

文章编号: 2096-1618(2022)04-0386-06

基于便携式自动气象站数据的插值方法研究

冯岩莹, 卢会国, 蒋娟萍

(成都信息工程大学电子工程学院, 四川 成都 610225)

摘要:便携式自动气象站常设在偏远无人区,受各种因素影响,其观测数据经常有缺测,使用数据插值方法提高数据连续性,可为后续研究打下重要基础。选取设在青藏高原西部无人区的 7 个便携式自动气象站 2020 年 11 月至 2021 年 10 月的气温、气压数据进行插值研究。气温在时间域、气压在空间域的插值方法选用最邻近插值、线性插值、三次样条插值及立方插值法。对于气温的时间插值,选取时间作为自变量,对应的气温值为因变量构建插值函数,结果表明,线性插值法插值效果最好,并且在温度较低的第一季度和第四季度插值精度高于第三季度、第二季度。对于气压的空间插值,选取海拔高度作为自变量,对应的气压值作为因变量构建插值函数,结果表明,立方插值效果最好。气温在空间域的插值方法选用反距离加权法、克里金插值法、样条函数法及基于三角形的三次插值法,通过插值目标站点周围台站的经纬度、气温值构建插值函数,结果表明,克里金插值法明显优于其他 3 种插值方法。

关键词:便携式自动站;气象数据处理;气温插值;气压插值

中图分类号:TP274+.2

文献标志码:A

doi:10.16836/j.cnki.jcuit.2022.04.004

0 引言

便携式自动气象站是一种集数据传感、采集、存储、通信、定位及电源管理等功能为一体无人值守的小型气象站。这些台站通常设在偏远无人区,受周边恶劣环境影响较大,常出现数据缺测现象,限制了气候分析的适用范围^[1]。数据插值是气象数据处理最常用的方法之一,为了使便携式自动气象站的数据更加具有准确性与代表性^[2],对便携式自动气象站的缺测数据进行插值处理,提高数据质量,可为后续研究打下重要基础。

数据插值一直是气象工作的一个重要组成部分,也是科研人员一直研究的热门课题^[3],采用空间插值方法可以有效获取研究区气象数据的空间连续分布^[4-5]。张超等^[6]针对云南区域站大密度站点的插值问题,利用 Surfer 平台自带的插值工具绘制云南区域站降水图,采用 Cressman 插值方法,研究发现 Cressman 的插值结果要优于 Surfer 软件平台自带的插值方法。Tang 等^[7]利用北美连续完整数据集,研究间隙填充优先空间插值的价值,采用 IDW、双线性插值、最邻近插值等方法对北美地区的降水和温度进行插值,结果发现降水对统计插值的改善最为明显,其次是最低气温和最高气温;当站点网络稀疏且采用简单的插值方法时,改进幅度更大。沈艳等^[8]利用最优插值对中国 3 万多个自动站小时降水数据进行空间分析,当站网密度差异较大时,通过确定影响台站数进行空间插值,一方面可以提高站点稠密区的插值效率,另一方面可以提高站点相对稀疏区的插值精度。钟雨萍^[9]对

深圳市风速时间序列和空间进行插值,在时间插值中,构建双层长短期记忆网络模型,实现风速的短期预测;在空间插值中,分别构建 3 个梯度提升树回归模型对反距离权重法、修正反距离权重法、梯度反距离权重法的插值结果进行校正,研究发现梯度提升树回归模型对部分未知区域的插值有校正效果。王艳萍等^[10]使用克里金插值、反距离权重插值和样条插值对宁夏 2018 年气温资料及逆行进行插值分析与评估,结果表明在台站分布稀疏时适合宁夏气温要素的插值方法是反距离权重法,在台站分布密集时适合宁夏气温要素插值的方法是克里金插值法。李叶等^[11]对大兴安岭山脉北段及其周边区域的气温数据进行空间插值,结果表明反距离权重法、普通克里金插值法、样条函数法 3 种气温插值方法可反映气温要素的空间分布情况,但不适合气象台站稀少且地形起伏较大的区域。Qiao 等^[12]针对土壤中污染物复杂的空间分布特征和较强的局部空间变异性,比较了 IDW 法、克里金法与径向基函数的插值效果,发现土壤层级不同,插值方法各有优异。Serrano-Notivol 等^[13]使用 k 近邻回归法对西班牙地区的温度进行估计,创建了网格化的温度数据集。

