

陕西省月平均气温空间插值方法研究

马秀霞, 黄领梅, 沈冰

(西安理工大学 西北旱区生态水利工程国家重点实验室培育基地, 陕西 西安 710048)

摘要: 为弥补陕西省气象站点数量有限、空间分布不均等问题,对常用空间插值方法进行比较研究,探寻适宜陕西省气温空间插值方法。根据陕西省内外 110 个气象站点 1981-2010 年的月平均温度,选用反距离权重法、普通克里金法与多元线性回归法进行空插值,利用实际验证法进行精度评估。结果表明,考虑了经度、纬度和高程的多元线性回归法的插值精度最高,普通克里金法次之,反距离权重法精度最低;多元线性回归分析法对陕西省气温插值最为适宜。

关键词: 月平均气温;空间插值;多元线性回归法;普通克里金法;反距离权重法

中图分类号:P468.0⁺21

文献标识码:A

文章编号:1672-643X(2017)05-0100-06

Study on spatial interpolation method of monthly mean temperature in Shaanxi Province

MA Xiuxia, HUANG Lingmei, SHEN Bing

(State Key Lab Cultivation Base of Northwest Arid Ecology and Hydraulic Engineering,
Xi'an University of Technology, Xi'an 710048, China)

Abstract: In order to compensate for the limited number of meteorological stations in Shaanxi Province and the uneven distribution, this paper makes a comparative study on the common spatial interpolation methods to explore the suitable spatial interpolation method of air temperature for Shaanxi Province. According to the monthly average temperature of 110 meteorological stations in Shaanxi Province from 1981 to 2010, the inverse distance weight (IDW), ordinary Kriging (OK) and multiple linear regression (MLR) were used to perform the null interpolation, and the accuracy was evaluated by the actual verification method. The results show that the interpolation accuracy of the MLR, which takes into account the longitude, latitude and elevation, is the highest, the accuracy of the OK ranks the second, and the IDW has the lowest precision. The MLR is most suitable for the temperature interpolation in Shaanxi Province.

Key words: monthly mean temperature; spatial interpolation; Multiple Linear Regression(MLR); Ordinary Kriging(OK); Inverse Distance Weighted(IDW)

1 研究背景

全球变暖已经是一个不争的事实^[1],近年来高温热浪、洪涝灾害、台风飓风等极端事件频发与全球变暖有一定的联系^[2-4]。为了更好地应对气候变化,需要掌握气象要素的变化规律,预测其发展趋势。对气象要素的研究依赖于实测站点的观测数据,而实际中气象站点的数量有限,把握气象要素变化规律则须对实测数据进行插值。

气温空间插值是基于有限的气象观测数据获得某一地区的气温数据分布规律,从而对未知点的数据进行预测^[5]。20世纪90年代随着GIS的发展与成熟,反距离权重、样条函数、克里金等空间插值方法广泛应用于土壤、农业、水文、生态、海洋、气象等领域,也为气候要素的空间插值提供了新方法。利用GIS对气温插值的精度依赖于站点的密度,然而,我国气象台站空间分布不均、密度不足^[6]。通常仅提供离散的、不规则分布的气象台站数据,难以反映

收稿日期:2017-04-25; 修回日期:2017-06-30

基金项目:国家自然科学基金项目(51679184);陕西省水利厅项目(2016slkj-12)

作者简介:马秀霞(1993-),女,宁夏中宁人,硕士研究生,研究方向为水文及水资源。

通讯作者:黄领梅(1972-),女,四川乐至人,博士,副教授,硕士生导师,研究方向为水文及水资源。

气候空间变化连续过渡的基本特征,每一个站点只能代表一定区域范围内的气象要素值^[7]。另一方面影响气温空间分布的因素很多,其中以海拔高度和地形条件的影响最显著^[8]。一般情况下,随着海拔高度的增加,气温下降,但其变化速率因山地性质和气候条件而不同。对于山区气温分布来说,不能忽视海拔高度对气温的影响,考虑海拔高度的影响可以明显提高气温的插值精度^[9]。结合数字高程模型(DEM),在分析平均气温与经度、纬度、海拔高度、坡度、坡向地形要素相关关系的基础上,应用多元回归模型取得了较好的插值效果^[5,10-13]。

