

气温插值中不同空间插值方法的适用性分析 ——以江苏省为例

彭彬¹, 周艳莲^{1,2*}, 高苹³, 居为民²

(1. 南京大学地理与海洋科学学院, 南京 210093;

2. 南京大学国际地球系统科学研究所, 南京 210093; 3. 江苏省气象台, 南京 210008)

摘要: 气温是最重要的气象因子之一, 空间插值为台站气象数据降尺度提供了有效方法。本文利用江苏省 67 个气象台站 2003 年的逐日气温资料计算逐月平均气温和年平均气温, 结合空间分辨率为 30m×30m 的 DEM 数据, 分别利用反距离权重法、张力样条插值法、普通克里格插值法和协同克里格插值法, 对月 and 年平均气温进行插值, 并利用交叉验证法对插值精度进行了验证。研究表明, 考虑空间自相关性的普通克里格法的插值精度明显优于反距离权重法和张力样条插值法; 而协同克里格法由于考虑了经纬度、距海岸距离和高程等影响气温空间分布的局地因素, 其插值精度一般略优于普通克里格法, 尤其是在站点稀疏的情况下, 协同克里格的优势更加明显; 由于受海陆分布和东亚季风的影响, 江苏省气温的空间分布与距海岸距离有关, 在利用协同克里格进行空间插值时, 距海岸距离也是一个值得关注的因素。

关键词: 协同克里格; 普通克里格; 插值; 气温

DOI: 10.3724/SP.J.1047.2011.00539

1 引言

气温对生态系统的结构和功能具有非常重要的影响, 是生态模型的主要输入因子之一。尽管气象台站通过长期观测获得了大量的气温资料, 但這些测站的数据只能代表气象台站所在位置有限范围内的气温状况, 在进行区域/全球生态系统模拟时只能通过空间插值的方法生成气温空间分布场^[1], 而气温插值的误差必然会导致生态模型的不确定性。20 世纪 80 年代以来, 全球气候变暖得到广泛关注, 不同地区在不同时期对全球变暖的响应不一样^[2]。受其分辨率的限制 GCM 尚无法给出中小尺度区域内气候变化的细节, 因而, 利用观测资料对区域气候变化的若干细节进行研究就显得很有必要。江苏地处江淮流域, 地跨温带、亚热带, 尤受季风进退异常及年际气候变化的显著影响, 属于

气候脆弱带^[3], 因此, 对江苏省气温时空分布特征的研究具有十分重要的意义。

地理信息科学为气温数据栅格化提供了有效的方法。近年来, 许多学者围绕气象数据栅格化问题展开一系列研究。蔡福等分别利用反距离权重法(Inverse Distance Weighting, IDW)、普通克里格法(Ordinary Kriging, OK)、趋势面模拟等方法, 进行月 and 年平均气温的插值^[4], 谭继强等对比了反距离移动平均法、反距离移动表面法和普通克里格法 3 种不同的空间插值方法在气温插值中的精度, 并指出克里格插值最为有效^[5]。由于气温空间分布受高程影响较大, 这些单纯利用气温数据进行插值的方法不可避免地造成插值误差, 因此, DEM 逐渐被应用到气温数据插值中。蔡迪花等提出基于 DEM 的多元线性回归插值方法, 并通过对比指出其精度优于传统的插值方法^[6], 廖顺宝等指出在气

收稿日期: 2010-11-30; **修回日期:** 2011-05-23.