本文通过对便携式自动气象站的数据进行分析研究,使用多种插值方法对气温、气压要素进行插值处理,并通过分析不同插值方法得到的插值估计数据与真实值间的误差,找到更适合便携式自动气象站数据的插值方法。

1 数据来源

使用的数据来自青藏高原西部无人区 7 个便携式

自动气象站 2020 年 11 月至 2021 年 10 月采集到的气温及气压数据,台站分布相对位置如图 1 所示。其中台站间距离最长为 378.62 km,距离最短为 34.89 km,且 7 个台站均位于高海拔无人区(均大于 4000 m),无公共通信条件,数据传输采用卫星通信,设定每 10 min 进行一次数据传输。由于周围环境恶劣,常出现不能正常上传的情况,获取的数据大部分不均匀,利用插值方法对这类缺失数据进行插值处理,以提高数据的连续性。



图 1 便携式自动气象站的相对位置

2 研究方法

由于气温要素在时间序列上的变化有一定规律,而气压要素在时间序列上的变化规律不明显,所以对气温要素分别进行时间域和空间域的插值,对气压要素进行空间域的插值。气温在时间域的插值方法选用最邻近插值、线性插值、三次样条插值及立方插值法。气温在空间域的插值方法选用反距离加权法、克里金插值法、样条函数法及基于三角形的三次插值法。而气压在空间域的插值方法选用最邻近插值、线性插值、三次样条插值及立方插值法。

2.1 插值方法

最邻近插值法是最简单的插值方法,主要通过寻找最邻近的数据进行插值。线性插值是一种插值函数为一次函数的多项式插值,相比其他插值方法更简单、方便,时间为自变量,温度为因变量,用线段依次连接相邻的点,每条线段内点的高度(即纵坐标)作为插值估计值。三次样条插值是一种分段多项式插值,最初应用在工程上,通过计算一组已知数据得到一条光滑曲线,既保持分段低次插值多项式,又提高了插值函数光滑性,得到的函数在每一个小区间内均为三次多项式。立方插值同样为一种分段多项式插值,但立方插值不仅区间单调,并且将样本点处的一阶导数值设置为两侧割线斜率的加权调和平均值,从而实现简单快速地求解各个样本点的一阶导数值,且在每个子区间内单调,进而使插值后的曲线更加平滑。

空间插值定义为同一时间维度下,对具有空间属

性的气象数据插值到特定的空间位置上。

反距离权重法(IDW)是一种简单常用的空间插值法,通过对采样点进行线性加权决定输出值的大小,输入与输出栅格的距离越远则被赋予的权重越小,对输出值的影响越小。文中根据各台站经纬度信息计算台站间的距离,插值点与样本间距离越大,权重越小,即样本点距离插值点越远,影响也越小。计算公式为

$$S=\frac{\sum_{i=1}^n S_i \frac{1}{d_i^k}}{\sum_{i=1}^n \frac{1}{d_i^k}} \tag{1}$$

其中: S 为插值点的值; n 为样本点个数; s_i 为第 i 个样本点的值; k 为指定的幂值,通常为 $2 \sim 3$; d_i 为第 i 个已知样本点到插值样本点的距离。