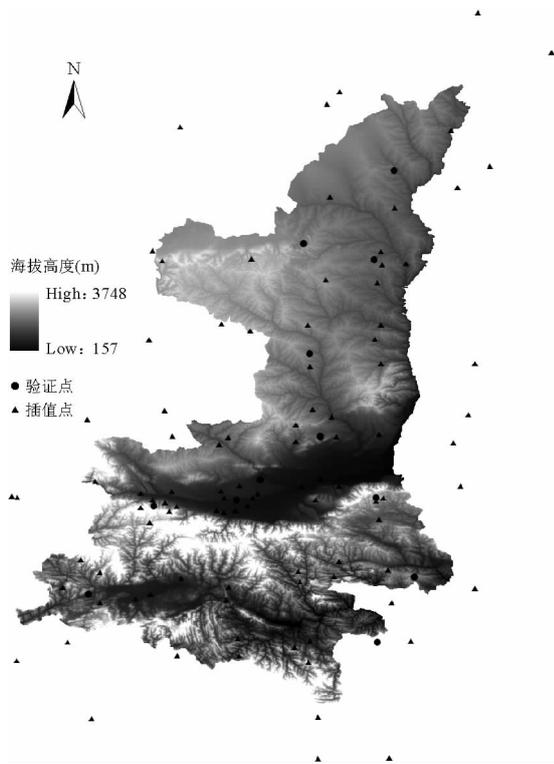


图1 陕西省内及其周边气象站点分布图

本文以陕西省为例,选取了陕西省内及周边气象站点的实测月平均气温,应用反距离权重法、普通克里金法、多元回归分析法进行插值,结合实际验证方法,以期获得适合陕西省月平均气温插值的计算方法;为进一步了解陕西省的月平均气温分布规律及应对气候变化提供参考依据。

2 研究区概况

陕西省地域狭长,地势南北高、中间低、由西向东倾斜。有高原、山地、平原和盆地等多种地形。作为中国南北气候分界线的秦岭山脉横贯全省东西。以秦岭为界,陕西省南部为长江水系,北部属黄河水系。由于陕西省横跨3个气候带,受高原地形和东

亚季风等影响,陕西南北天气气候复杂多变,降水时空分布极为不均,气候变化趋势差异明显^[14]。

实验数据下载于中国气象科学数据共享服务网,站点包括陕西省内及周边110个气象站(分布见图1),研究期为1981-2010年,时段为月。基础数据为研究期内的各站逐日平均气温,此外还包括各气象站点的经度、纬度以及海拔高度。其中,98个站点(占89%)的数据用于插值模型的构建,神木、横山、子洲、甘泉、白水、宝鸡县、礼泉、耀县、华阴、勉县、商南和白河12个站点(占11%)的数据用于插值模型的验证。

3 研究方法

空间插值就是根据已知的离散数据或分区数据,按照特定的数学关系推求出其他未知点或未知区域数据的计算过程。空间插值方法大体可分为全局方法和局部方法。全局方法是指利用所有已知样点的特征值估算未知点的数值;局部插值的方法是用邻近于未知样点的少数已知点的特征值来估算该未知点的特征值,只考虑插值区域内的局部特性而不受其他区域的影响^[15]。趋势面法、回归法等属于全局法;反距离权重法、样条函数法、克里金法等属于局部法。

3.1 反距离权重法

反距离权重法(Inverse Distance Weighted,简称IDW),其原理是已知样本点对插值样本点的影响随着距离增大而减小,用权重表示其影响程度,距离越大,对权重赋予权值越小,影响越小。计算公式为:

$$Z = \frac{\sum_{i=1}^n \left(z_i \frac{1}{d_i^k} \right)}{\sum_{i=1}^n \frac{1}{d_i^k}} \quad (1)$$

式中: Z 为区域内未知点的值; n 为区域内对未知点有影响的已知点数; z_i 为第 i 个已知点的值; d_i 为第 i 个已知点到该未知点的距离; k 为指定的幂。插值时应选择平均绝对误差值最小的 k 值,一般取2~3^[16],本文 k 值取Arcgis的统计模块Geostatistical Wizard计算出的最优值。

