

利用时空 Kriging 进行气温插值研究

李 莎^{1,2} 舒 红¹ 徐正全¹

(1 武汉大学测绘遥感信息工程国家重点实验室, 武汉市珞喻路 129 号, 430079)

(2 湖北第二师范学院机械与电气工程系, 武汉市光谷二路 29 号, 430205)

摘 要:以黑龙江省近 37 a 的月均气温数据为研究对象,介绍了一类积分式时空协方差(变异)函数模型进行时空 Kriging 插值。针对月均气温呈现出的明显季节变化,对各站点的气温进行去季节项处理,并在此基础上建立时空变异函数。将空间维的普通 Kriging 插值扩展至时空维,同时考虑空间和时间相关性对研究变量进行时空估计,并将估计结果与空间 Kriging 插值效果进行了比较。结果表明,时空插值效果理想,插值精度较空间 Kriging 更高。

关键词:Kriging 插值;时空相关;协方差;变异函数;气温

中图法分类号:P208

全球气候变暖给人类及生态系统带来了显著的影响,如极端天气、冰川消融、永久冻土层融化、海平面上升、生态系统改变等。因此国内外许多研究人员对气温数据进行了大量的研究,如 Willmott 和 Robeson 利用高分辨率气候站点生成的气温数据,并结合空间插值的年均气温偏差提出了气候辅助反距离权重插值方法^[1];Koike 等将神经网络的 Kriging 方法应用于温度空间分布的研究分析中^[2];Nalder 和 Wein 在进行气候数据插值研究时提出了梯度距离平方反比法(GIDS)^[3];陈华等利用 Mann-Kendall 检验方法和空间插值方法分析了汉江流域 53 a 的年和四季气温的时空变化趋势^[4]。

Kriging 估计法^[5,6]是一种常用的空间插值方法,源自于地统计学^[7,8],它以建立变量的协方差(变异)函数为前提,根据变量的空间自相关性对待估点进行插值。然而气温是时空分布的变量,它不仅具有空间相关性,在时间分布上也同样具有相关性。如果只采用空间相关性及空间邻近点进行插值估计,而忽略变量在时间分布上的有效信息,将不利于插值精度的提高。将 Kriging 法应用于时空变量的插值研究中,一方面需要将 Kriging 法进行时空扩展,另一方面,应该首先建

立有效的时空协方差(变异)函数模型。目前,关于时空协方差函数模型的研究主要可分为两大类:可分离型^[9,10]和不可分离型^[11-16]。前者主要通过将空间协方差函数与时间协方差函数相加或相乘得到,构建简易,但分割了时空空间的相关信息,主要应用于早期的时空变量插值中;而后者虽然构建相对复杂,但更有效地描述了变量的时空变异结构,因此,不可分离型模型愈来愈频繁地应用于时空统计研究中。本文以黑龙江省 32 个气象站点 1972~2008 年的月均气温数据为例,采用了一类积分变异函数模型来拟合变量的时空变异结构,将原本空间维的普通 Kriging 法扩展至时空维,考虑气温的时空相关性,进行时空插值研究。

1 研究区域概况

黑龙江省位于东经 121°11′~135°05′,北纬 43°25′~53°33′,是我国最东北部。北部、东部以黑龙江、乌苏里江为界,与俄罗斯相望,西部毗邻内蒙古自治区,南部与吉林省接壤,面积约 46 万 km²。该省地形复杂,西北部有大兴安岭山地,北部有小兴安岭山地,东南部有张广才岭、老爷岭、

太平岭和完达山等山地,西南有嫩江、松花江南北贯穿该省,形成了东北部的三江平原、西南部的松嫩平原,东南还有兴凯湖。黑龙江省是全国气温最低的省份,一月平均气温在 $-30.9\text{ }^{\circ}\text{C}\sim-14.7\text{ }^{\circ}\text{C}$,夏季平均气温在 $18\text{ }^{\circ}\text{C}$ 左右,极端最高气温达 $41.6\text{ }^{\circ}\text{C}$ 。

本实验数据取自中国气象科学数据共享服务网上黑龙江省1972年1月至2008年12月的月均气温数据,共32个气象站点,其中一站点的部分数据缺失,因此,实验中实际采用了31个站点的月均气温数据参与插值预测。

