

文章编号:1671-9352(2008)03-0001-05

山东省多年气象要素空间插值方法比较研究

徐超,吴大千,张治国*

(山东大学生命科学院,山东 济南 250100)

摘要:针对山东省境内气象站点分布比较稀疏的情况,以1990~2001年山东省17个气象站点的年平均降水和温度为基础数据,分别建立了平均降水和温度与海拔、经纬度和植被指数(NDVI)之间的回归方程;在此基础上建立模拟站点以增加气象数据的信息量。分别采用反距离加权法(IDW)、径向基函数法(RBF)和普通克里格法(OK)3种插值方法进行比较插值分析。结果表明:模拟站点的加入大大提高了插值的精度;采用的3种插值方法中,普通克里格法比反距离加权法和径向基函数法具有更为理想的插值效果。

关键词:空间插值;温度;降水;模拟站点

中图分类号:P426.6; P423.5 **文献标志码:**A

Comparative study of spatial interpolation methods on weather data in Shandong Province

XU Chao, WU Da-qian, ZHANG Zhi-guo*

(School of Life Science, Shandong University, Jinan 250100, Shandong, China)

Abstract: Average precipitation and temperature data of 17 meteorological stations in Shandong Province from 1990 to 2001 were employed for spatial interpolation. The relationships between the average precipitation, temperature and elevation, normalized difference vegetation index (NDVI), longitude and latitude using stepwise regression equations were set up. Due to the sparse distribution of meteorological stations in Shandong Province, some suppositional meteorological stations were added to provide additional spatial information of meteorological data. Three methods, inverse distance weighted (IDW), radial basis function (RBF) and ordinary Kriging (OK), were adopted for interpolation. The results show that the suppositional meteorological stations can greatly improve the interpolation accuracy, and the ordinary Kriging method was the best among the three selected methods.

Key words: spatial interpolation; temperature; rainfall; suppositional meteorological stations

0 引言

随着现代生态学研究的发展,越来越多的区域蒸散和生态系统生产力模型被开发出来^[1],对空间化的气象要素数据的要求也越来越高。通过空间插值来生成所需区域的气象数据成为一种主要的解决途径。由于经济和人的原因,气象观测站点的数量是有限的,特别是具有长期观测记录的站点,且站

点的空间分布也是不均衡的^[2]。如何利用有限的资料得到科学的气象要素的空间分布规律,一直是气象学、生态学研究关注的热点^[3]。

针对这一问题,国内外学者做了大量的研究改进,研究思路大致可以分为2类:一类是方法和数据上的改进,增加与气象要素相关的地理信息或者对插值模型进行改进以提高插值精度^[4-7];另一类是在站点稀疏区域增加模拟站点以提高原观测数据的空间信息量,以达到更好的插值效果^[8]。

收稿日期:2007-06-20

基金项目:山东省环境保护科技计划基金资助项目(2006049)

作者简介:徐超(1984-),男,硕士研究生,现从事植被生态学与遥感应用的研究.Email: chaoxu@mail.sdu.edu.cn

*通讯作者:张治国(1963-),男,副教授,主要从事植物种群生态学、数量生态学和3S技术研究.Email: zgzhang@sdu.edu.cn

作为中国北方人口和经济大省的山东省,未曾做过相关的气象要素插值研究。山东省面积 0.1567 Mm^2 ,但只分布有 25 个气象站点,且分布不均匀。采用相关辅助地理信息及增加模拟站点的方法,对山东省多年气象数据的插值方法进行优化,建立山东省高分辨率的多年气温和降水栅格数据集,对于探索气象要素空间分布的规律具有理论和实践意义。

1 数据来源

收集了山东省内 25 个国家级气象站点的气象数据,其中 17 个站点具有 1961 ~ 2001 年的连续观测数据,而另外 8 个站点存在观测数据缺失的情况,因此选用其中 17 个气象站点 1990 ~ 2001 年 12 年间的年均降水、气温数据;高程模型(DEM)数据为美国马里兰大学全球土地覆盖数据库(the global land cover facility, GLCF)提供的 90 m 分辨率的数字高程图;另外还采用了同期的遥感植被指数数据,即 1990 ~ 2001 年逐旬的 NOAA/AVHRR-NDVI 数据,空间分辨率为 8 km ,此数据来自美国地球资源观测系统(the earth resources observation systems, EROS)数据中心的探路者数据库,经过大气纠正、云检验、质量控制、图像配准等处理流程^[9-11]。提取每年 NDVI 最大值,对 12 年的最大值进行平均处理后用于数据分析。采用对数变换方式,对降水数据进行了预处理以提高插值的精度^[12]。