基金项目: 中国科学院资源与环境信息系统国家重点实验室开放基金(2009-2011); 江苏高校优势学科建设工程资助项目; 中央高校基本科研业务费专项资金。

作者简介: 彭彬(1989-), 男, 南京大学地理学国家理科基础科学研究与人才培养基地本科生, 目前主要从事定量遥感方面的学习。E-mail: 112233pengbin@163.com

*** 通讯作者:** 周艳莲(1980-), 女, 南京大学地理与海洋科学学院讲师, 博士, 主要从事陆地生态系统碳水循环研究。
E-mail: zhouyl@nju.edu.cn

温插值中 DEM 数据误差对栅格化结果会产生影响^[1],夏智宏等也利用 DEM 对气象资料进行垂直订正^[7]。但是,目前的研究多利用气温资料本身的信息进行插值,或仅考虑 DEM 对气温插值的影响,对其他辅助信息的利用较少,不可避免地会造成插值误差。协同克里格法(Co-Kriging,CK)既考虑变量的空间自相关性,又能充分利用其他辅助信息,可以提高气温插值的精度,杨凤海等将其用于气温数据插值中,取得了较好的效果^[8]。

空间插值方法虽很多,但插值方法在不同研究区域的适宜性分析不足,同一种插值方法在不同地区的应用效果存在很大的差别。由于受到海陆分布和东亚季风的影响,江苏省气温的空间分布与距海岸的距离有关,且季节性差异明显,如果能将距海岸距离等反映海洋对气温空间分布影响的信息应用到插值过程中,势必能够提高江苏省气温插值精度。

本文利用江苏省 2003 年的气温资料,对比了 IDW、样条插值法和 OK 插值法在年平均气温、1 月和 7 月平均气温中的精度;分析了逐月平均气温与经度、纬度、海拔高程、距海岸线最短距离 4 个辅助变量之间的相关性,及 4 个辅助变量之间的独立性,通过主成分变换提取出强相关因子作为辅助变量对逐月平均气温进行协同克里格插值,并评价协同克里格方法在气温插值中的能力,以期对气温资料的降尺度提供参考。

2 研究区与数据资料

2.1 研究区概况

江苏省位于长江、淮河下游,是长江三角洲地区的重要组成部分。其地理位置介于东经 116°18′~121°57′,北纬 30°45′~35°20′之间(图 1),是全国地势最低的一个省区,绝大部分地区在海拔 50m 以下,低山丘陵集中在北部和西南部,占全省总面积的 14.3%,省内最高峰海拔 624.4m。江苏省高程变化小,其对气温的空间分布影响小,省内气象要素的纬向分布特征明显,但由于受到海陆分布和东亚季风的影响,气象要素的空间分布与距海岸距离有关,且季节性差异明显,因此,采用合理的插值方法以反映这些因素对气温的影响,是提高气温插值精度的有效途径。

2.2 数据资料

本研究所采用的地理信息资料是由国家基础地理信息系统网站(<http://nfgis.nsd.gov.cn/>)提供的 1:400 万国界、省界和地区界限 shpfile 文件。所采用的气象资料来源于江苏省气象局,包括江苏省 2003 年 67 个气象台站(见图 1)的逐日气温数据,每个台站的数据记录含有站点编号、经纬度信息、逐日的平均气温、最高气温、最低气温等。采用的 DEM 数据是 ASTER GDEM 数据,由 ASTER 官方网站(<http://asterweb.jpl.nasa.gov/gdem.asp>)提供,空间分辨率是 30m×30m,对地理信息文件添加 Albers 等面积投影,中央经度为 E110°,标准纬线为 N25°和 N47°,并在 ArcGIS 9.2 中从全国的省界和地区界文件中裁剪出江苏省部分(见图 1);将获得的 DEM 数据通过数据拼接、投影转换和数据裁剪等处理,提取出江苏省的 DEM 数据,并利用 Arc GIS 9.2 中的“Extract To Points”工具获得气象台站的高程。计算各个气象台站 2003 年的逐月平均气温和年平均气温,并利用 Arc GIS 9.2 中的地统计学模块对气温数据进行空间探索性分析^[9]。其结果显示数据质量良好,各月平均气温近似呈正态分布,且没有空间变异异常点。

