

孙越,王海军,周月华,等.2023.三种插值方法对区域自动气象站日气温缺测数据插补的适用性研究[J].暴雨灾害,42(1): 97-104. SUN Yue, WANG Haijun, ZHOU Yuehua, et al. 2023. Applicability of three interpolation methods in estimating daily temperature with missing data from regional automatic weather station [J]. *Torrential Rain and Disasters*, 42(1): 97-104 (in Chinese). doi:10.12406/byzh.2022-183

## 三种插值方法对区域自动气象站日气温缺测数据插补的适用性研究

孙越<sup>1</sup>,王海军<sup>1</sup>,周月华<sup>2</sup>,严婧<sup>1</sup>,刘莹<sup>1</sup>

(1. 湖北省气象信息与技术保障中心,武汉 430074;2. 武汉区域气候中心,武汉 430074)

**摘要:**为解决气温观测记录缺测的问题,选择反距离权重插值(Inverse Distance Weighted, IDW)、普通克里金插值(Ordinary Kriging, OK)和多元线性回归(Multiple Linear Regression, MLR)三种方法,以湖北省2020年为例,对全省逐日平均气温( $T$ )、最高气温( $T_{\max}$ )和最低气温( $T_{\min}$ )进行空间插补,并采用平均绝对误差(Mean Absolute Error, MAE)对3种方法的插补结果进行检验。结果表明:用MLR插补得到的 $T_{\max}$ 、 $T_{\min}$ 、 $T$ 的MAE最小,分别为0.41℃、0.31℃和0.20℃,其中 $T$ 的插补误差在1℃以内的站点比例高达100%;相比IDW和OK,MLR插补结果的MAE空间分布均匀,其不仅随海拔高度变化较小,随季节变化也相对较小。单站试验结果表明,当用于MLR模型的样本数量越多、时间离散度越大时,MLR对气温的插补效果越好。总体上,对日气温缺测数据的插补效果,MLR最优,IDW次之,OK最差;对于建立气象站点长时间连续气温数据集而言,MLR更适合解决区域自动气象站日气温数据缺测问题。

**关键词:** 气温;记录缺测;数据插补方法;平均绝对误差;多元线性回归

中图分类号: P423

文献标志码: A

DOI: 10.12406/byzh.2022-183

### Applicability of three interpolation methods in estimating daily temperature with missing data from regional automatic weather station

SUN Yue<sup>1</sup>, WANG Haijun<sup>1</sup>, ZHOU Yuehua<sup>2</sup>, YAN Jing<sup>1</sup>, LIU Ying<sup>1</sup>

(1. Hubei Meteorological Information and Technological Support Center, Wuhan 430074; 2. Wuhan Regional Climate Center, Wuhan 430074)

**Abstract:** In order to solve the problem of missing record of temperature observation, taking the Hubei Province in 2020 as an example, we selected three methods, that is, Inverse Distance Weighted (IDW), Ordinary Kriging (OK) and Multiple Linear Regression (MLR), to interpolate the missing values of daily mean temperature ( $T$ ), maximum temperature ( $T_{\max}$ ), and minimum temperature ( $T_{\min}$ ). Based on this interpolation results, using the average absolute error (MAE), we evaluated the interpolation results obtained by these three methods. The results show that the MAE of  $T_{\max}$ ,  $T_{\min}$  and  $T$  obtained with the MLR interpolation is the lowest, which are 0.41 °C, 0.31 °C, and 0.20 °C, respectively. Meanwhile, the interpolation errors of  $T$  at all stations are less than 1 °C. Compared with IDW and OK, the MAE spatial distribution of interpolation results obtained with MLR is more uniform with slight changes with altitude and seasons. Single station test shows that the more samples used for MLR model is and the greater the sample time dispersion is, the better the interpolation effect of MLR on temperature is. On the whole, the interpolation effect of MLR on missing values of daily temperature from regional stations is the best, IDW is the second, and OK is the worst. For establishing long-term continuous temperature datasets of meteorological stations, MLR is more suitable for solving the problem of missing records of daily temperature from regional automatic weather station (AWS).

**Key words:** temperature; missing observation record; interpolation method; mean absolute error; multiple linear regression

### 引言

高质量气象站点数据是编制长时间连续且均一性较好的历史气象数据集,也是开展气候分析、灾害风险普查等工作的基础(黄俊杰等,2018;娄珊珊等,

2019;陶玮等,2021)。由于受自然条件或人为因素影响,各地地面观测数据停测、缺测、漏测现象时有发生,使得在建立连续气象数据集的过程中,必然要面临如何对以往缺测气象观测记录进行插补的问题(郭彦等,2015;李楠等,2018;王少影等,2020)。国内外已

收稿日期: 2022-09-19;定稿日期:2022-11-21

资助项目: 湖北省气象局科技发展基金(重点)项目(2022Z05-01)

第一作者: 孙越,主要从事气象数据质量控制与评估工作。E-mail: 376262384@qq.com

有人研究指出气象观测记录的缺测对气候分析或其他相关工作的影响。如Stooksbury等(1999)指出,日最高、最低气温记录中断(即记录缺测)对计算分析月平均最高、最低气温的时空变化规律存在不利影响;李庆祥等(2005)认为,气象数据缺测或中断对气候变化及其趋势研究、气候评估及评价会产生影响。因此,如何寻找到对缺测记录进行插补的有效方法,引起了国内外许多学者的关注,如Thornton等(1997)、Acock和Pachepsky(2000)、王海军等(2008)曾对短时间缺测记录插补方法做过研究。对缺测记录进行插补,多采用空间插值方法,包括确定性插值方法和统计学插值方法。前者基于已知点之间的相似程度进行插值,如,反距离加权插值法、趋势面法、样条函数法等(刘宇等,2006;王海军等,2007;李框宇等,2019;潘霄等,2019);后者基于样本点的统计规律构建空间结构模型进行插值,如克里金插值法(Li and Heap, 2014;马诺等,2020)和回归插值方法等。除空间插值外,还有回归插值方法,该方法包括一元回归和多元回归插值法等,其原理是通过空间已知的求解得到模型或函数来最大限度地逼近插值点。Huth和Nemešová(1995)针对不同天气类型建立相应的回归模型来插补缺测的日气温数据。姬兴杰等(2020)采用线性回归方法建立了中岳嵩山高山国家基准气候站1956—2017年时间序列连续的月平均气温资料。已有研究表明,对于气温缺测记录的插补,不同插值方法的适用性与空间尺度和时间尺度有关。姜晓剑等(2010)分析了全国1951—2005年559个气象站逐月第15日平均最高、最低气温的插值结果,指出采用薄盘样条插值法得到的最高、最低气温插值误差最小,分别为1.02℃和1.12℃;张莉莉等(2012)对1979—2008年海南岛18个气象站年平均气温进行空间插值,结果表明混合插值法精度最高,样条插值法精度最低;王新宇和黄鹏程(2020)对比了克里金插值、反距离权重插值和张力样条函数插值对全国660个气象站年平均气温的插值精度,其中克里金法最优。上述研究针对的主要是国家级气象站,而目前对于区域自动气象站日气温插补的研究相对较少,特别是适用于各种地形气候条件下常用地面气象要素的最优插补方法更少。