2 方法

2.1 模拟站点的建立

对参与插值的 17 个实测站点进行地统计学分析,选择 40 km 作为抽样间隔,依照有效变异函数的建立原则,建立模拟站点。具体过程为:对原始观测站点生成 40 km 为半径的影响范围区,在无观测站点影响范围覆盖区域,增设以 40 km 为半径的圆形影响区,尽量将山东省内空白区域覆盖完全,取其圆心作为模拟站点的位置,以提高站点分布的均匀度,进而改进局部估计的精度^[8]。共增加了 20 个模拟站点,其分布如图 1 所示。

对多年平均温度(T)、降水(P)、站点经度(L_o)、纬度(L_a)、DEM 值(E)及 NDVI 最大值 N_{\max} 进行多元逐步回归分析,回归方程为

$$T = 41.438 - 0.147L_o - 0.206L_a - 0.015N_{\max} - 0.03E \quad (R^2 = 0.46, p = 0.042), \quad (1)$$

$$P = -3353.7 + 23.48L_o + 23.13L_a + 7.41N_{\max} -$$

$$5.20E \quad (R^2 = 0.57, p = 0.019). \quad (2)$$

由式(1)、(2)可以计算出模拟站点的多年平均温度和降水值。

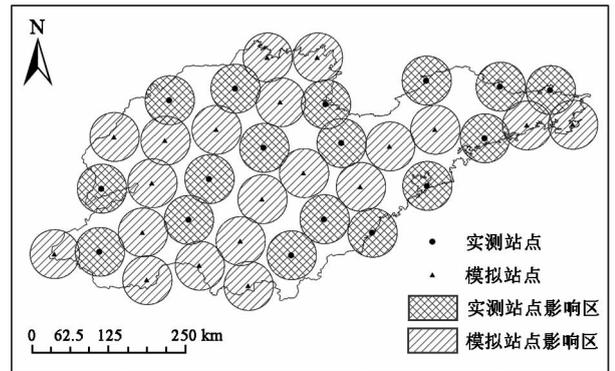


图 1 实测和模拟站点及其影响区分布图

Fig.1 Spatial distribution map of the actual meteorological stations and suppositional meteorological station sites and their affecting areas

2.2 插值方法的选择

2.2.1 反距离加权法(IDW, inverse distance weighted)

反距离加权插值法是基于相近相似原理:即 2 个物体离得越近,它们的性质就越相似;反之,离得越远则相似性越小。其一般公式为

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i). \quad (3)$$

式中, $\hat{Z}(s_0)$ 为 s_0 处的预测值; $Z(s_i)$ 是在 s_i 处获得的测量值; N 为预计算过程中要使用的预测点周围样点的数量; λ_i 为预测计算过程中使用的各样点的权重,确定权重的计算公式为

$$\lambda_i = d_{i0}^{-p} / \sum_{i=1}^N d_{i0}^{-p}, \quad (4)$$

$$\sum_{i=1}^N \lambda_i = 1. \quad (5)$$

式中, d_{i0} 为预测点 s_0 与各已知样点 s_i 之间的距离^[13]。随着样点与预测点之间距离的增加,权重值的降低用参数 p 来控制。通过求均方根预测误差(root-mean-square prediction error, RMSPE)的最小值,能够得到 p 的最佳值。一般情况下该值取 $2^{[13-16]}$,即反距离平方插值。

2.2.2 径向基函数法(radial basis functions, RBF)

所谓径向基函数,就是某种沿径向对称的标量函数。通常定义为空间中任一点 x 到某一中心 x_0 之间欧氏距离的单调函数,其作用往往是局部的,即当 x 远离 x_0 时函数取值很小。最常用的径向基函数是高斯核函数。

影响插值区域内各点预测值的主要因素有:平面坐标值 X, Y 及其邻近的几个样点的测量值。其中,

邻近的几个样点的测量值对该点的作用与距离 d 成负相关,另考虑计算的简单性,只取临近的 4 个样点,即对于插值点的值 Z ,若其最邻近的 4 个采样点的值为 Z_1, Z_2, Z_3, Z_4 ,距离为 d_1, d_2, d_3, d_4 。则有

$$Z = \sum_{p=1}^n Z_p G(d_p) / \sum_{p=1}^n G(d_p), \quad n = 4. \quad (6)$$

式中, $G(d)$ 为一递减函数,且与距离 d 成负相关。令 $G(d) = (1/d)k (k > 0)$, k 的确定是解决问题的关键。由于地理要素分布的复杂性,对于整个插值空间, k 并不是常数。设 Z_1, Z_2 和 Z_3 分别表示 $k = 1, 2, 3$ 时的 Z 值,以输入参量 (Z_1, Z_2, Z_3) 模拟邻近的几个未知点的值。