2 时空协方差函数

假设 $Z(s, t) \in R^k \times T$ 是定义在时空域的随机过程,其中 R^k 代表 k 维的欧氏空间, T 代表时间,同时定义 $(s_i, t_i) (i=1, \dots, n)$ 为时空场中任意样本点的位置, h_s 为样本点间的空间距离, h_t 为时间距离。当 $Z(s, t)$ 满足二阶平稳或固有假设时,其协方差函数记为:

$$C(h_s, h_t) = \text{cov}(Z(s+h_s, t+h_t), Z(s, t)) \quad (1)$$

对应的变异函数记为:

$$\gamma(h_s, h_t) = \frac{1}{2} E (Z(s+h_s, t+h_t) - Z(s, t))^2 = \sigma^2 - C(h_s, h_t) \quad (2)$$

式中, σ^2 为 $Z(s, t)$ 的方差。

由式(1)、式(2)可见,当随机过程 $Z(s, t)$ 满足二阶平稳或固有假设时,其协方差和变异函数只取决于样本点距离 h_s 和 h_t ,而与样本的具体位置无关。本文采用文献[13, 16]介绍的一类积分式协方差模型:

$$C_{s,t}(h_s, h_t) = \int_V [k_1 C_s(h_s; a) C_t(h_t; a) + k_2 C_s(h_s; a) + k_3 C_t(h_t; a)] d\mu(a) \quad (3)$$

对应的变异函数是:

$$\gamma_{s,t}(h_s, h_t) = \int_V [(k_2 + k_1 C_t(0; a)) \gamma_s(h_s; a) + (k_3 + k_1 C_s(0; a)) \gamma_t(h_t; a) - k_1 \gamma_s(h_s; a) \gamma_t(h_t; a)] d\mu(a) \quad (4)$$

其中, $d\mu(a) = \varphi(a) da$, $\varphi(a)$ 为对应的密度函数; $C_{s,t}$ 为时空协方差, C_s 为空间协方差, C_t 为时间协方差; $\gamma_{s,t}$ 、 γ_s 、 γ_t 分别是对应的变异函数; $C_{s,t}(0, 0)$ 、 $C_s(0)$ 、 $C_t(0)$ 分别是对应的基台值。由变异函数的性质可得: $\gamma_{s,t}(0, 0) = \gamma_s(0) = \gamma_t(0) = 0$,将 $\gamma_s(h_s; a)$ 、 $\gamma_t(h_t; a)$ 进行标准化,即设定其基台值 $C_s(0; a) = 1$, $C_t(0; a) = 1$,并设 $C_{s,0}(0, 0)$ 、 $C_{0,t}(0, 0)$ 分别是 $\gamma_{s,t}(h_s, 0)$ 、 $\gamma_{s,t}(0, h_t)$ 的基台值,可

得:

$$C_{s,t}(0, 0) = k_1 + k_2 + k_3 \quad (5)$$

$$C_{s,0}(0, 0) = k_1 + k_2 \quad (6)$$

$$C_{0,t}(0, 0) = k_1 + k_3 \quad (7)$$

3 时空 Kriging 插值

Kriging 插值是一种线性的无偏估计方法,要求估计误差的方差最小。本文采用普通 Kriging 进行时空扩展,即

$$Z^*(s_0, t_0) = \sum_{i=1}^n \lambda_i Z(s_i, t_i) \quad (8)$$

式中, $Z^*(s_0, t_0)$ 为时空点 (s_0, t_0) 处的估计值; λ_i 为邻近观测值 $Z(s_i, t_i)$ 的加权系数。Kriging 插值是以协方差(或变异)函数为基本前提的,式(8)中,参数 λ_i 由时空变异函数确定。引入拉格朗日系数 μ 可得参数 λ_i 的计算式为:

$$\sum_{i=1}^n \lambda_i \gamma[(s_i, t_i) - (s_j, t_j)] + \mu = \gamma[(s_j, t_j) - (s_0, t_0)], j = 1, \dots, n$$

$$\sum_{i=1}^n \lambda_i = 1$$

与空间 Kriging 法不同,时空插值时,空间和时间上的相邻样本点都参与估计,如对泰来站点2007年10月的月均气温进行插值估计,所选的邻近观测点有克山、富裕、齐齐哈尔、明水、哈尔滨的当月数据以及泰来2007年8、9、11、12月的数据。