3.2 普通克里金法

普通克里金法(Ordinary Kriging,简称OK),是克里金法(Kriging)的一种,亦称为局部估计或空间局部插值法,是地统计学的主要方法之一。该方法考虑了数据的空间自相关性,能够对已知采样点的误差做出理论估计。其实质是利用区域化变量的取

值进行线性无偏、最优估计^[17]。

假设 x 为研究区域内的任一点, $Z(x)$ 是该点的测量值, 在研究区域内共有 n 个实测点, 即 x_1, x_2, \dots, x_n , 那么, 对于任意待估点或待估块 V 的实际值 $Z_V(x)$, 其估计值 $Z_V^*(x)$ 可通过待估点或待估块段内的 n 个有效样本值 $Z_V(x_i)$ ($i = 1, 2, \dots, n$) 的线性组合来表示, 即

$$Z_V^*(X) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (2)$$

$$\sum_{i=1}^n \lambda_i = 1 \quad (3)$$

式中: λ_i 为权重系数, 是各已知样本 $Z(x_i)$ 在估计值 $Z_V^*(x)$ 影响大小的系数, 而 $Z_V^*(x)$ 的好坏取决于怎样计算或选择权重系数 λ_i , 通常权重系数由方程组(2)、(3) 联立解出^[18]。

设 $Z(x)$ 为区域化变量, 满足二阶平稳和本征假设, 其数学期望为 m , $Z(x)$ 在空间位置 x 和 $x+h$ 处的观测值分别为 $Z(x)$ 和 $Z(x+h)$, 其协方差函数

$c(h)$ 和变异函数 $\gamma(h)$ 存在。即:

$$c(h) = E[Z(x)Z(x+h)] - m^2 \quad (4)$$

$$\gamma(h) = \frac{1}{2}E[Z(x) - Z(x+h)]^2 \quad (5)$$

式中: $\gamma(h)$ 值越大, 表明点与点之间的相关性越小。

3.3 多元线性回归法

气温的插值结果不仅受地域的影响, 还受海拔高度及下垫面等的影响^[19]。当自变量较多时, 可利用多元线性回归法(Multiple Linear Regression, 简称 MLR)分析各自变量对温度的影响。将经度(x_1 , ($^\circ$))、纬度(x_2 , ($^\circ$))、海拔(x_3 , m) 作为自变量, 以多年月平均气温 T_i (摄氏度) 为因变量, 建立多元线性回归模型:

$$T_i = a_{1i}x_{1i} + a_{2i}x_{2i} + a_{3i}x_{3i} + c \quad (6)$$

式中: T_i 为第 i 个月的平均温度; a_{1i} 、 a_{2i} 、 a_{3i} 分别为第 i 个月的经度、纬度及海拔的回归系数; c 为常数项, 各系数及常数项均无量纲。本文运用 SPSS 软件中的“多元线性回归分析”工具, 求得各月回归方程系数见表 1。

表 1 各月多元线性回归方程系数表

月份	常数项	a_{1i}	a_{2i}	a_{3i}	R	月份	常数项	a_{1i}	a_{2i}	a_{3i}	R
1	77.967	-0.215	-1.541	-0.004	0.984	7	31.099	-0.104	0.263	-0.006	0.934
2	68.910	-0.240	-1.088	-0.004	0.982	8	35.027	-0.061	-0.03	-0.006	0.950
3	54.448	-0.202	-0.633	-0.005	0.971	9	31.083	-0.014	-0.211	-0.005	0.944
4	35.611	-0.087	-0.263	-0.005	0.948	10	35.529	-0.001	-0.557	-0.005	0.940
5	25.321	-0.067	0.132	-0.005	0.950	11	53.839	-0.059	-1.129	-0.004	0.966
6	25.443	-0.096	0.353	-0.006	0.918	12	64.598	-0.096	-1.492	-0.003	0.981