湖北省现有区域自动气象站(以下简称区域站)2 000余个,由于受仪器故障、观测环境改变等因素的影响,其部分观测数据不时出现缺测。为最大限度保证区域站资料的完整性和可用性,必须认真对待并妥善解决区域站数据缺测的问题。从上述相关研究中看到,各种插值方法虽能在一定程度上解决区域站数据缺测问题,但不同插值法对于不同要素其适用性不同,

且对不同地形条件也有差异。因此,为了便于气象科研业务人员在实际工作中有效使用各种插值方法,需对这些方法的适用性进行科学分析和研判。为此,本文选择反距离权重法、普通克里金法、多元线性回归法这3种典型插值方法,以2020年湖北省区域站为例,对当年全省各区域站日平均气温、日最高气温和日最低气温缺测数据进行空间插值,从插补结果误差的空间分布、季节变化及样本数量对插补结果的影响3个方面,探讨3种插值方法对气温缺测插补的适用性,期望能为编制高质量气温观测数据集和进一步开展气候分析、气象灾害综合风险评估等工作提供数据支撑。

## 1 资料与方法

### 1.1 资料说明

2020年1月1日—12月31日湖北省有气温观测数据的区域站共计1 690个,其数据平均完整率(全年有记录的天数与全年应观测天数之比)为95.24%。本文选择完整率大于90%的区域站(共1 464个)逐日气温数据作为研究对象,包括日平均气温( $T$ )、日最高气温( $T_{\max}$ )和日最低气温( $T_{\min}$ )资料,其来源于湖北省气象信息与技术保障中心。上述资料使用前,已对其进行了气象要素界限值、气候极值、内部一致性和时间一致性等质量控制。

### 1.2 插值方法介绍

(1)反距离权重插值法(Inverse Distance Weighted, IDW)。IDW是以插值点与样本点间的距离为权重进行加权平均(Thornton et al., 1997),离插值点越近的样本点赋予的权重越大,其目标站点( $p$ )的估计值( $x_p$ )按下式计算。

$$x_p = \frac{\sum_{i=1}^N \frac{1}{(d+5)^2} [x_i + \beta(h_p - h_i)]}{\sum_{i=1}^N \frac{1}{(d+5)^2}} \quad (1)$$

式(1)中, $x_i$ 为第*i*个邻近参考站点的要素值; $N$ 为邻近站个数; $d$ 为目标站点和邻近参考站点的距离(单位:km),本文参考Nalder和Wein(1998)的研究成果,权重距离幂次数取2,同时为避免某一邻近参考站距离过近导致权重过大,对*d*添加了圆滑参数5,以此保证无任一邻近参考站被赋予全部权值; $h_p$ 、 $h_i$ 分别是目标站点和邻近站点的海拔高度; $\beta$ 为气温随海拔高度的垂直变化率,本文取常数 $-0.006\text{ }^\circ\text{C}\cdot\text{m}^{-1}$ 。

(2)普通克里金插值法(Ordinary Kriging, OK)。OK同样是使用已知点的属性加权值来得到目标点的预测值(Matheron, 1963),其目标站点的估计值( $\hat{y}$ )按下式计算。

$$\hat{y} = \sum_{i=1}^N \lambda_i x_i \quad (2)$$

式(2)中, $x_i$ 为第*i*个邻近参考站点的要素值; $\lambda_i$ 为第*i*个

邻近参考站的权重。不同于IDW,OK是使用广义最小二乘对上式进行无偏和最优估计,并建立如下方程组来确定权重系数(岳文泽等,2005)。

$$\begin{cases} \sum_{i=1}^N \lambda_i = 1 \\ \sum_{i=1}^N \lambda_i C(x_i, x_j) - C(x_i, x_p) - \mu = 0 (j=1, 2, \dots, N) \end{cases} \quad (3)$$

式(3)中, $C(x_i, x_j)$ 为第*i*个邻近站点和第*j*个邻近站点之间的协方差函数; $C(x_i, x_p)$ 为第*i*个邻近站点和目标点之间的协方差函数; $\mu$ 为拉格朗日乘数。OK假设半方差函数只与两个位置的空间距离有关,可进一步通过半方差函数的拟合函数求解得到权重值。

(3) 多元线性回归插值法(Multiple Linear Regression, MLR)。MLR是利用回归方程对多个自变量和因变量之间的关系进行建模。即

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i \quad (i=1, 2, \dots, N) \quad (4)$$

式(4)中, $\hat{y}$ 为目标站要素估计值; $x_i$ 为第*i*个邻近参考站点的要素值; $\beta_0, \beta_1, \dots, \beta_N$ 为回归模型参数,回归参数通过普通最小二乘法求解方程组得到(王海军等,2008)。

以上3种插值方法中的邻近站个数*N*,对于同一目标站,不同插补方法选择的邻近站个数相同。本文插补试验中,邻近站为目标站半径30 km内的其它气象观测站,当邻近站的个数*N*<5时,将半径动态扩大直至邻近站个数*N*≥5个。

### 1.3 误差分析

首先,通过IDW、OK和MLR插补得到“缺测”的日气温数据;然后,用平均绝对误差(Mean Absolute Error, *E*)对插补结果进行评价。*E*的计算式如下

$$E = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (5)$$

其中, $x_i$ 为第*i*天观测值; $y_i$ 为第*i*天插补值;*n*为插补天数。分析过程中,统计*E*的最大值( $E_M$ )和标准差(*S*),以此分别表征误差的极值和误差稳定性。*S*越小,表明误差离散程度小,误差越稳定。同时,还统计了平均误差( $E_A$ ),以此表征误差的无偏性,其计算公式如下

$$E_A = \frac{1}{n} \sum_{i=1}^n (x_i - y_i) \quad (6)$$

当 $E_A$ 接近0时,表明误差是无偏的。此外,在误差空间分布中进一步分析*E*在不同区间的概率分布以及不同海拔区间平均*E*分布特点,以便更好地对比不同插补方法的适用性。

