行为意向的关系: 以崇礼密苑云顶滑雪场为例. 干旱区资源与环境, 2018, 32(5): 202-208.
- [26] Lozowski E, Szilder K, Poirier L. A bobsleigh ice friction model. International Journal of Offshore and Polar Engineering, 2014, 24(1): 52-60.
- [27] Nachbauer W, Kaps P, Mössner M, et al. Effects of ski and snow properties on the turning of Alpine skis—a computer simulation. Journal of Biomechanics, 2006, 39: S188.
- [28] You J, Tarboton D G, Luce C H. Modeling the snow surface temperature with a one-layer energy balance snowmelt model. Hydrology and Earth System Sciences, 2014, 10(12): 15071-15118.
- [29] Baeurle L, Kaempfer U, Szabo D, et al. Sliding friction of polyethylene on snow and ice: contact area and modeling. Cold Regions Science and Technology, 2007, 47(3): 276-289.
- [30] Steiger R, Mayer M. Snow making and climate change. Mountain Research and Development, 2008, 28(3): 292-298.
- [31] Hamilton L C, Brown C, Keim B D. Ski areas, weather and climate: time series models for New England case studies. International Journal of Climatology, 2010, 27: 2113-2124.
- [32] 王才勇, 曹立春. 全球变暖对雪上赛事的影响及对策. 冰雪运动, 2011, 33(3): 72-74.

(上接48页)

- 计与实现. 南京: 南京信息工程大学, 2016.
- [10] 侯枫. 气象数据质量控制与监控系统的设计. 哈尔滨: 哈尔滨工程大学, 2017.
- [11] 李涛, 郁美辰, 陆正邦, 等. 基于关联规则挖掘的气象观测设备一致性检测算法. 电子测量与仪器学报, 2017, 31(10): 1568-1573.
- [12] 维克托·迈尔-舍恩伯格. 大数据时代: 生活、工作与思维的大变革. 杭州: 浙江人民出版社, 2013.
- [13] 沈文海. 再析气象大数据及其应用. 中国信息化, 2016, (1): 85-96.
- [14] Quinlan J R. C4.5 Programs for machine learning. San Mateo, CA: Morgan Kaufmann, 1992.
- [15] Breiman L, Friedman J H, Olshen R, et al. Classification and regression trees. Encyclopedia of Ecology, 1984, 40(3): 582-588.
- [16] Breiman L. Random forests. Machine Learning, 2001, 45(1): 5-32.
- [17] Anil K. Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 2010, 31 (8): 651-666.
- [18] 王千, 王成, 冯振元, 等. K-means聚类算法研究综述. 电子设计工程, 2012, 20(7): 21-24.
- [19] Rousseeuw P J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 1999, 20(20): 53-65.
- [20] 刘华婷, 郭仁祥, 姜浩. 关联规则挖掘Apriori算法的研究与改进. 计算机应用与软件, 2009, 26(1): 146-149.
- [21] Onal A C, Sezer O B, Ozbayoglu M, et al. Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning// IEEE International Conference on Big Data. IEEE, 2018.
- [22] 张齐东, 熊雄. 空间回归检验方法在地面气象资料质量控制中的应用——以逐时气温资料为例. 内燃机与配件, 2017(10): 152-153.
- [23] 尹婧姣, 江志红, 吴息, 等. 空间差值检验方法在地面气象资料质量控制中的应用. 气候与环境研究, 2010, 15(3): 229-236.