注: 表中的 R 表示相关系数。

3.4 插值结果验证方法

在对气象数据插值结果的精度进行检验时, 一般采用“实际验证”和“交叉验证”法^[9-10]。“实际验证”是将大部分气象站点的数据作为“训练样本”用于插值计算, 剩下少量的气象站点的数据样本作为“验证样本”, 不参加插值计算, 利用“训练样本”的站点数据进行插值, 将插值结果与“验证样本”的实际观测数据进行对比; “交叉验证”是假设每一个气象站点的实测值都是未知的, 利用周围若干已知站点进行插值模拟, 计算所有站点模拟值与实测值之间的差值的平均值作为交叉验证的结果^[20]。本文采用实际验证法。

选取平均绝对误差(mean absolute error, MAE) 和平方根误差(rootmean squared error, RMSE) 作为插值效果的评价指标。MAE 和 RMSE 的值越小, 表示插值的精度越高^[19]。MAE 和 RMSE 的计算公式分别为:

$$MAE = \frac{1}{n} \sum_{i=1}^n \text{ABS}(T_{oi} - T_{ei}) \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (T_{oi} - T_{ei})^2} \quad (8)$$

式中: T_{oi} 为第 i 个站点的气温观测值, $^\circ\text{C}$; T_{ei} 为第 i 个站点的气温预测值, $^\circ\text{C}$ 。

4 结果分析

对于选取的 110 个气象站 1981-2010 年逐日平均温度, 分别计算各站的多年月平均气温。其中 98 个气象站点的多年月平均温度用于 3 种方法的参数率定或系数确定; 然后用其余 12 个站点的多年月平均温度对插值结果进行验证。利用 98 个气象站点的各月平均温度插值出陕西省的各月平均温度。由于篇幅所限, 本文仅列举了 3 月、6 月、9 月及 12 月的插值结果, 见图 2 和图 3。由图 2 和 3 可知,