在MLR插补试验中,对每个站点全年记录数据每隔一天(1 d)设为缺测,使用约50%的数据于多元线性回归建模,再对剩余日期进行插补。对此可简单表述为以第1, 3, 5, ..., 2*n*-1天的数据用于建模,插补得到

第2, 5, 6, ..., 2*n*天的数据;然后,以第2, 4, 6, ..., 2*n*天的数据用于建模,插补得到第1, 3, 5, ..., 2*n*-1天的数据。这样既能保证有足够多的样本数,又能检验MLR方法在全年不同月份的表现。在IDW和OK插补试验中,假设全年真实观测数据缺测,然后使用邻近站数据对其进行插补,最后通过对比插补值和真值的差异来衡量3种插值方法对日气温缺测数据插补的适用性。

以上插补试验基于完整率大于90%的区域站,而实际上2020年湖北省有26个区域站日气温数据完整率小于50%,如建始小溪口气象观测站(Q5806),全年只有4—6月有85个日气温记录。为进一步探究样本数量是否对气温插补结果存在影响,以小溪口站为例,进行如下试验设计(表1)。

表1 基于建始县小溪口单站样本数量的气温插补结果影响试验设计

Table 1 Experiments on effects of sample size on interpolated temperature based on Xiaoxikou station in Jianshi County of Hubei Province

试验方案	多元线性回归插值法(MLR)训练建模样本来源
Exp1	小溪口站4—6月有气温记录日(85 d)平均气温
Exp2	随机选择该站85 d日平均气温
Exp3	随机选择该站一年记录中10%的日平均气温
Exp4	随机选择该站一年记录中50%的日平均气温
Exp5	随机选择该站一年记录中90%的日平均气温

表1中,Exp1中用于MLR建模的样本是4—6月的连续数据;Exp2—Exp5中用于MLR建模的样本是随机选择的分布在全年的数据。然后,用不同方法对剩余“缺测”日期的气温进行插补。同时,对比IDW、OK和IDW+MLR方法的插补结果。IDW+MLR方法是先用IDW方法对“缺测”日期气温进行插补,得到全年数据用于MLR训练建模,再用回归模型对“缺测”日期气温进行插补。

## 2 结果与分析

本文分别采用IDW、OK与MLR对2020年湖北省1 464个区域站逐日 $T$ 、 $T_{max}$ 和 $T_{min}$ 进行插值。表2给出用上述3种插补方法得到的湖北全省区域站气温的 $E$ 、 $E_M$ 、 $E_A$ 和*S*的平均值。由 $E_A$ 值接近0可知,3种插补方法的误差均是无偏的,其中MLR方法对气温的插补结果总体偏低,OK方法的插补结果总体偏高。再由 $E_M$ 统计结果可知,3种插补方法的误差极大值总体在2℃左右。从误差无偏性和极大值看,用这3种方法对缺测气温进行插补是可行的。3个气温要素中, $T$ 的 $E$ 、 $E_M$ 、 $E_A$ 和*S*均小于 $T_{max}$ 和 $T_{min}$ 的,其原因是 $T_{max}$ 和 $T_{min}$ 受局地因素影响较大,使得目标站与邻近站其差别也较大,因而插补结果误差偏大(李庆祥等,2005;王海军

等,2008)。对比3种插补方法,无论是从其 $E$ 和 $E_M$ 还是 $S$ (误差稳定性)和 $E_A$ (误差无偏性)看,MLR表现均最优,IDW次之,OK最差。为进一步对比不同插补方法的表现,本文将以 $T$ 的插补结果为例分析其MAE在时间和空间上的特征。

表2 基于IDW、OK和MLR法的湖北省2020年日平均气温( $T$ )、日最高气温( $T_{max}$ )和日最低气温( $T_{min}$ )插补结果的误差统计(单位:  $^{\circ}\text{C}$ )

Table 2 The error statistics (unit:  $^{\circ}\text{C}$ ) of the daily mean temperature ( $T$ ), daily maximum temperature ( $T_{max}$ ) and daily minimum temperature ( $T_{min}$ ) interpolated respectively by the IDW, OK and MLR three methods over Hubei province in 2020

插补方法	温度	$E$	$E_M$	$S$	$E_A$
MLR	$T$	0.20	0.84	0.27	$-0.28 \times 10^{-3}$
	$T_{max}$	0.41	1.91	0.54	$-0.50 \times 10^{-3}$
	$T_{min}$	0.31	1.46	0.43	$-0.66 \times 10^{-3}$
IDW	$T$	0.47	1.36	0.45	$20.00 \times 10^{-3}$
	$T_{max}$	0.70	2.20	0.72	$-10.00 \times 10^{-3}$
	$T_{min}$	0.68	1.95	0.70	$40.00 \times 10^{-3}$
OK	$T$	0.73	1.44	0.47	$50.00 \times 10^{-3}$
	$T_{max}$	1.01	2.29	0.74	$30.00 \times 10^{-3}$
	$T_{min}$	0.85	2.03	0.72	$60.00 \times 10^{-3}$

注:  $E$ 为平均绝对误差;  $E_M$ 为 $E$ 最大值;  $S$ 为 $E$ 标准差;  $E_A$ 为平均误差

### 2.1 气温插补误差的空间分布

图1给出基于IDW、OK和MLR三种方法插补得

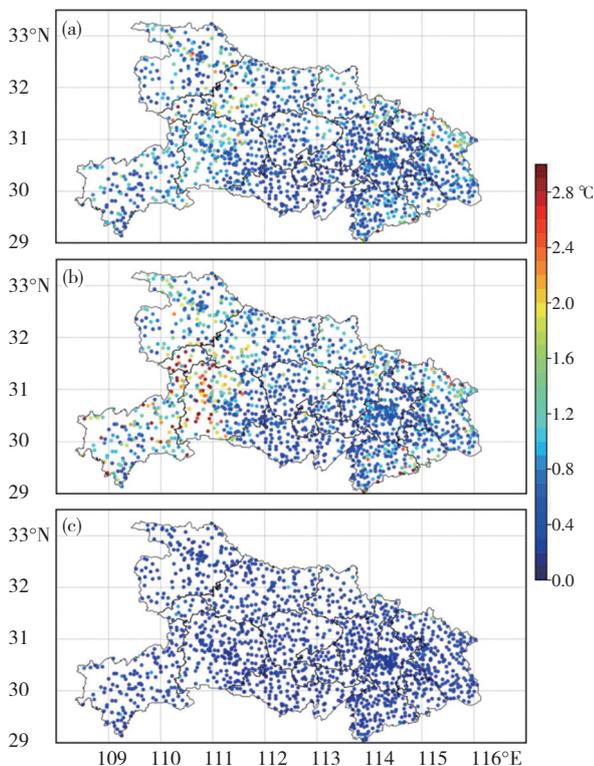


图1 基于IDW(a)、OK(b)和MLR(c)三种方法插补的2020年湖北省气象站点日平均气温的平均绝对误差分布

Fig.1 Distributions of the mean absolute error of the daily mean temperature interpolated respectively by the IDW, OK and MLR three methods at the weather stations over Hubei in 2020