综上所述,以 (X, Y, Z_1, Z_2, Z_3) 为输入参量,用已知样点的值作为期望输出进行训练,可以模拟地表空间坐标的空间分布^[17]。

2.2.3 普通克里格法(ordinary Kriging, OK)

以区域化变量理论为基础,半变异函数为分析工具的克里格法在地统计学中被广泛使用,对空间分布具有随机性与结构性变量的研究具有其独特的优点^[18]。其插值公式为

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i). \quad (7)$$

式中, λ_i 是站点气象要素值 $Z(s_i)$ 的权重,用来表示各站点要素值 $Z(s_i)$ 对估计值 $\hat{Z}(s_0)$ 的贡献, s_i 表示站点的位置。为达到更好的插值效果,权重系数由“克里格方程组”决定。

$$\begin{cases} \sum_{i=1}^n \lambda_i C(s_i, s_j) - \mu = C(s_i, s_0), \\ \sum_{i=1}^n \lambda_i = 1. \end{cases} \quad (8)$$

式中, $C(s_i, s_j)$ 为样点间的协方差, $C(s_i, s_0)$ 为样点与插值点间的协方差, μ 为极小化处理时的拉格朗日乘子。协方差与半变异函数的关系为

$$C(h) = \delta^{*2} - \gamma(h). \quad (9)$$

式中, δ^{*2} 为实验方差。

克里格插值的权重取决于半变异函数,其表达式为

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(s_i) - Z(s_i + h))^2. \quad (10)$$

式中, $\gamma(h)$ 为变量 Z 以 h 为距离间隔的半方差, $N(h)$ 为被 h 分隔的实验数据对的数目。

本研究采用的是球形函数模型,表达式为

$$\gamma(h) = \begin{cases} C \left[\frac{3}{2} \frac{h}{\alpha} - \frac{1}{2} \frac{h^3}{\alpha^3} \right] & (h \leq \alpha), \\ C & (h > \alpha). \end{cases} \quad (11)$$

式中, α 为变程, C 为基台值。

2.3 插值效果验证

采用交叉验证法(cross-validation)来验证插值的效果^[19]。它通过逐一剔除站点的观测数据,利用其他站点数据进行模拟生成该站点的预测值,将其与实测数据进行比较,以分析估计误差在各个站点的分布情况。一般情况下,采用平均绝对误差(mean absolute error, MAE)和插值平均误差平方的平方根(root mean squared interpolation, RMSIE)作为评估不同插值方法的标准^[3]。MAE 可以估量模拟值可能的误差范围,RMSIE 可以反映利用样点数据的估值灵敏度和极值效应。其表达式为

$$MAE = \frac{\sum_{i=1}^n ABS(Z_{a,i} - Z_{e,i})}{n}, \quad (12)$$

$$RMSIE = \sqrt{\frac{\sum_{i=1}^n (Z_{a,i} - Z_{e,i})^2}{n}}. \quad (13)$$

式中, $Z_{a,i}$ 为第 i 个站点的实际观测值, $Z_{e,i}$ 为第 i 个站点的插值估计值。 n 为用于参与验证的站点的数目。

3 结果与分析

3.1 基于模拟站点的插值效果分析

使用 3 种插值方法(IDW, RBF, OK)分别对 17 个真实站点和真实站点加模拟站点共 37 个站点的多年温度和降水数据进行插值,温度和降水数据插值的交叉验证结果见表 1。

表 1 插值方法的交叉验证结果
Table 1 Cross-validation results of the interpolation methods

插值方法	降水/mm		温度/(°C)	
	MAE	RMSIE	MAE	RMSIE
IDW-R	49.31	68.53	0.4806	0.577
IDW-S	33.72	46.90	0.3154	0.392
RBF-R	50.27	65.26	0.4683	0.571
RBF-S	32.81	45.05	0.3153	0.396
OK-R	52.50	64.75	0.4263	0.503
OK-S	32.97	44.01	0.3032	0.384

注: IDW-R、RBF-R、OK-R 表示基于真实站点的插值方法, IDW-S、RBF-S、OK-S 表示基于真实站点加模拟站点的插值方法。

由表 1 中可以看出,基于真实站点加模拟站点的 3 种插值方法的 MAE 和 RMSIE 值均比基于真实站点的 3 种插值方法的 MAE 和 RMSIE 值要低,这说明在插值过程中加入模拟站点的方法可以显著地提高插值的效果。