陕西省气温南北差异大,气温分界线大致沿秦岭山脉,秦岭以北区域温度较秦岭以南气温低。

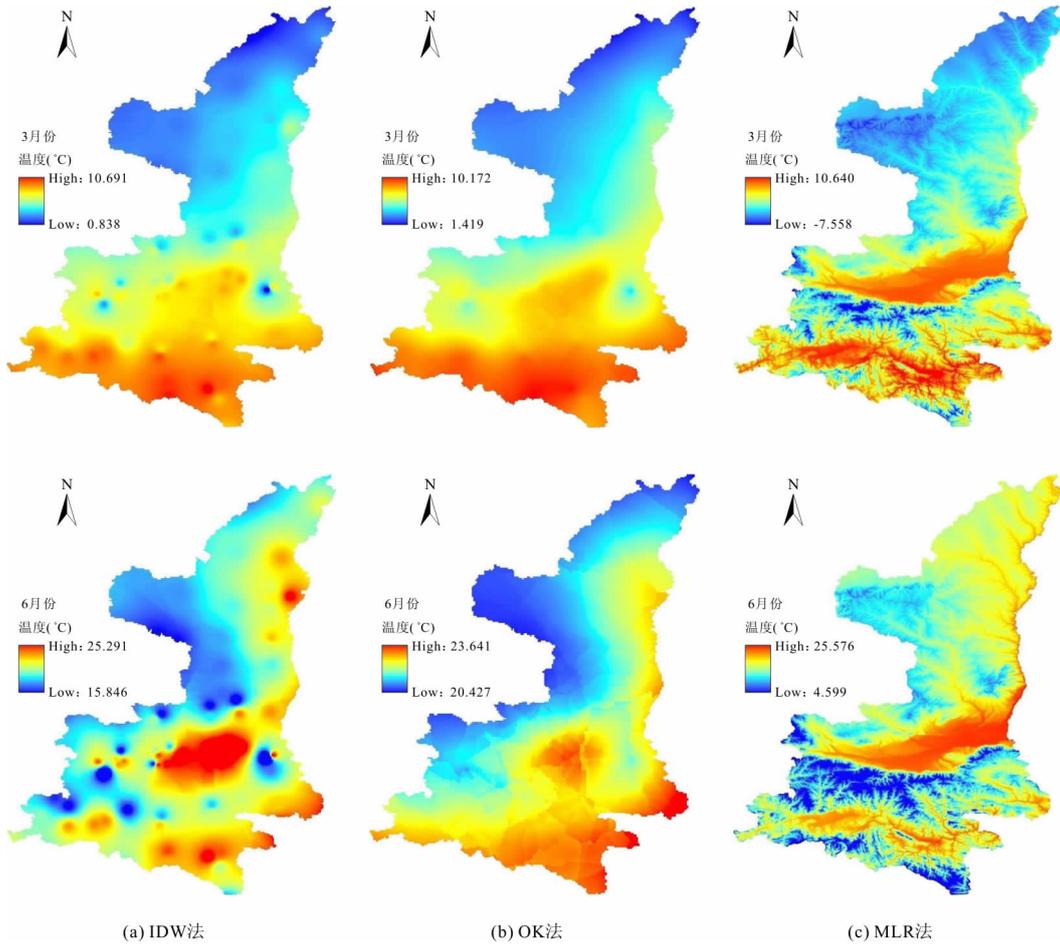


图 2 陕西省 3 月与 6 月各方法的月平均温度插值图

根据选用的验证方法,对 12 个验证站点月平均温度的插值与实测值,计算评价指标,结果见表 2。

表 2 3 种插值方法的评价指标 (MAE 和 RMSE) °C

月份	IDW 法		OK 法		MLR 法	
	MAE	RMSE	MAE	RMSE	MAE°C	RMSE
1	0.843	1.252	0.723	1.130	0.538	0.628
2	1.018	1.682	0.825	1.464	0.343	0.386
3	0.913	0.374	0.766	1.234	0.262	0.297
4	0.599	0.829	0.546	0.649	0.302	0.394
5	1.062	1.989	0.798	0.977	0.395	0.462
6	1.025	2.160	1.053	1.306	0.364	0.485
7	0.969	1.979	0.918	1.114	0.304	0.426
8	0.899	1.828	0.775	0.971	0.349	0.439
9	0.804	1.556	0.702	0.859	0.356	0.426
10	0.702	1.161	0.549	0.716	0.407	0.516
11	0.786	1.151	0.681	1.052	0.572	0.656
12	0.750	1.055	0.597	0.941	0.570	0.663
平均	0.864	1.418	0.744	1.034	0.397	0.482

由表 2 可知,多元线性回归法 (MLR 法) 优于反距离权重法 (IDW 法) 和普通克里金法 (OK 法)。对于平均绝对误差 MAE 指标, IDW 法的变化范围 0.599 ~ 1.062°C, 均值为 0.864°C; OK 法的变化范围 0.546 ~ 1.053°C, 均值为 0.744°C; 多元线性回归法的变化范围 0.262 ~ 0.572°C, 均值为 0.397°C, MAE 值由大到小为: IDW 法 > OK 法 > 多元线性回归法, 即反距离权重法的误差最大, 普通克里金法的误差次之, 多元线性回归法的误差最小。

对于平方根误差值 RMSE 指标, IDW 法的变化范围 0.374 ~ 2.160°C, 均值为 1.418°C; OK 法的变化范围 0.649 ~ 1.464°C, 均值为 1.034°C; 多元线性回归法的变化范围 0.297 ~ 0.663°C, 均值为 0.482°C。RMSE 值由大到小: IDW 法 > OK 法 > 多元线性回归法, 即反距离权重法的误差最大, 普通克里金法的误差次之, 多元线性回归法的误差最小。