到的2020年湖北省日平均气温的平均绝对误差( $E$ )的空间分布。从中看到,采用IDW方法插补的各站 $T$ 的 $E$ 空间分布较为一致(图1a),多数区 $E < 1.0^{\circ}\text{C}$ ,仅个别站点出现 $E > 2.0^{\circ}\text{C}$ ;采用OK方法插补的 $T$ 的 $E$ 空间分布各站存在明显差异(图1b),平原地区,如荆门、荆州、天门、仙桃、武汉等市,其误差较小,而湖北西部,如宜昌市和神农架林区,以及湖北与邻省边界处,如咸宁以南和黄冈以北区域,其误差较大。三种方法相比,MLR方法插补的 $T$ 的 $E$ 各站整体相对较小(图1c),且其空间分布均匀,表明平均绝对误差随空间位置无明显变化。

为进一步揭示采用三种插补方法获得的日平均插值气温误差的空间分布特征,图2统计得到的3种插补结果的不同误差区间平均绝对误差概率分布及其对应的频率(频数/总数)。从中看到,IDW、OK和MLR插补气温平均绝对误差在 $1.0^{\circ}\text{C}$ 以内的站点比例分别达到84.01%、75.75%和100%;IDW和OK的平均绝对误差概率分布更离散(图2a、b),主要分布在 $0.2 \sim 0.5^{\circ}\text{C}$ 区间内,同时两者均存在明显的极端误差( $E > 2.0^{\circ}\text{C}$ )的站点,且OK插补结果的极端误差更突出(占比达7.31%);MLR插补结果的平均绝对误差分布十分集中(图2c),其主要分布在 $0.1 \sim 0.2^{\circ}\text{C}$ 区间,无极端误差。从平均绝对

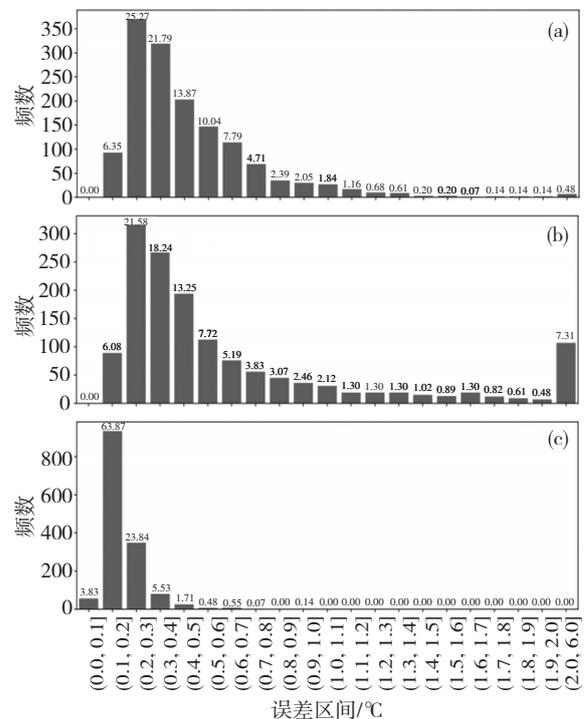


图2 基于IDW(a)、OK(b)和MLR(c)三种方法插补得到的2020年湖北省日平均气温平均绝对误差频率分布(柱顶数字表示在相应误差区间的概率,单位: %)

Fig.2 The frequency distributions of the mean absolute error of the daily mean temperature interpolated respectively by the IDW, OK and MLR three methods at the weather stations over Hubei in 2020. The numbers on the top of the histograms are the corresponding probabilities (unit: %) at the different error ranges

误差区间频数分布看,对区域气温使用MLR插补得到的结果误差小于IDW和OK方法插补得到的结果误差。

从平均绝对误差空间分布看出其大值主要出现在鄂西和鄂东的山区,表明插补结果误差对海拔高度变化较为敏感。为分析海拔高度对各站插补方法得到结果的影响,图3给出基于MLR、IDW和OK三种插补方法的不同海拔高度区间各站点 $T$ 的 $E$ 分布。从中可见,总体上,3种方法插补结果的 $E$ 均在低海拔地区较小、在高海拔地区较大,尤其是海拔高度1500 m以上的区域 $E$ 显著增加。对比3种插补方法,在同一海拔高度区间,MLR方法的插值结果均明显优于其它两种方法。另外,随着海拔高度增加,MLR插补结果的 $E$ 增幅也明显小于其它两种方法,表明MLR方法受地形

因素影响较小,比IDW和OK方法表现更稳定。其中,OK方法的插补结果误差对海拔高度变化最敏感,在200 m以下区域 $E < 1\text{ }^{\circ}\text{C}$ ,在1500 m以上区域 $E$ 迅速增加到 $2\text{ }^{\circ}\text{C}$ 以上;OK方法本身更适用于空间连续变化且平展的下垫面,而高海拔地区往往地形复杂、坡度变化大,用其进行温度插补的效果较差。IDW的插补结果在不同海拔区间总体较为稳定,海拔高度2000 m以下地区 $E < 1\text{ }^{\circ}\text{C}$ ;IDW方法用于插值时虽然引入温度垂直变化率( $\beta$ )考虑了海拔高度对气温的影响,但 $\beta$ 只是一经验值,所以此法用在山区也存在一定的局限性。同样,分析MLR、IDW和OK方法对不同海拔高度区间日最高气温和日最低气温的插补结果,也得到了与平均气温类似的结论。