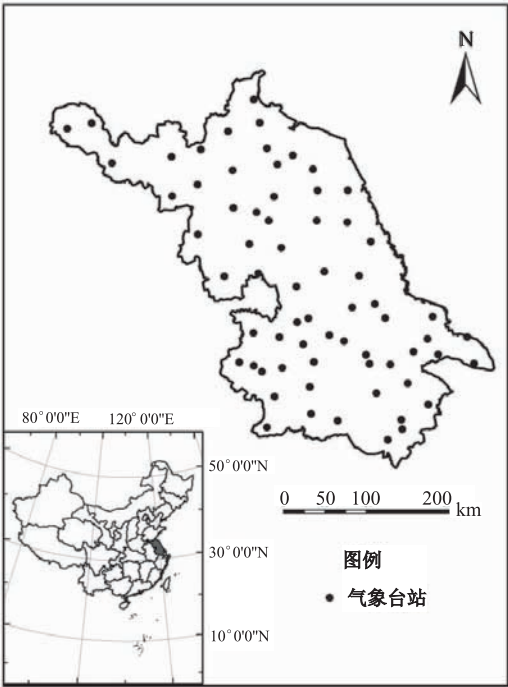


图 1 研究区域和气象站点分布图
Fig. 1 Map of the study area and the distribution of the meteorological stations

3 空间插值与精度验证方法及其适用性分析

空间插值的理论依据是 Tobler 地理学定律,即空间上距离越小的地理数据比那些空间距离更大的数据有更大的相似性^[10-11]。目前,常用的插值方法有最邻近法、算术平均法、反距离权重法、多项式插值法(包括局部多项式插值和全局多项式插值)、样条插值法、高次曲面插值、克里格插值系列方法,以及径向基函数插值法等。其中,反距离权重法、样条插值法和克里格插值系列方法属于精确插值^[12](见表 1)。

表 1 常用插值方法的主要特征比较
Tab. 1 Major characteristics of the commonly used interpolation methods

插值方法	整体拟合	局部拟合	确定性	随机性	精确插值	非精确插值
趋势面	√		√			√
多元回归	√			√		√
密度估算		√	√			√
IDW		√	√		√	
Spline		√	√		√	
Kriging		√		√	√	

由于算术平均法、最邻近法、多项式插值和高次曲面插值等方法没有考虑变量的空间自相关性,本文不予采用。而样条插值是径向基函数插值方法的一个特例。故下文主要分析考虑了空间自相关性并且为精确插值的反距离权重法、样条插值法和克里格插值法的精度,并进一步评价协同克里格插值方法在气温资料降尺度中的效果。

3.1 空间插值方法及应用分析

(1)反距离权重法(IDW)

反距离权重插值法是距离权重系数方法系列中最具代表性的方法,该方法对待估值点周围的观测点进行线性加权获得估计点的值,权重系数与距离成反比,如果用 v_e 表示待估计点的值,则有:

$$v_e = \sum_{j=1}^n w_j v_j \tag{1}$$

其中, $v_j (j = 1, 2 \cdots, n)$ 为待估值点周围的点 (x_j, y_j) 处的变量值; w_j 是其对应的权重系数,可用如下公式计算:

$$w_j = \frac{f(d_{ej})}{\sum_{j=1}^n f(d_{ej})} \tag{2}$$

其中, n 为参与插值的临近点数, $f(d_{ej})$ 是待插值点 (x_e, y_e) 与邻近点 (x_j, y_j) 之间的距离 d_{ej} 的权重函数,最常用的为距离倒数函数或距离倒数平方函数^[11]。

(2)样条插值法(Spline)

样条插值法从根本上属于多项式插值,该方法主要是通过估计方差,利用一些特征节点,用多项式拟合的方法来产生平滑的插值曲线。这种方法适用于逐渐变化的表面,如温度、高程、地下水位高度或污染浓度等。如果用 v_e 表示待估计点的值,则有:

$$v_e = \sum_{j=1}^n A_j d_{ej}^2 \log d_{ej} + ax_e + by_e + c \tag{3}$$

其中, n 为参与插值的临近点数, d_{ej} 为待插值点 (x_e, y_e) 与邻近点 (x_j, y_j) 之间的距离, A_j, a, b, c 都是相应的参数^[13-14]。