4 实验及结果分析

4.1 去季节变化

气温的时空分布并不平稳,空间上呈现由北向南逐渐递增的趋势,但差异相对不大,而时间分布上具有显著的季节变化特点(如图1所示)。这种周期变化显著影响了变量分布的平稳性,同时造成时间变异函数拟合时出现明显的孔穴效应,不利于时空变异函数的建模,因此在插值实验时,应首先将季节因素去掉,待插值之后再将对应的季节项加上。从图1可知,原始气温随着一年四季呈周期变化,冬季气温最低,夏季气温最高,采用时序分解中的加法模型,将变量的季节因素提取出来,余下的部分用于时空插值实验。用自相关图检法^[17]判断去季节项数据的平稳性,如图2所示,序列的自相关系数(ACF)随延迟时期数很快衰减到 ± 0.1 以内,这说明去季节项数据近似

平稳。

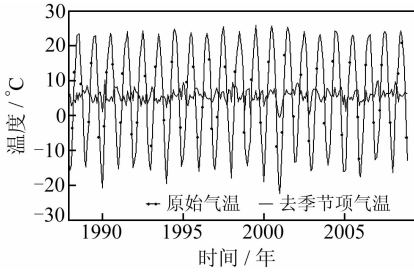


图 1 泰来站点的气温时序分布图

Fig. 1 Temperature Distribution in Time of Tailai Station

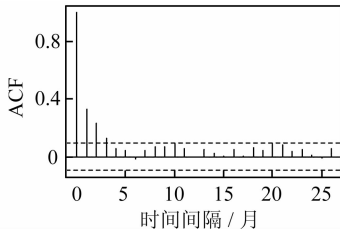


图 2 气温去季节项序列自相关图

Fig. 2 ACF of Seasonal Removed Temperature

4.2 时空协方差建模

分别对去季节项后的数据计算纯空间和纯时间变异函数,并进行拟合(如图 3 所示),由此确定其对应的标准化变异函数分别为:

$$\gamma_s(h_s; a) = 1 - \exp(-a(|h_s|/0.39)^2) \quad (9)$$

$$\gamma_t(h_t; a) = 1 - \exp(-ah_t/2.06) \quad (10)$$

$$\gamma_{s,t}(h_s, h_t) = 5 - 1.82 \left[\frac{1.5}{1.5 + \left(\frac{|h_s|}{0.39}\right)^2} \right]^3 - 0.48 \left[\frac{1.5}{1.5 + \frac{h_t}{2.06}} \right]^3 - 2.7 \left[\frac{1.5}{1.5 + \left(\frac{|h_s|}{0.39}\right)^2 + \frac{h_t}{2.06}} \right]^3 \quad (12)$$

4.3 插值结果分析

现用前文所述的变异函数模型对 2007 年 10 月的各个站点进行时空插值,插值过程采用交叉验证法进行。将插值结果加上之前去除的对应季节项得到最终估计值,并与原始观测值进行对比,如图 4(a)所示,其相关系数为 0.914,而图 4(b)为该月的纯空间插值效果,相关系数只有 0.903。此外,由于是在时空域中对变量进行变异函数建模和插值估计,所以可以非常方便地对研究范围内的任意时空点进行估计。图 5 为时空 Kriging 插值误差分布图,图中显示了 2007 年 4 月、7 月和 10 月三个时间点的所有站点的估计误差,其中,每个气泡表示该站点的估计误差,气泡越大,表示该点的误差越大。

图 5 中,绝大多数的气泡较小,大部分站点的月均气温插值误差的绝对值小于 0.7 °C,插值效

实际拟合的空间变异函数的块金值和基台值分别为 0.37、4.52,时间变异函数的块金值和基台值分别为 1.53、3.18,时空基台值取时空样本变异函数的最大值 5,代入式(5)~式(7)得: $k_1 = 2.7, k_2 = 1.82, k_3 = 0.48$ 。为了在积分式时空模型推导过程中消去参数 a ,经过参数的反复调整,取密度函数为:

$$\varphi(a) = \frac{\beta^{n+1}}{\Gamma(n+1)} a^n e^{-\beta a} = \frac{1.5^3}{2!} a^2 e^{-1.5a} \quad (11)$$