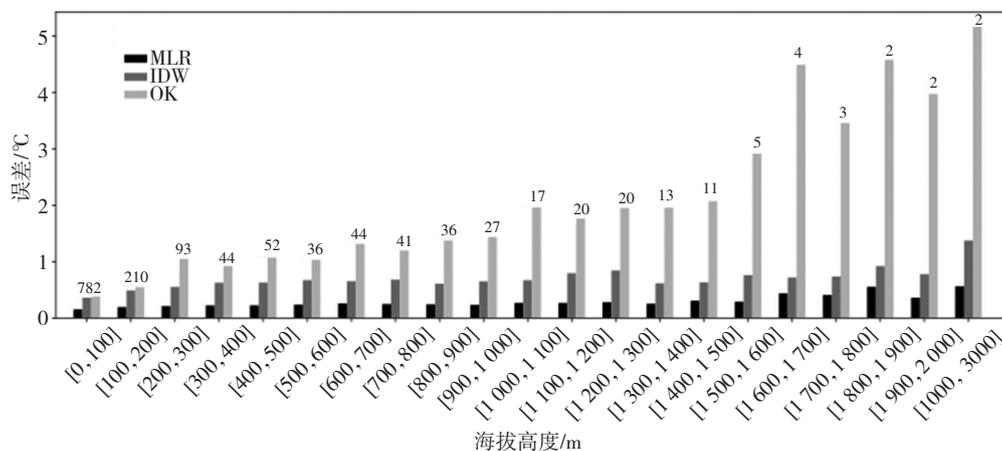


图3 基于IDW、OK和MLR三种方法插补得到的2020年湖北省日平均气温平均绝对误差(单位:  $^{\circ}\text{C}$ )随海拔高度(单位: m)的变化(柱顶数字表示在相应海拔区间的站点数)

Fig.3 Variations of the mean absolute error (unit:  $^{\circ}\text{C}$ ) of the daily mean temperature interpolated respectively by the IDW, OK and MLR three methods with the altitude (unit: m) at the weather stations over Hubei in 2020. The numbers on the top of the histograms are the corresponding station numbers at the different altitude ranges

综上气温插补误差空间分布特征的分析表明:MLR插补结果误差最小,误差空间分布均匀,受海拔高度和地形因素影响最小,适用于各种地区;IDW通过引入经验系数考虑了气温随海拔高度的变化,其插补误差随海拔变化较小;OK方法插补结果误差的空间变化明显,不适用于高海拔和地形变化较大的区域。

## 2.2 气温插补误差的季节变化

为揭示各种插值方法在不同季节获得的气温插值结果的误差大小,图4给出基于IDW、OK和MLR三种方法插补得到的2020年各月湖北省 $T$ 、 $T_{\max}$ 和 $T_{\min}$ 的平均绝对误差( $E$ )变化。对比三种气温, $T$ 的插补误差总体小于 $T_{\min}$ 和 $T_{\max}$ 。图4a显示 $T$ 的平均插补误差随月份变幅较小,MLR和OK插值结果的 $E$ 在8月最大,而IDW的是在11月最大,三种方法的 $E$ 在春季(3—5月)最高;图4b显示 $T_{\max}$ 的平均插补误差存在明显的季节变化, $E$ 在夏季(6—8月)最高,其次是春季,冬季

(12—2月)最低;图4c显示 $T_{\min}$ 平均插补误差的季节变化显著, $E$ 在春季(3—5月)最高,其次是冬季,夏季最低。图4b和图4c的相反变化与夏季易出现极端高温、冬春季节易出现极端低温有关,反映了插补误差对极值的敏感性。冬季冷空气活动频繁,冷气团移动速度和路径不同会导致一定范围内气温时空变化不一致,使得插补站和邻近站最低气温差异较大;同样地,在夏季,局地强对流天气往往导致区域站之间最高气温差异较大,会使插补站和邻近站最高气温存在较大差异。对于IDW和OK,其插补结果取决于当前时刻邻近站气温的加权结果,当插补站与邻近站受到冷空气活动或局地强对流天气的影响而使其气温相差较大时必然导致温度插补误差较大。而对于MLR,线性回归模型是基于全年数据,能较好地反映插补站与邻近站气温的变化关系,即使插补站和邻近站气温差异较大时。所以,对于极值气温的插补,MLR效果最优。

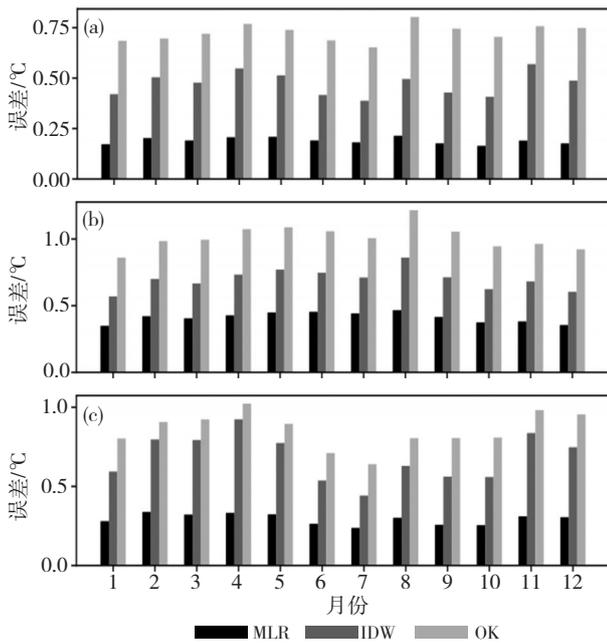


图4 基于IDW、OK和MLR三种方法插补得到的2020年湖北省日平均(a)、日最高(b)和日最低(c)气温平均绝对误差(单位:  $^{\circ}\text{C}$ )月变化

Fig.4 Monthly variations of the mean absolute error (unit:  $^{\circ}\text{C}$ ) of (a) daily mean temperature, (b) daily maximum temperature and (c) daily minimum temperature interpolated respectively by the IDW, OK and MLR three methods at the weather stations over Hubei in 2020

对比3种插补方法对各月 $T$ 、 $T_{\max}$ 和 $T_{\min}$ 的插补结果可知,MLR的插补误差均最小,IDW次之,OK插补误差最大。进一步计算3种插补方法对各月 $T$ 、 $T_{\max}$ 和 $T_{\min}$ 插补结果的标准差(S)表明,用MLR插补的 $T$ 、 $T_{\max}$ 和 $T_{\min}$ 的S均最小,IDW插补得到的 $T$ 和 $T_{\max}$ 的S最大,OK插补得到的 $T_{\max}$ 的S最大。