采用 IDW 法插值时, IDW 法只考虑样本点与目标点的距离, 距离越近所占权重越大, 插值结果过分

依赖于样本点。其插值结果容易受极值影响,经常会有某一点的值高于周围的值,即“牛眼”现象;从图2和图3可以看到夏季与秋季的“牛眼”现象更

加明显。而造成该现象的另一原因是样本点分布不均,“牛眼”附近的样本点比较稀疏。

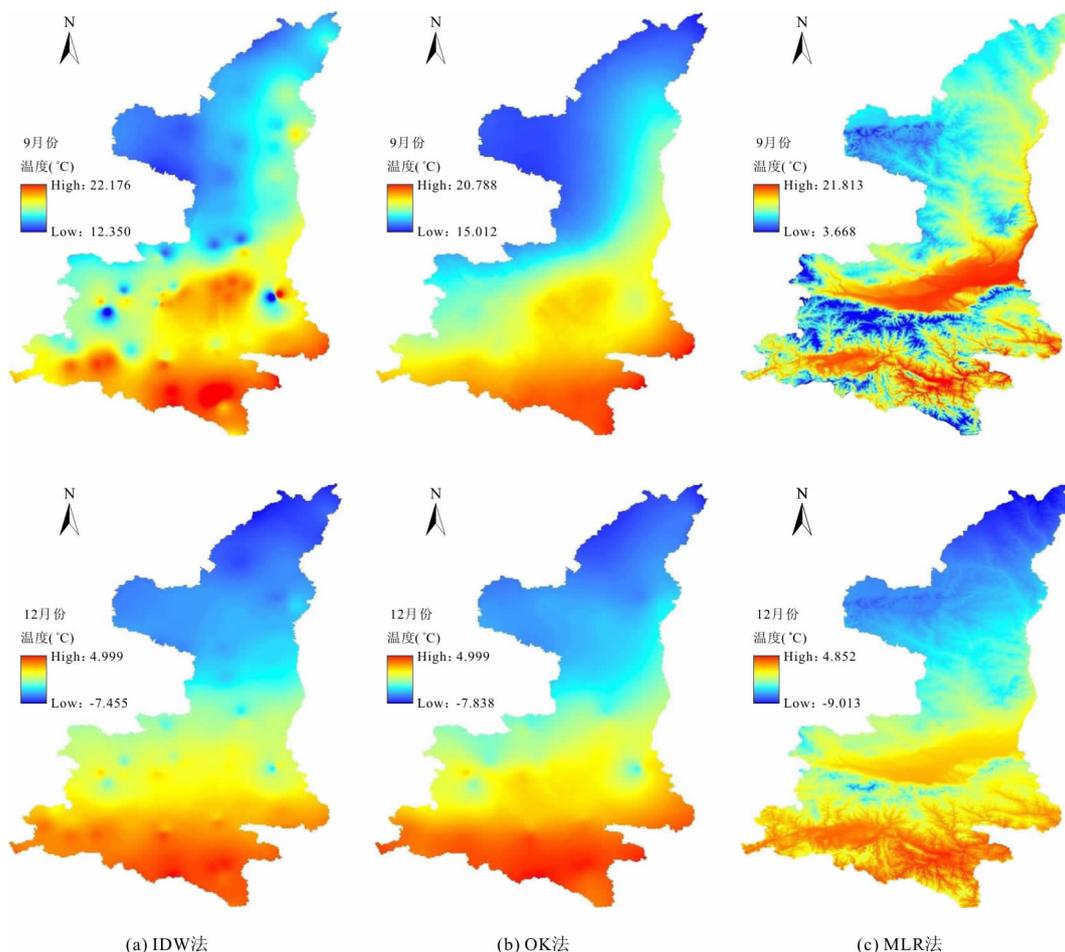


图3 陕西省9月与12月各方法的月平均温度插值图

OK法的插值精度相对IDW得到了提高,该法兼顾了样本点与插值点之间距离及样本点与插值点间的空间相关性,从而避免了目标点周边的部分样本点过于密集而形成样本族群的问题。但是当周边的样本点稀少时,用该法进行插值的优点得不到体现,且插值结果图会产生明显的凹凸现象。另该方法计算量大,须人为选定变异函数,变异函数可供选择的模型较多,如何选择模型争议较大^[16]。由于多元线性回归法属全局插值方法,利用全部样本点确定自变量的系数后,其插值结果不受样本密度与极值点的影响。可以预测许多未知区域的温度,且精度较高。

5 结论

(1) 本文综合分析了影响气温插值的因素,选

取了多元线性回归分析法、反距离权重法和普通克里金法,利用1981-2010年的实测气温数据,对陕西省月平均气温进行空间插值计算。以平均绝对误差和平方根误差作为插值结果精度评价的依据。结果表明3种方法均可以较好地应用于气温的预测。