克里金插值是一种对已知样本数据加权平均,使估计值与真实值的方差最小并具有相同数学期望的统计学过程。克里金插值法与反距离权重法相似,均通过样本点数据加权预测插值点数据,为确定性插值。计算公式为

$$\hat{Z}(S_0)=\sum_{i=1}^N \lambda_i Z(S_i) \tag{2}$$

其中: $\hat{Z}(S_0)$ 为插值点估计值; $Z(S_i)$ 为第 i 个位置的真实值; N 为样本个数; λ_i 为第 i 个位置处的权重。

样条函数用于空间插值的原理与时间插值相同。为生成恰好经过输入点的平滑表面,样条函数的空间插值法使用可最小化整体表面曲率的数学函数计算估计值。基于三角形的三次插值法是一种多项式插值法,是逐次通过曲线 $\varphi(t)=a_0+a_1t+a_2t^2+a_3t^3$ 的极小点逼近寻找函数 $f(t)$ 的极小点的方法。

2.2 验证方法

气象数据的插值结果误差验证方法分为实际验证法和交叉验证法。本文选取实际验证法,将所有数据集划分为训练序列和验证序列,验证序列中的实测数据不参与插值过程,只对插值方法进行评估,将训练序列称为样本点(样本台站),验证序列称为插值点(插值台站)。根据平均绝对误差(MAE)、均方根误差(RMSE)及插值的估计值与实测值的相关系数(R^2)对不同插值法的效果进行评估。绝对误差确切地表明了估计值偏离实际值的实际大小,平均绝对误差即为一组估计数据的绝对误差平均值,可以更好地反应估计值误差的实际情况。均方根误差可反应估计值和实测值间差异程度,相关系数(R^2)可以反应估计值和实测值间的相关性。平均绝对误差、均方根误差越接近 0,表示插值方法精度越高。相关系数越接近 1,表示估计值与实测值相关性越大,插值方法精度越高。

3 结果分析

气温的高低受多种因素影响,其中太阳辐射可引

起气温明显的日变化。即一天中,早上温度最低,随着太阳上升,大气层吸收太阳辐射,气温逐渐上升,由于此时阳光是斜射到地球表面,穿过的大气层较长,会损失相对较多的热量。随着太阳继续上升,阳光穿过的大气层逐渐变短,热量损失逐渐变少,地表吸收的热量越来越多,气温也就越来越高。下午两点左右,地面空气除了受太阳的短波辐射外,还吸收地面的长波辐射,热量会达到最大值,气温也相应达到日最高气温,之后逐渐下降,到了次日凌晨又会降到日最低气温。根据这种气温随时间的变化规律对气温进行时间插值,时

间作为自变量,对应的气温作为因变量,构造插值函数,对气温在时间域上进行插值。

大气压强在数值上等于单位面积上从大气上界到地表这一段垂直空气柱的重量。气压值的大小与海拔高度、气温等因素有关,海拔高度越高,空气柱的重量越小,气压也就越低,海拔高度越低,空气柱的重量越大,气压也就越高。文中选取的台站均处于高海拔地区,且台站间海拔高度差异大,海拔信息如表 1 所示。因此可将海拔高度作为自变量,对应站点的气压值作为因变量,构造插值函数,对气压在空间域进行插值。

表 1 便携式自动气象站的海拔信息

台站号	2 号	3 号	4 号	5 号	6 号	11 号	13 号
海拔/m	4280	4957	4320	4220	5145	4854	3628

选取插值的验证站点的方法为:对于气温的插值,选取 7 个台站中地理位置相对中心的 3 号台站作为气温插值的目标站点,分别在时间域和空间域进行插值;对于气压的插值,选取 7 个台站中海拔高度处于相对中间位置的 5 号台站作为气压插值的目标站点,进行插值处理。