3.2 基于模拟站点加真实站点的 3 种插值方法的误差比较分析

表 1 中给出了基于模拟站点加真实站点的 3 种

插值方法的交叉验证的结果。可以看出:对于温度插值,3种插值方法的MAE的排序为IDW > RBF > OK,其值分别为0.315 4 mm,0.315 3 mm和0.303 2 mm;RMSIE的排序为RBF > IDW > OK,其值分别为0.396 mm,0.392 mm和0.382 mm。OK方法的MAE和RMSIE均为最低,取得了较好插值效果。

降水插值的MAE的排序为IDW > OK > RBF,其值分别为33.72 °C,32.97 °C和32.81 °C;RMSIE的排序为IDW > RBF > OK,其值分别为46.90 °C,45.05 °C和44.01 °C。OK方法的MAE虽然不是最低值,但其值与RBF方法的MAE值极为接近,仅相差0.03 °C;而OK方法的RMSIE值取得了最低值,并且要比RBF方法低1.04 °C。对空间进行插值,最重要的是在保证插值精度的前提下,最大可能性的捕获插值变量的空间变化特征,在平均绝对误差(MAE)相差不大的情况下,OK方法的插值平均误差平方的平方根(RMSIE)明显优于RBF方法,说明OK方法在一定程度上优于RBF方法。

3.3 基于模拟站点加真实站点的普通克里格插值结果

基于模拟站点加真实站点的山东省多年平均降水和温度的普通克里格插值结果见图2。

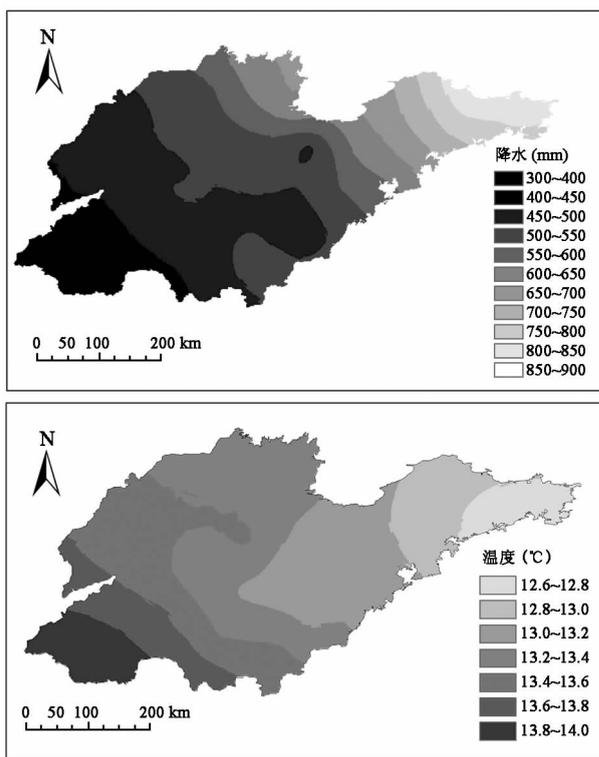


图2 基于模拟站点加真实站点的山东省多年平均降水、温度普通克里格插值栅格图

Fig. 2 Precipitation and temperature map in Shandong Province based on the ordinary Kriging method with both actual meteorological stations and suppositional meteorological stations involved

从图2中可以看出:山东省多年平均降水量的空间分布受海洋的影响较大,降水量由海岸向内陆递减,并且沿海岸线呈现明显的条带状分布。降水量最高的区域覆盖了威海市的大部分以及烟台、青岛市的小部分,其受海洋的影响出现较大的降水量;而降水量最小的区域出现在远离海岸线的菏泽市、聊城市和德州市的部分区域;在潍坊市境内出现了一个降水相对稀少的区域,可能是由于降水分布的纬度地带性和海洋气候的综合影响导致的。插值结果与山东省近几十年的年均降水的实际情况基本相吻合。

山东省多年平均温度的空间分布也同样受到海洋的影响,同时还受到地形因素的影响。由于海洋的影响,从海岸线到内陆呈现递减的规律。济南、淄博等城市的气温较高是由于这些城市的工业化和城市化程度较高引起的。由于泰山的原因导致泰安附近气温的降低趋势明显减缓。

4 讨论

气象要素插值方法发展至今,仍然无法摆脱气象站点缺少所带来的难题。结合增加辅助信息和增加辅助模拟站点2种方法的优点,以多元逐步回归模型为基础,增设模拟站点反演降水和温度数据,并用实测站点数据进行回归检验,以寻求达到较好的插值效果。