综合以上结果表明:MLR对气温缺测数据的插补效果受季节影响相对较小,其线性回归模型(即式(4))能更好地反映区域站之间气温的相关性,该方法适用于对不同季节气温缺测数据尤其是对 $T_{\max}$ 和 $T_{\min}$ 的插补;而IDW和OK,对气温缺测数据的插补误差随季节变化明显,特别是当插补站和邻近站受到局地高温、低温等极端天气过程影响而使其气温相差较大时,这两种方法的插补结果误差会随之增大,因此不适用对缺测 $T_{\max}$ 和 $T_{\min}$ 的插补。

### 2.3 样本数量对插补结果的影响

综上所述,MLR对区域站日气温数据的插补结果优于IDW和OK。具体表现是:(1)MLR插补误差小,平均绝对误差均值( $E_A$ )仅 $0.2^{\circ}\text{C}$ , $E \leq 1.0^{\circ}\text{C}$ 的站点占比高达100%, $E \leq 0.2^{\circ}\text{C}$ 的站点占比达到67.7%;(2)MLR的插补结果稳定,无明显极端误差出现,也不受海拔高度和季节变化的影响。究其原因,MLR主要是通过构建回归模型来反映目标站和邻近站的关系,即使两

站气温相差较大,如若两站气温时间变化呈较好的线性关系,通过其回归模型就能达到较好的插补效果。对气温而言,其时空变化是连续的,目标站和邻近站气温的线性相关性较好,MLR更适用于对缺测气温进行插补。反观其他两种方法,IDW的插补结果误差大于MLR,其原因可能是IDW插补误差主要取决于邻近参考站的距离权重,对局地性较强或地理环境相差较大的两个站点,即使空间距离近,气温也可能存在较大差异,从而使得该方法在这种情况下会产生较大的插补误差。OK插补结果误差大于IDW的可能原因则是,此法主要是通过空间各处随机变量集合构成随机函数来推断目标站点的值,但本文中它相比IDW未考虑海拔高度的影响,所以在海拔高度和坡度较大地区构建的函数无法准确描述气温的空间分布,从而导致误差较大。

为进一步探究用于多元线性回归建模的样本数量是否对3种气温插补结果存在影响,以小溪口站为例,进行不同插补方法的样本数量对插补结果影响对比试验。考虑到小溪口站( $109.83^{\circ}\text{E}$ , $30.56^{\circ}\text{N}$ ,海拔高度 $560.0\text{ m}$ )气温缺测记录过多而不利于交叉检验,选择距该站 $12\text{ km}$ 、地理环境相似的建始国家基本气象站( $109.72^{\circ}\text{E}$ , $30.61^{\circ}\text{N}$ ,海拔高度 $609.2\text{ m}$ )的日平均气温进行相关对比试验。根据本文第1.3节中的试验设计方案,得到基于IDW、OK、MLR和IDW+MLR的2020年建始站日平均插值气温的平均绝对误差见表3。

表3 基于IDW、OK、MLR和IDW+MLR方法插补得到的2020年建始站日平均气温平均绝对误差(单位:  $^{\circ}\text{C}$ )

Table 3 Mean absolute error (unit:  $^{\circ}\text{C}$ ) of the daily mean temperature interpolated respectively by the IDW, OK, MLR and IDW+MLR methods at the Jianshi station of Hubei in 2020

试验方案	IDW	OK	MLR	IDW+MLR
Exp1	0.37	0.70	0.28	0.34
Exp2	0.38	0.68	0.26	0.33
Exp3	0.37	0.68	0.26	0.35
Exp4	0.36	0.69	0.25	0.28
Exp5	0.35	0.58	0.24	0.26

从表3中可见,所有试验方案中,用MLR插值的日平均气温误差均小于IDW和OK的;对比不同试验看到,OK的插值误差变化幅度最大(达 $0.12^{\circ}\text{C}$ ),IDW的插值误差变化幅度最小(仅 $0.03^{\circ}\text{C}$ );对于OK,插补样本数少时(Exp5)和插补样本数多时(Exp3)的误差相差约 $0.1^{\circ}\text{C}$ ,而MLR和IDW的误差相差仅 $0.02^{\circ}\text{C}$ ,表明OK的插补结果误差的不确定性大于MLR和IDW,且随着插补样本数量的变化存在一定的波动。对于MLR,其插补误差一方面受建模样本数量的影响,另一方面会受建模样本时间跨度的影响。

进一步对比表3中Exp3、Exp4、Exp5的 $E$ 看到,其训练数据越多,所建模型越准确,气温缺测插补误差越小。再如,Exp1和Exp2两个试验方案,其训练样本数量相同,但Exp1中是用4—6月的数据进行建模,然后对1—3月和7—12月气温进行插补,Exp2中训练样本则是随机分布在1—12月,再对1—12月缺测气温进行插补。可见,训练样本的时间分布越分散,即时间离散度越大,OK和MLR模型越能更好地反映目标站和邻近参考站气温的相关性,其插补误差越小。

为改善当站点可用率过低时MLR训练样本量少的问题,本试验引入IDW和MLR相结合的IDW+MLR方法。此法既增加了训练样本数量,又能保证样本在全年各月的分布。试验结果表明,IDW+MLR的插补结果的 $E$ 仍然大于MLR。结合IDW的插补结果看,是其给IDW+MLR模型中带来了误差,导致IDW+MLR的插补误差介于IDW和MLR之间。可见,人为增加建模样本数量和时间跨度并不能降低MLR的插补误差。值得注意的是,自2013年湖北省区域站数据被广泛使用以来,2013—2020年全省区域站日平均气温的年完整率 $>90\%$ 的站点数占当年总站点数的比例平均为 $77.6\%$ ,完整率 $<10\%$ 的站点数占比平均为 $0.8\%$ 。因此,在多数情况下,MLR都能很好地解决区域站日气温数据缺测的问题,而对缺测过多的情况,可采用IDW+MLR对缺测数据进行插补。

### 3 结论与讨论

本文对比了IDW、OK和MLR三种方法对湖北省区域站日气温缺测插补的适用性,以2020年完整率大于 $90\%$ 的区域站为例,从空间分布和季节变化两个方面分析了3种插补方法得到的日平均气温( $T$ )、日最高气温( $T_{\max}$ )和日最低气温( $T_{\min}$ )的平均绝对误差( $E$ ),再通过对试验研究了样本数量对插补结果误差的影响,主要结论如下:

(1) 基于MLR对区域站插补试验的当年 $T_{\max}$ 、 $T_{\min}$ 和 $T$ 的 $E$ 分别为 $0.41\text{ }^{\circ}\text{C}$ 、 $0.31\text{ }^{\circ}\text{C}$ 和 $0.20\text{ }^{\circ}\text{C}$ ,用IDW插补的 $T_{\max}$ 、 $T_{\min}$ 和 $T$ 的 $E$ 分别为 $0.70\text{ }^{\circ}\text{C}$ 、 $0.68\text{ }^{\circ}\text{C}$ 和 $0.47\text{ }^{\circ}\text{C}$ ,用OK插补的 $T_{\max}$ 、 $T_{\min}$ 和 $T$ 的 $E$ 分别为 $1.01\text{ }^{\circ}\text{C}$ 、 $0.85\text{ }^{\circ}\text{C}$ 和 $0.73\text{ }^{\circ}\text{C}$ 。MLR、IDW和OK插补的 $T$ 的 $E$ 在 $1.0\text{ }^{\circ}\text{C}$ 以内的站点比例分别为 $100\%$ 、 $84.01\%$ 和 $75.75\%$ ;  $E$ 在 $2.0\text{ }^{\circ}\text{C}$ 以上的站点比例分别为 $0$ 、 $0.48\%$ 和 $7.31\%$ 。无论对 $T_{\max}$ 、 $T_{\min}$ 或 $T$ 的插补结果,均是MLR最优,IDW次之,OK最差。

(2) 基于MLR插补的 $T$ 受海拔高度影响较小,其 $E$ 小于 $1.0\text{ }^{\circ}\text{C}$ ,而基于IDW插补的 $T$ 在海拔 $2\text{ }000\text{ m}$ 以上 $E > 1.0\text{ }^{\circ}\text{C}$ ,基于OK插补的 $T$ 在海拔 $1\text{ }500\text{ m}$ 以上 $E >$

$2.0\text{ }^{\circ}\text{C}$ 。另外,用MLR插补的 $T$ 、 $T_{\max}$ 和 $T_{\min}$ 的 $E$ 均随季节变化幅度最小,用IDW插补得到的 $T$ 和 $T_{\min}$ 的 $E$ 变幅最大,用OK插补得到的 $T_{\max}$ 的 $E$ 变幅最大。这表明MLR回归模型受海拔高度和季节变化影响较小,其插补气温度差的时空分布均匀,相比IDW和OK,MLR用于气温插补更具稳定性。

(3) 对于长时间连续气温缺测,MLR的插补效果仍是最好的,其插补结果主要受建模样本影响,即样本数量越多、时间离散度越大时,模型插补结果越准确、插补误差越小。

行业标准《地面气象观测规范》(中国气象局,2003)对地面气象观测仪器基本技术性能的要求是,我国自动气象站对气温的分辨力为 $0.1\text{ }^{\circ}\text{C}$ ,可达到的业务准确度为 $\pm 0.2\text{ }^{\circ}\text{C}$ 。据此,从本文对区域自动气象站日气温缺测数据插补方法的比较研究结果看,MLR对日平均气温的插补总体上在可接受范围内,且因气温在时空上具有较好的连续性,其邻近站气温存在较好的线性关系,此法相比其它方法适用性更强。需要指出的是,上述研究结果仅是以2020年湖北省气温观测为例获得的,各种插补方法的优劣还需在其它年份作进一步试验评估;另外,各种插补方法是否适用于对区域站其它气象要素(如降水)数据缺测的插补、其效果如何,还有待于今后开展相关试验研究,从而获得各种高质量地面气象观测资料,为推进气象灾害风险评估工作提供数据支撑。

### 参考文献(References):

- 郭彦,林秀芝,侯素珍,等.2015.基于EMD和BP算法的降水数据插补[J].水资源与水工程学报,26(2):16—21. Guo Y, Lin X Z, Hou S Z, et al. 2015. Interpolation of precipitation data based on empirical mode decomposition and back propagation algorithm [J]. Journal of Water Resources and Water Engineering, 26(2):16—21 (in Chinese). doi: 10.11705/j.issn.1672-643X.2015.02.04
- 黄俊杰,周悦,周月华,等.2018.近46 a汉江流域地表干湿状况变化及其影响因素[J].暴雨灾害,37(2):181—186. Huang J J, Zhou Y, Zhou Y H, et al. 2018. Variation of surface dry and wet conditions and its influencing factors in the Hanjiang River basin from 1971 to 2016 [J]. Torrential Rain and Disasters, 37(2):181—186 (in Chinese). doi: 10.3969/j.issn.1004-9045.2018.02.010
- 姜晓剑,刘小军,黄芬,等.2010.逐日气象要素空间插值方法的比较[J].应用生态学报,21(3):624—630. Jiang X J, Liu X J, Huang F, et al. 2010. Comparison of spatial interpolation methods for daily meteorological elements [J]. Chinese Journal of Applied Ecology, 21(3):624—630. (in Chinese) doi:10.13287/j.issn.1001-9332.2010.0088
- 姬兴杰,左璇,徐文慧.2020.中岳嵩山高山站平均气温资料插补及其变化趋势分析[J].气象与环境学报,36(6):58—67. Ji X J, Zuo X, Xu W H. 2020. Interpolation and variation trend analysis of air temperature data at Songshan mountain station in Henan province [J]. Journal of Meteorology and Environment, 36(6):58—67 (in Chinese). doi: 10.13287/j.issn.1001-9332.2020.0067