(2) 通过对比分析发现多元线性回归分析法的预测效果优于反距离权重法和普通克里金法;该方法考虑了经纬度及海拔,能够直观地反映海拔与温度的关系。反距离权重法插值结果易受极值的影响,而陕西省地形复杂,多种地貌并存,温度差异较大,对插值精度影响较大。普通克里金插值法在样本容量较少的情况下,插值结果图会产生明显的凹凸现象。因此综合分析比较后,多元回归分析法更适用于陕西省的气温插值研究。

(3) 一种空间插值方法不能保证适用于所有案

例,对于具体的问题要具体分析。在今后研究中可尝试将多元线性回归法与其余方法耦合,取长补短;也可尝试在多元线性回归法中增加坡度、坡向等因素,建立更高精度的空间插值模型,为气象预报、工农业生产、分布式水文模型研究等提供数据支撑。

参考文献:

- [1] IPCC. Climate change 2013: the physical science basis [M]. Cambridge: Cambridge University Press, 2013.
- [2] 赵国永,韩艳,郭雪莲,等. 1961-2013年河南省气温要素时空变化特征[J]. 中国农学通报, 2015, 31(13): 249-257.
- [3] 张磊,潘婕,陶生才. 1961-2011年临沂市气温变化特征分析[J]. 中国农学通报, 2013, 29(5): 204-210.
- [4] 任璇,郑江华,穆晨,等. 不同气象插值方法在新疆草地NPP估算中的可靠性评价[J]. 草业科学, 2017, 34(3): 439-448.
- [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应用[J]. 生态学报, 2010, 21(3): 624-630.
- [6] 杜朝正. 基于GIS的传统插值方法比较研究——以山东省多年平均气温为例[J]. 安徽农业科学, 2013, 41(33): 12939-12941.
- [7] 李军,游松财,黄敬峰. 中国1961-2000年月平均气温空间插值方法与空间分布[J]. 生态环境, 2006, 15(1): 109-114.
- [8] 陈冬花,邹陈,王苏颖,等. 基于DEM的伊犁河谷气温空间插值研究[J]. 光谱学与光谱分析, 2011, 31(7): 1925-1929.
- [9] 蔡迪花,郭锐,李崇伟. 基于DEM的气温插值方法研究[J]. 干旱气象, 2009, 27(1): 10-17+28.
- [10] 王铁男,范永刚,李兄莲. 基于地形因素插值分析内蒙古敖汉旗气温要素分布[J]. 北京农业, 2016(2上旬刊): 148-150.
- [11] 石大明,姜忠宝,张晨琛. 吉林省站点气象要素精细化插值方法研究[J]. 吉林气象, 2015, 22(4): 36-38.
- [12] 白红英,马新萍,高翔,等. 基于DEM的秦岭山地1月气温及0℃等温线变化[J]. 地理学报, 2012, 67(11): 1443-1450.
- [13] 张海静,周秉荣,金元锋,等. 基于GIS技术的青海省最低气温空间插值方法探讨[J]. 草业科学, 2010, 27(9): 5-10.
- [14] 钱永兰,吕厚荃,张艳红. 基于ANUSPLIN软件的逐日气象要素插值方法应用与评估[J]. 气象与环境学报, 2010, 26(2): 7-15.
- [15] 于凤鸣,卓义,姜春龙,等. GIS空间插值方法在内蒙古气温插值中的应用[J]. 内蒙古气象, 2013(3): 11-14.
- [16] 李俊晓,李朝奎,殷智慧. 基于ArcGIS的克里金插值方法及其应用[J]. 测绘通报, 2013(9): 87-90+97.
- [17] 王智,师庆东,常顺利,等. 新疆地区平均气温空间插值方法研究[J]. 高原气象, 2012, 31(1): 201-208.
- [18] 谭剑波,李爱农,雷光斌. 青藏高原东南缘气象要素Anusplin和Cokriging空间插值对比分析[J]. 高原气象, 2016, 35(4): 875-886.
- [19] 许民,王雁,周兆叶,等. 长江流域逐月气温空间插值方法的探讨[J]. 长江流域资源与环境, 2012, 21(3): 327-334.
- [20] 刘宇,陈泮勤,张稳,等. 一种地面气温的空间插值方法及其误差分析[J]. 大气科学, 2006, 30(1): 146-152.