由于采用的3种插值方法均为基于搜索半径的插值方法,模拟站点的插入,增加了搜索半径内的气象站点个数,并且增加了气象数据的空间变异性信息,因而有利于提高插值的精度。插值结果表明:增设的模拟站点对插值结果有明显的改善,大大增加了插值的精度;3种插值方法的比较分析表明普通克里格方法具有较理想的插值效果。

引入第三方数据进行辅助插值需要有2个前提:第一,第三方数据必须有很好的空间分辨率和空间延续性;第二,第三方数据必须和插值变量存在真实可靠的显著的相关关系。采用海拔高度、经纬度和NDVI值作为第三方数据,这几个变量不仅有着很好的空间分辨率和空间延续性,而且它们对于降水和平均温度的影响是严格证明过的和广泛认可的。尽管在模拟站点数据反演的过程中使用到第三方数据,但这些数据并没有对原实测数据的变异特征产生显著的影响^[8],即插值方法仅利用了第三方数据加大信息量,并没有影响实测数据的空间变异特征。交叉验证结果明确证明了这一点。

由于受到数据资源的限制,在模拟站点的增设中只考虑了气象数据与海拔、经纬度和植被指数之间的关系,考虑更多的地学因子对气象的影响须进一步探讨。

参考文献:

- [1] 李忠武,蔡强国,唐政洪,等.作物生产力模型及其应用研究[J].应用生态学报,2002,13(9):1174-1178.
- [2] 尚宗波,高琼,杨奠安.利用中国气候信息系统研究年降水量空间分布规律[J].生态学报,2001,21(5):689-694.
- [3] 潘耀忠,龚道溢,邓磊,等.基于DEM的中国陆地多年平均温度插值方法[J].地理学报,2004,59(3):366-374.
- [4] MARQUENEZ J, LASTRA J, GARCA P. Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis[J]. Journal of Hydrology, 2003, 270(1):1-11.
- [5] 孙鹏森,刘世荣,李崇巍.基于地形和主风向效应模拟山区降水空间分布[J].生态学报,2004,24(9):1910-1915.
- [6] HOFIERKA J, PARAJKA J, MITASOVA H, et al. Multivariate interpolation of precipitation using regularized spline with tension[J]. Transactions in GIS, 2002, 6(2):135-150.
- [7] WONG K W, WONG P M, GEDEON T D, et al. Rainfall prediction model using soft computing technique[J]. Soft Computing, 2003, 7(6):434-438.
- [8] 辜智慧,史培军,陈晋.气象观测站点稀疏地区的降水插值方法探讨——以锡林郭勒盟为例[J].北京师范大学学报:自然科学版,2006,42(2):204-208.
- [9] TOWNSHEND J R G. Global data sets for land applications from the advance very high resolution radiometer: an introduction[J]. International Journal of Remote Sensing, 1994, 15(17):3319-3332.
- [10] JAMES M E, KALLURI S N V. The Pathfinder AVHRR land data set: An improved coarse resolution data set for terrestrial monitoring[J]. International Journal of Remote Sensing, 1994, 15(17):3347-3363.
- [11] SMITH P M, KALLURI S N V, PRINCE S D, et al. The NOAA/NASA Pathfinder AVHRR 8 km land data set[J]. Photogrammetric Engineering and Remote Sensing, 1997, 63(1):12-31.
- [12] MARC G G, REINHARD F. Analysis of rainfall data by robust spatial statistic using S+ SPATIALSTATS[J]. Journal of Geographic Information and Decision Analysis, 1998, 2(2):116-126.
- [13] 蔡福,于贵瑞,祝青林,等.气象要素空间化方法精度的比较研究——以平均气温为例[J].资源科学,2005,27(5):173-179.
- [14] PATRICK M B, KELLER C P. Multivariate interpolation to incorporate thematic surface data using inverse distance weighting(IDW)[J]. Computer & Geosciences, 1996, 22(7):795-799.
- [15] GOOVAERTS P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall[J]. Journal of Hydrology, 2000, 22(8):113-129.
- [16] 范银贵.空间插值方法在绘制降雨量等值线图中的应用[J].水利水电科技进展,2002,22(3):48-50.
- [17] 杨海荣.基于RBF神经网络的空间插值[J].长沙交通学院学报,2006,22(1):68-71.
- [18] 林忠辉,莫兴国,李宏轩,等.中国陆地区域气象要素的空间插值[J].地理学报,2002,57(1):47-56.
- [19] HOLDAWAY M R. Spatial modeling and interpolation of monthly temperature using Kriging[J]. Clim Res, 1996, 6:215-225.

(编辑:孙培芹)