3.1 不同季度气温插值精度

对 3 号台站的气温进行时间插值。将 2020 年 11 月至 2021 年 10 月的气温数据使用留出法进行划分,划

分比例为 0.67,即全部数据的 2/3 用来构造插值函数,余下的 1/3 用来验证方法的优劣,不参与插值过程。为分析不同季度的插值估计的误差情况,按季度进行插值估计。第一季度、第四季度温度较低,第三季度温度最高,其中第一季度的平均温度为-4.7℃,第二季度的平均温度为 4.2℃,第三季度的平均温度为 13.5℃,第四季度的平均温度为-3.8℃。使用最邻近插值、线性插值、三次样条插值及立方插值法进行插值估计,输入为时间,输出为对应时刻的气温,构建插值函数,插值误差如图 2 所示,精度对比结果如表 2 所示。

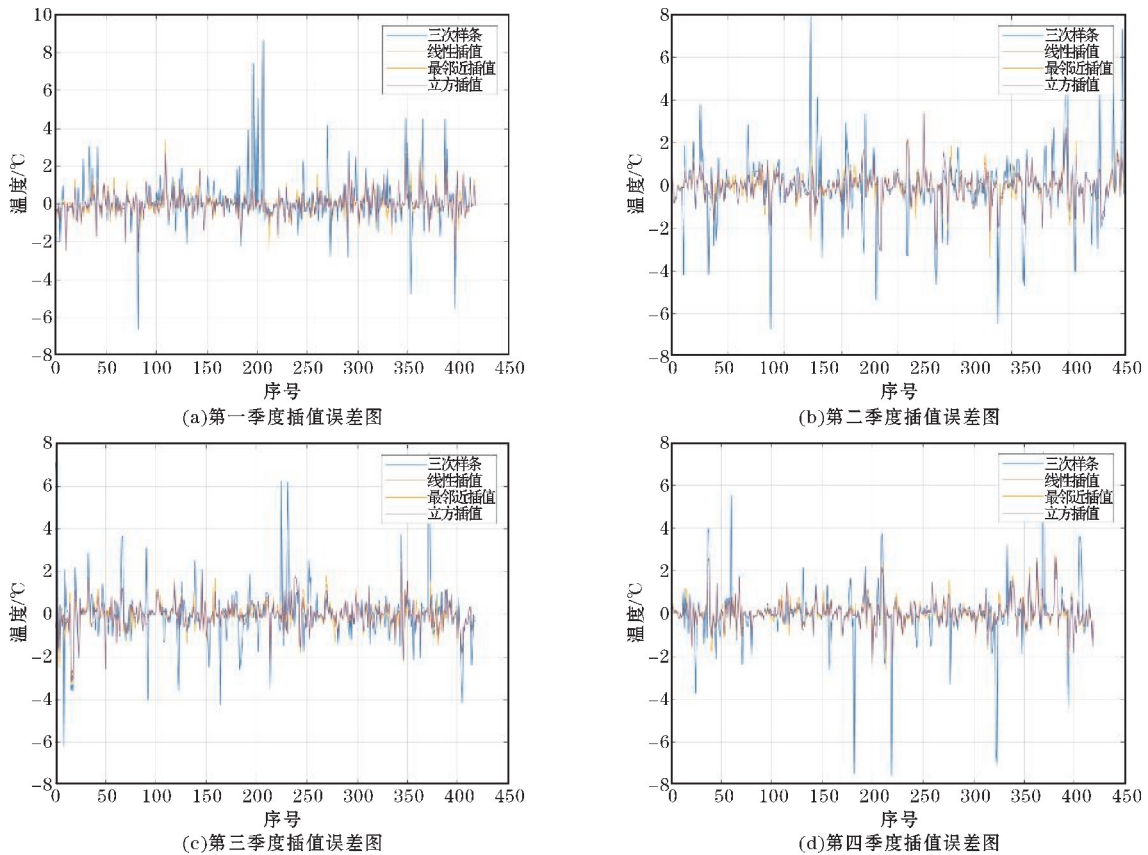


图 2 气温时间插值误差图

表2 气温要素时间插值估计结果检验

插值时间	平均温度/℃	验证方法	最邻近插值	线性插值	三次样条插值	立方插值
第一季度	-4.7	MAE	0.46	0.40	0.68	0.42
第二季度	4.2		0.54	0.49	0.86	0.49
第三季度	13.5		0.49	0.43	0.74	0.43
第四季度	-3.8		0.44	0.38	0.67	0.39
第一季度	-4.7	RMSE	0.71	0.61	1.20	0.64
第二季度	4.2		0.81	0.73	1.45	0.74
第三季度	13.5		0.72	0.64	1.30	0.65
第四季度	-3.8		0.68	0.59	1.28	0.61
第一季度	-4.7	R^2	0.79	0.83	0.61	0.82
第二季度	4.2		0.75	0.79	0.57	0.78
第三季度	13.5		0.76	0.80	0.59	0.79