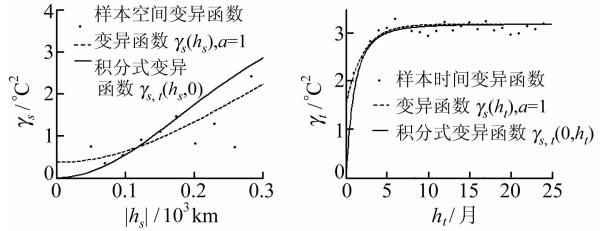


图 3 纯空间和纯时间变异函数及其拟合模型

Fig. 3 Spatial Variogram and Temporal Variogram and Their Fitted Models

式(11)中,参数 $n \geq 0, \beta > 0$,它们的值通过以下步骤确定^[13]:①将 $\varphi(a)$ 代入式(4)得到 $\gamma_{s,t}(h_s, 0), \gamma_{s,t}(0, h_t)$;②反复调整 n, β 的值,直到 $\gamma_{s,t}(h_s, 0), \gamma_{s,t}(0, h_t)$ 分别对样本空间变异函数和样本时间变异函数的拟合效果最佳。参数的确定可采用最小二乘法,也可根据经验观察函数拟合效果来确定。

将式(11)代入式(4)得:

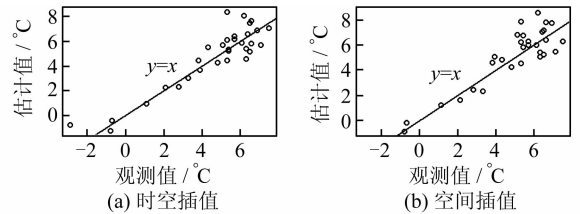


图 4 估计值与观测值对比散点图

Fig. 4 Scatter Plots of Estimate Value vs. Observed Value

果比较满意。而少部分误差较大的站点主要分布于区域的边界,如黑龙江省最北端的漠河站点、东部边界的鹤岗、富锦、宝清、虎林站点以及南部的牡丹江站点等。这些站点由于地处研究区域的边界,周围的邻近站点分布较少,缺乏可供插值的空间信息,致使插值误差相对于其他站点更大些。

时空插值由于同时考虑了空间和时间相关

metry[C]. Statistical Methods for Spatio-Temporal Systems, Chapman & Hall/CRC, Boca Raton, 2007

[17] 王燕. 应用时间序列分析 [M]. 2 版. 北京: 中国人民大学出版社, 2008

第一作者简介: 李莎, 讲师, 博士, 研究方向为地质统计、时空数据分析等。

E-mail: ls_li8202@126.com

Interpolation of Temperature Based on Spatial-temporal Kriging

LI Sha^{1,2} SHU Hong¹ XU Zhengquan¹

(1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 Department of Mechanical and Electrical Engineering, Hubei University of Education, 29 Guanggu 2nd Road, Wuhan 430205, China)

Abstract: A kind of integrated covariance (variogram) model is introduced for spatial-temporal Kriging interpolation of monthly average temperature of 37 years in Heilongjiang province. As monthly average temperature display obvious seasonal change, seasonal part is removed before interpolation. Spatial-temporal variogram is built based on the ones of pure space and pure time. Extending ordinary Kriging into space-time and considering variable's correlation both in space and time, the monthly average temperatures of all stations in 2007 are estimated, and the effect is compared with spatial Kriging's. The result shows that spatial-temporal interpolation is practical, and its accuracy is better than that of spatial Kriging.

Key words: Kriging interpolation; space-time correlation; covariance; variogram; temperature

About the first author: LI Sha, lecturer, Ph. D. majors in geostatistics and spatial-temporal data analysis.

E-mail: ls_li8202@126.com

(上接第 228 页)

Feature Matching from Network Data at Different Scales Based on Similarity Measure

AN Xiaoya¹ SUN Qun¹ YU Bohu²

(1 Institute of Surveying and Mapping, Information Engineering University, 66 Middle Longhai Road, Zhengzhou 450052, China)

(2 Northwest Nuclear Institute of Technology, Xi'an 710024, China)

Abstract: An algorithm for feature matching from network data at different map scaled based on similarity measure is presented. The whole strategy of matching is the first pre-matching of nodes and arcs, followed by accurate matching through similarity of node-arc topologies and discrete Fréchet distance. The matching process combines the matches in geometry, semantics, topology, nodes and arcs effectively. Finally, the different matching results are displayed to facilitate the human-computer interaction. The experimental results show that this method can match correspondent roads under complicated conditions effectively, and heighten the correctness and the speed of the feature matching.

Key words: data integration; similarity measure; map matching; Hausdorff distance; discrete Fréchet distance

About the first author: AN Xiaoya, Ph. D candidate, majors in geography data updating.

E-mail: xya2001@tom.com