(上接52页)

平均气温直减率排序为八仙观站>南岩站>太子坡站>紫霄宫站。

5) 武当山景区日平均气温 $\geq 0^{\circ}\text{C}$ 与 $\geq 10^{\circ}\text{C}$ 的积温垂直变化趋势同平均气温相近, 二者随着海拔升高而下降, 在低山地区更突出。

#### 参考文献

- [1] 管延龙, 王让会, 李成, 等. 天山北麓1963—2010年0 cm最高与最低地表温度变化特征. 干旱气象, 2015, 33(4): 587-594.
- [2] 王朋岭, 唐国利, 曹丽娟, 等. 1981—2010年青藏高原地区气温变化与高程及纬度的关系. 气候变化研究进展, 2012, 8(5): 313-319.
- [3] 王海军, 张勃, 靳晓华, 等. 基于GIS的祁连山区气温和降水的时空变化分析. 中国沙漠, 2009, 29(6): 1196-1202.
- [4] 林之光, 安顺清, 吴其勋, 等. 武夷山区气温垂直梯度的研究. 气象, 1983, 9(4): 18-19.
- [5] 郑成洋, 方精云. 福建黄岗山东南坡气温的垂直变化. 气象学报, 2004, 62(2): 251-255.
- [6] 黄承标, 卢立华, 温远光, 等. 大青山林区不同海拔高度主要气象要素的变化. 贵州农业科学, 2011, 39(1): 90-95.
- [7] 丁焯毅, 高益波, 胡波, 等. 四明山区域气温的垂直变化特征分析. 浙江气象, 2017, 38(4): 42-44.
- [8] 田杰, 王庆伟, 于大炮, 等. 长白山北坡气温的垂直变化. 干旱区资源与环境, 2013, 27(4): 65-69.
- [9] 张锦斌, 蔡志林, 姚健, 等. 雁荡山区气候垂直变化分析. 浙江农业科学, 2014, 55(3): 403-405.
- [10] 周礼华, 黄力, 钱深华, 等. 金佛山西坡气温的垂直变化特征. 山地研究, 2019, 37(6): 818-827.
- [11] 曾琦, 任国玉. 湖北省主要山系高影响天气指标空间特征分析与建模. 暴雨灾害, 2020, 39(2): 192-200.
- [12] 王海军, 杨志彪, 杨代才, 等. 自动气象站实时资料自动质量控制方法及其应用. 气象, 2007, 33(10): 102-109.
- [13] 任芝花, 张志富, 孙超, 等. 全国自动气象站实时观测资料三级质量控制系统设计. 气象, 2015, 41(10): 1268-1277.
- [14] 范宏飞. 自动气象站资料质量控制系统设计. 吉林大学学报(信息科学版), 2021, 33(4): 470-478.
- [15] 黄嘉佑. 气候状态变化趋势与突变分析. 气象, 1995, 21(7): 54-57.
- [16] 周淑贞. 气象学与气候学(第三版). 北京: 高等教育出版社, 1997: 56-58.
- [17] 杨萍, 肖子牛, 刘伟东. 北京气温日变化特征的城郊差异及其季节变化分析. 大气科学, 2013, 37 (1): 101-112.
- [18] 王岱, 游庆龙, 江志红, 等. 中国极端气温季节变化对全球变暖减缓的响应分析. 冰川冻土, 2016, 38(1): 36-46.
- [19] 熊敏述. 基于集合预报系统的日最高和最低气温预报. 气象学报, 2017, 75(2): 211-222.
- [20] 陈锋, 董美堂, 冀春晓. 综合分析法在复杂地形气温精细格点化中的应用. 高原气象, 2016, 35(5): 1376-1388.
- [21] 武丽梅, 陈璐, 王凯, 等. 中国大陆1983—2012年年极端气温时空变化特征. 干旱气象, 2016, 34(6): 945-951.
- [22] 段春锋, 缪启龙, 曹雯, 等. 以高山站为背景研究城市化对气温变化趋势的影响. 大气科学, 2012, 36(4): 811-822.
- [23] 汪凯, 叶红, 唐立娜, 等. 气温日较差研究进展: 变化趋势及其影响因素. 气候变化研究进展, 2010, 6(6): 417-423.
- [24] 董丹宏, 黄刚. 中国最高、最低温度及日较差在海拔高度上变化的初步分析. 大气科学, 2015, 39(5): 1011-1024.
- [25] 刘少华, 严登华, 翁白莎, 等. 近50 a中国 $\geq 10^{\circ}\text{C}$ 有效积温时空演变. 干旱区研究, 2013, 30(4): 689-696.