第四季度	-3.8		0.80	0.84	0.60	0.83

由图2、表2可知,三次样条插值法产生的误差最大,且误差曲线波动幅度最大、最不稳定,线性插值法、最邻近插值法、立方插值法产生的误差波动幅度相对较小。其中最邻近插值法误差在-3.4℃~3.5℃,线性插值法误差在-3.10℃~3.12℃,三次样条法误差在-7.61℃~8.69℃,立方插值法误差在-3.20℃~3.37℃。4种插值方法均在第四季度的插值估计误差最小,第二季度的插值估计误差最大,造成第二季度插值误差最大的原因可能是第二季度的4~6月是天气变化幅度最大的季节,气温的变化幅度相应较大,因此对插值估计结果产生了影响。除三次样条插值法在第一季度的均方根误差比第四季度小外,其他3种插值方法在第一季度的均方根误差均比第四季度大。综合来看,插值方法的效果从优到劣的排序为:第四季度>第一季度>第三季度>第二季度,进而认为低温条件下,有利于以上4种方法的插值。综合表2可知,

平均绝对误差 MAE:线性插值<立方插值<最邻近插值<三次样条插值。

均方根误差 RMSE:线性插值<立方插值<最邻近插值<三次样条插值。

估计值与实测值 R^2 :线性插值>立方插值>最邻近插值>三次样条插值。

3.2 气象要素精度比较

3.2.1 气温要素空间插值估计结果

对3号台站气温要素进行空间插值,在所有样本台站2020年11月至2021年10月的观测数据中选取同一时间维度下的气温数据进行插值函数的构造。同样使用留出法对全部数据进行划分,划分比例为0.67。使用反距离加权法、克里金插值法、样条函数法及基于三角形的三次插值法进行插值估计。输入为除3号台站外其他6个便携式自动站在同一时刻的气温观测值,输出为3号台站对应时刻的估计值,构建插值函数。插值所得的估计值与实测值的散点图如图3所示,4种插值方法所得到的误差如图4所示。