- rology and Environment, 36(6): 58–67 (in Chinese). doi:10.3969/j.issn.1673-503X.2020.06.007
- 马诺,唐治,常存,等.2020.基于DEM的气温空间插值方法比较[J].干旱气象,38(3):465–471. Ma N, Tang Y, Chang C, et al. 2020. Comparison of spatial interpolation methods of temperature based on DEM data [J]. Journal of Arid Meteorology, 38(3):465–471 (in Chinese). doi: 10.11755/j.issn.1006-7639(2020)-03-0465
- 李楠,崔耀平,刘素洁,等.2018.基于多源遥感数据的地表温度空值插补[J].测绘,41(2):57–61. Li N, Cui Y P, Liu S J, et al. 2018. The interpolation of land surface temperature based on multi-source remote sensing data [J]. Surveying and Mapping, 41(2):57–61 (in Chinese)
- 李框宇,周梅,陈玖英,等.2019.一种适用于气温空间插值的改进梯度距离平方反比法[J].中国科学院大学学报,36(4):491–497. Li K Y, Zhou M, Chen J Y, et al. 2019. An approach of improved gradient plus inverse distance squared for spatial interpolation of temperature [J]. Journal of University of Chinese Academy of Sciences, 36(4):491–497 (in Chinese). doi: 10.7523/j.issn.2095-6134.2019.04.008
- 李庆祥, Matthew J M, Claude N W, 等. 2005. 利用多模式对中国气温序列中不连续点的检测[J]. 气候与环境研究, 10(4):736–742. Li Q X, Matthew J M, Claude N W, et al. 2005. Detection of discontinuities in Chinese temperature series using a multiple test approach [J]. Climatic and Environmental Research, 10(4):736–742 (in Chinese). doi: 10.3969/j.issn.1006-9585.2005.04.004
- 娄珊珊,陈光舟,陆雅君,等.2019.安徽省近40 a 极端温度事件的气候特征分析[J].暴雨灾害,38(1):72–78. Lou S S, Chen G Z, Lu Y J, et al. 2019. Climate characteristics of extreme temperature events in recent 40A in Anhui [J]. Torrential Rain and Disasters, 38(1):72–78 (in Chinese). doi: 10.3969/j.issn.1004-9045.2019.01.009
- 刘宇,陈泮勤,张稳,等.2006.一种地面气温的空间插值方法及其误差分析[J].大气科学,30(1):146–152. Liu Y, Chen P Q, Zhang W et al. 2006. A spatial interpolation method for surface air temperature and its error analysis [J]. Chinese Journal of Atmospheric Sciences, 30(1): 146–152 (in Chinese). doi: 10.3878/j.issn.1006-9895.2006.01.12
- 潘霄,叶小岭,熊雄,等.2019.基于三次样条插值的探空气温质量控制研究[J].气象研究与应用,40(2):90–93,97. Pan X, Ye X L, Xiong X, et al. 2019. Research on quality control of air temperature based on cubic spline interpolation [J]. Journal of Meteorological Research and Application, 40(2): 90–93, 97 (in Chinese). doi: CNKI:SUN:GXQX.0.2019-02-021
- 陶玮,郭婷,吴瑞姣,等.2021.安徽省暴雨灾害预警等级的划分及其应用[J].暴雨灾害,40(2):174–181. Tao W, Guo T, Wu R J, et al. 2021. Classification of early warning levels for rainstorm disaster and its application in Anhui Province [J]. Torrential Rain and Disasters, 40(2):174–181 (in Chinese). doi: 10.3969/j.issn. 1004-9045.2021.02.009
- 王海军,杨志彪,杨代才,等.2007.自动气象站实时资料自动质量控制方法及其应用[J].气象,33(10):102–109. Wang H J, Yang Z B, Yang D C, et al. 2007. The method and application of automatic quality control for real time data from automatic weather stations [J]. Meteorological Monthly, 33(10):102–109 (in Chinese). doi: 10.3969/j.issn.1000-0526.2007.10.015
- 王海军,涂诗玉,陈正洪.2008.日气温数据缺测的插补方法试验与误差分析[J].气象,34(7):83–91. Wang H J, Tu S Y, Chen Z H. 2008. Interpolation method for missing data of daily air temperature and its error analysis [J]. Meteorological Monthly, 34(7):83–91 (in Chinese). doi: 10.7519/j.issn.1000-0526.2008.7.012
- 王少影,张宇,孟宪红,等.2020.机器学习算法对涡动相关缺失通量数据的插补研究[J].高原气象,39(6):1348–1360. Wang S Y, Zhang Y, Meng X H, et al. 2020. Fill the gaps of eddy covariance fluxes using Machine Learning Algorithms [J]. Plateau Meteorology, 39(6): 1348–1360 (in Chinese). doi: 10.7522/j.issn.1000-0534.2019.00142
- 王新宇,黄鹏程.2020.基于GIS的气象要素插值方法比较研究[J].测绘与空间地理信息,43(5):167–170. Wang X Y, Huang P C. 2020. Comparative study of interpolation methods of meteorological factors based on GIS [J]. Geomatics & Spatial Information Technology, 43(5):167–170 (in Chinese). doi: CNKI:SUN:DBCH.0.2020-05-047
- 岳文泽,徐建华,徐丽华.2005.基于地统计方法的气候要素空间插值研究[J].高原气象,24(6):974–980. Yue W Z, Xu J H, Xu L H. 2005. A study on spatial interpolation methods for climate variables based geostatistics [J]. Plateau Meteorology, 24(6):974–980 (in Chinese). doi: 10.3321/j.issn.1000-0534.2005.06.018
- 张莉莉,陶忠良,张京红.2012.基于GIS海南岛气象要素空间插值法比较分析[J].热带农业科学,32(4):57–61. Zhang L L, Tao Z L, Zhang J H. 2012. Analysis and comparison of the spatial interpolation methods for climate factors in Hainan Island [J]. Chinese Journal of Tropical Agriculture, 32(4): 57–61 (in Chinese). doi: 10.3969/j.issn.1009-2196.2012.04.012
- 中国气象局.2003.地面气象观测规范[M].北京:气象出版社. China Meteorological Administration. 2003. The criterion of surface meteorological observation [M]. Beijing: China Meteorological Press
- Acock M C, Pachepsky Y A. 2000. Estimating missing weather data for agricultural simulations using group method of data handling [J]. Journal of Applied Meteorology, 39(7):1176–1184. doi: 10.1175/1520-0450(2000)0392.0.CO;2
- Huth R, Nemešová I. 1995. Estimation of missing daily temperatures: Can a weather categorization improve its accuracy? [J]. Journal of Climate, 8(7):1901–1916. doi:10.1175/1520-0442(1995)008<1901:EOMDTC>2.0.CO;2
- Li J, Heap A D. 2014. Spatial interpolation methods applied in the environmental sciences: A review [J]. Environmental Modelling & Software, 53: 173–189. doi: 10.1016/j.envsoft.2013.12.008
- Matheron G. 1963. Principles of Geostatistics [J]. Economic Geology, 58: 1246–1266. doi: 10.2113/gsecongeo. 58.8.1246
- Nalder I A, Wein R W. 1998. Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest [J]. Agricultural and Forest Meteorology, 92(4): 211–225. doi: 10.1016/S0168-1923(98)00102-6
- Stooksbury D E, Idso C D, Hubbard K G. 1999. The effects of data gaps on the calculated monthly mean maximum and minimum temperatures in the Continental United States: A spatial and temporal study [J]. Journal of Climate, 12(5):1524–1533. doi: 10.1175/1520-0442(1999)0122.0.CO;2
- Thornton P E, Running S W, White M A. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain [J]. Journal of Hydrology, 190(3–4): 214–251. doi: 10.1016/S0022-1694(96)03128-9