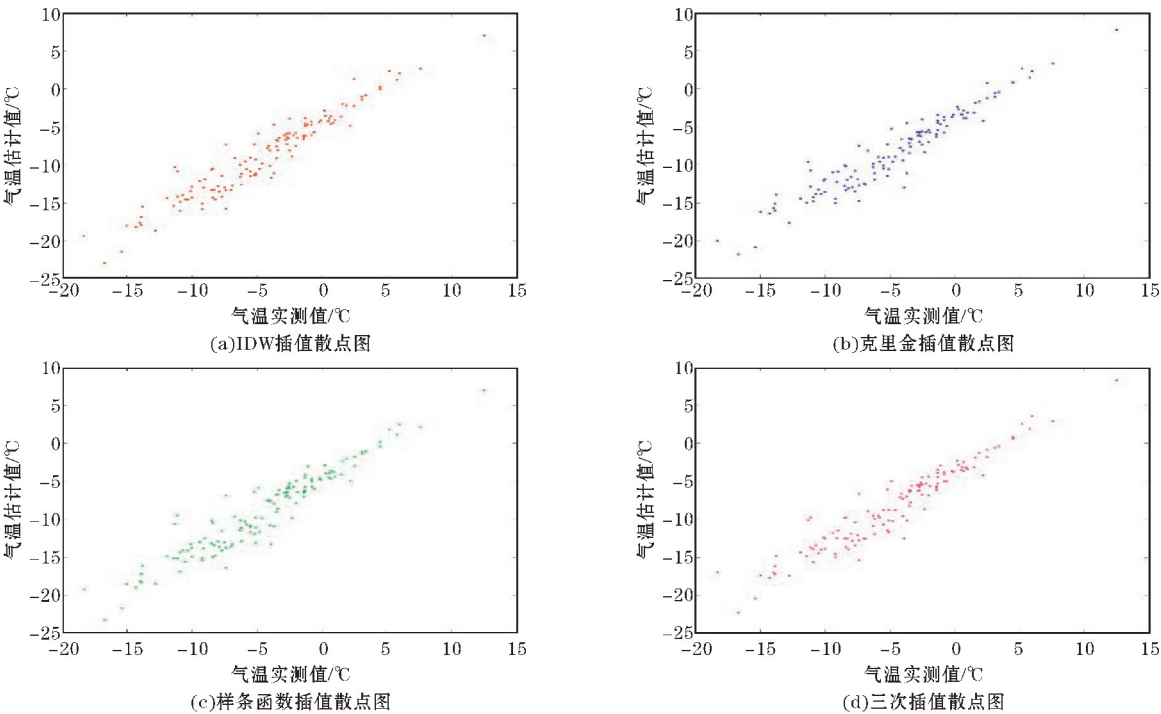


图3 气温空间插值散点图

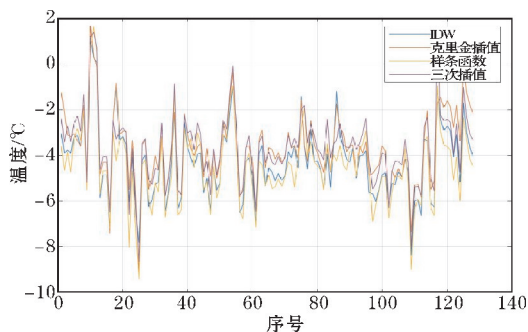


图4 气温空间插值误差图

由图3、图4可知,4种插值方法的估计值与实测值具有明显线性相关关系,误差折线变化趋势一致,其中样条函数法最为稳定,误差最小。IDW插值法误差在 $-8.36\text{ }^{\circ}\text{C}\sim 1.12\text{ }^{\circ}\text{C}$,克里金插值法误差在 $-9.11\text{ }^{\circ}\text{C}\sim 1.67\text{ }^{\circ}\text{C}$,三次插值法误差在 $-8.65\text{ }^{\circ}\text{C}\sim 1.38\text{ }^{\circ}\text{C}$,样条函数法误差在 $-9.39\text{ }^{\circ}\text{C}\sim 1.64\text{ }^{\circ}\text{C}$ 。由表3可知,4种插值方法的估计值与实测值间的相关系数均大于0.9,具有很强的相关关系,克里金插值的均方根误差和平均绝对误差均小于其他3种插值方法,样条函数和IDW插值误差最大。综合表3可知,

平均绝对误差MAE:克里金插值<基于三角形的三次插值<IDW插值<样条函数插值。

均方根误差RMSE:克里金插值<基于三角形的三次插值<样条函数插值<IDW插值。

估计值与实测值 R^2 :克里金插值=基于三角形的三次插值=IDW插值>样条函数插值。

表3 气温要素空间插值估计结果检验

参数	IDW	克里金插值	样条函数插值	三次插值
MAE	4.28	3.70	4.52	3.71
RMSE	4.55	3.59	4.48	3.64
R^2	0.96	0.96	0.95	0.96

3.2.2 气压要素空间插值估计结果

对5号台站气压要素进行空间插值,使用的数据集与气温要素空间插值相同,选取同一时间维度下的气压数据进行插值函数的构造。利用最邻近插值、线性插值、三次样条插值及立方插值法进行插值估计。海拔高度是影响气压的主要因素,因此使用海拔高度构造气压插值函数,输入为海拔高度,输出为对应的大气压力值。4种插值方法得到的误差如图5所示,精度对比结果如表4所示。

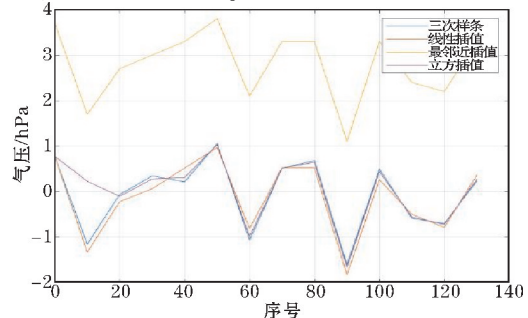


图5 气压插值误差图

表4 气压插值估计结果检验

参数	最邻近插值	线性插值	三次样条插值	立方插值
MAE	2.85	0.55	0.57	0.51
RMSE	2.92	0.70	0.69	0.63
R^2	0.9805	0.9816	0.9810	0.9839

由图5、表4可知,上述4种插值方法均具有较好的插值效果。其中,三次样条插值、线性插值和立方插值效果比最邻近插值效果好。最邻近插值法误差在1.1~3.8 hPa,线性插值误差在 $-1.83\sim 0.97\text{ hPa}$,三次样条插值误差在 $-1.61\sim 1.05\text{ hPa}$,立方插值误差在 $-1.66\sim 1.03\text{ hPa}$ 。4种插值方法的估计值与实测值间的相关系数 R^2 均大于0.98,具有非常强的相关关系。立方插值法效果最好,其均方根误差和平均绝对误差在4种插值方法中最小,最邻近插值法误差最大。综合表4可知,

平均绝对误差MAE:立方插值<线性插值<三次样条插值<最邻近插值。

均方根误差RMSE:立方插值<三次样条插值<线性插值<最邻近插值。

估计值与实测值 R^2 :立方插值>线性插值>三次样条插值>最邻近插值。

4 总结与讨论

针对便携式自动气象站数据缺测、缺失问题,利用几种经典插值方法对气温要素在时间域和空间域、对气压要素在空间域进行插值研究。气温时间插值部分,选择最邻近插值、线性插值、三次样条插值及立方插值法,通过实验结果表明,线性插值效果最优,并且在温度较低的第一季度、第四季度表现更好。4种插值方法从优到劣排序为:线性插值>立方插值>最邻近插值>三次样条插值。气温的空间插值部分,选择反距离加权法、克里金插值法、样条函数法及基于三角形的三次插值法,通过实验结果表明,克里金插值法效果最优,插值估计误差最小,4种插值方法从优到劣排序为:克里金插值>基于三角形的三次插值>IDW插值>样条函数插值。气压的插值部分,选用最邻近插值、线性插值、三次样条插值及立方插值法,结果表明,4种方法估计误差均具有较好的插值效果,其中立方插值效果最优,4种插值方法从优到劣排序为:立方插值>线性插值>三次样条插值>最邻近插值。

总体而言,对数据进行插值处理可以有效提高数据的连续性,有助于提高后续研究的统计精度和长期趋势估计。不同气象要素需选用不同插值方法,同时,同一种插值方法不能保证适用于任何情况,在不同环境条件下,可能具有不同的插值精度。并且便携式自

动气象站自身性能具有差异性,在后续研究中可结合机器学习的一些算法,建立回归模型,提高插值精度。

参考文献:

- [1] Eischeid J K, Pasteris P A. Creating a Serially Complete, National Daily Time Series of Temperature and Precipitation for the Western United States [J]. *Journal of Applied Meteorology*. 2000, 39 (9):1580-1591.
- [2] 王丽娜. 气象要素空间插值算法的研究及其应用[D]. 重庆:重庆理工大学,2014.
- [3] 刘琰琰. 气象要素插值的空间化精度提高方法研究[J]. *气象科学*,2017,37(2):278-282.
- [4] Burrows M T, Schoeman D S, Buckley L B, et al. The Pace of Shifting Climate in Marine and Terrestrial Ecosystems[J]. *Science*,2011,334:652-655.
- [5] 姜晓剑,刘小军,黄芬,等. 逐日气象要素空间插值方法的比较[J]. *应用生态学报*,2010(3):7.
- [6] 张超,孙绩华,段玮. 云南区域站降水资料利用 Surfer 软件实现 Cressman 插值的研究[J]. *成都信息工程大学学报*,2018,33(1):84-90.
- [7] Tang G, Clark M P, Papalexiou S M. The use of serially complete station data to improve the temporal continuity of gridded precipitation and temperature estimates[J]. *Journal of Hydrometeorology*,2021(5).
- [8] 沈艳,潘旻,徐宾,等. 最优插值法对中国自动站降水量空间分析中的参数优化[J]. *成都信息工程大学学报*,2012,27(2):219-224.
- [9] 钟雨萍. 基于机器学习的城市风场时序预测与空间插值[D]. 深圳:中国科学院大学(中国科学院深圳先进技术研究院),2021.
- [10] 王艳萍,李新庆,刘焱,等. 不同插值算法对气温空间插值效果评估分析[J]. *信息技术*,2020,44(6):31-35.
- [11] 李叶,张艳红,陈子琦,等. 中高纬度山区气温空间化的方法比较研究——以大兴安岭北麓为例[J]. *山地学报*,2021,39(2):174-182.
- [12] Qiao P, Li P, Cheng Y, et al. Comparison of common spatial interpolation methods for analyzing pollutant spatial distributions at contaminated sites [J]. *Environmental Geochemistry and Health*,2019,41(7).
- [13] Serrano-Notivol R, Santiago Beguería, Martín de Luis. STEAD: a high-resolution daily gridded temperature dataset for Spain[J]. *Earth System Science Data*,2019,11(3):1171-1188.

Research on Interpolation Method based on Portable Automatic Weather Station Data

FENG Yanying, LU Huiguo, JIANG Juanping

(College of Electronic Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Portable automatic weather stations are often deployed in remote uninhabited areas. Due to various factors, the observation data are often missing. Using data interpolation method to improve data continuity can provide an important basis for subsequent research. In this paper, the temperature and pressure data of seven portable automatic stations deployed in the unmanned area of the western Qinghai-Tibet Plateau from November 2020 to October 2021 are selected for interpolation research. The nearest interpolation, linear interpolation, cubic spline interpolation and cubic interpolation method are selected for the interpolation of temperature in time domain and air pressure in space domain. For the time interpolation of temperature, the time is selected as the independent variable, and the corresponding temperature value is the dependent variable to construct the interpolation function. The results show that the linear interpolation method has the best interpolation effect, and the interpolation accuracy in the first and fourth quarters of the lower temperature is higher than that in the third and second quarters; for the spatial interpolation of air pressure, the altitude is selected as the independent variable, and the corresponding air pressure value is used as the dependent variable to construct the interpolation function. The results show that the cubic interpolation effect is the best. The inverse distance weighting method, Kriging interpolation method, spline function method and triangle-based cubic interpolation method are selected for the interpolation of temperature in the spatial domain. The interpolation function is constructed by interpolating the longitude and latitude of stations around the target site and the temperature value. The results show that the Kriging interpolation method is significantly better than the other three interpolation methods.

Keywords: portable automatic weather stations; meteorological data processing; air temperature interpolation; atmospheric pressure interpolation