样条插值法又分为张力样条插值法 (Spline With Tension) 和规则样条插值法 (Regularized Spline)。为了避免产生极值现象,本文选择张力样条插值法。

(3)普通克里格法(OK)

克里格系列插值法包括简单克里格法、普通克里格法、协同克里格法、指示克里格法、因子克里格法、泛克里格法,以及块克里格法等,其以区域化变量理论和空间自相关理论为基础,对满足二阶平稳或固有假设的变量具有很好的估计精度,是一种最优线性无偏的估计方法 (Best Linear Unbiased Estimator, BLUE)^[11, 15-17]。使用普通克里格插值方法,需先计算样本变异函数,再根据样本变异函数的类型选择合适的变异函数理论模型进行模拟,最后根据模拟的变异函数对待估计点进行线性估计并给出估计方差作为不确定性的度量指标。如果用 v_e 表示待估计点的值,则有:

$$v_e = \sum_{j=1}^n r_j v_j \tag{4}$$

其中, $v_j (j = 1, 2 \cdots, n)$ 为待估值点周围的点 (x_j, y_j) 处的变量值; r_j 是其对应的权重系数,该系数是通过样本变异函数计算值进行模拟得到的,样本变异函数值可以通过下式来计算:

$$r^*(h) = \frac{1}{2N_h} \sum_{i=1}^{N_h} [v(x_i + h) - v(x_i)]^2 \tag{5}$$

其中, h 为分离距离, N_h 是在 $(x_i + h, x_i)$ 之间用来计算样本变异函数值的样本点对数。变异函数的理论模型有球状模型、指数模型、高斯模型、幂函数模型和 DeWijisian 模型等, 选用何种模型进行变异函数的模拟由样本变异函数的计算值分布类型决定。

(4) 协同克里格法 (CK)

协同克里格法是一种多变量估值方法, 其主要思想是既利用待估计变量的空间自相关性, 又利用不同变量之间的相关性, 这种方法可以解决现实中待估计变量的观测值太少的问题, 从而能提高估值的精度^[15, 18-20]。运用协同克里格方法的关键是计算变量之间的协变异函数, 如果用 $v_1(x)$, $v_2(x)$, $\cdots v_m(x)$ 分别表示变量 $v_1, v_2, \cdots v_m$ 在位置 x 处的值, 即 $V(x) = [v_1(x), v_2(x), \cdots v_m(x)]$, 如果 $x_1, x_2, \cdots x_n$ 是样本点的位置并且其测量值为 $V(x_1), V(x_2), \cdots V(x_n)$, 那么协同克里格估值方法可以表示为^[15]:

$$V_e(x) = \sum_{i=1}^n V(x_i) \Gamma_i \tag{6}$$

其中, Γ_i 是权重向量, 由协变异函数给出, 协变异函数的计算公式如下:

$$\begin{aligned} \gamma_{12}(h) = E\{[v_1(x+h) - v_1(x)] \\ [v_2(x+h) - v_2(x)]\} \end{aligned} \tag{7}$$

3.2 精度对比——交叉验证法

交叉验证法是一种常用的精度验证方法, 该方法假设参与插值过程的某一个观测值为未知值, 让余下的数据点全部参与插值而得到该假设点的预测值, 将预测值与测量值对比就可以评价插值方法的优劣程度。该方法的优点在于最大限度地利用了观测值, 而避免了因预留用于验证精度的观测值减少了参与插值过程的观测值数量而对总体精度的负面影响。交叉验证法常用的用于评价精度的参数是平均绝对误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Square Error, RMSE), 其计算公式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n abs(T_{oi} - T_{ei}) \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (T_{oi} - T_{ei})^2} \tag{9}$$

式中, n 为测站数, T_{oi} 和 T_{ei} 分别表示第 i 个台站位置上的实际观气温和预测气温值。

4 不同插值方法的精度对比

4.1 IDW、Spline 和 OK 的插值精度分析

选择 2003 年 1 月、7 月和年平均气温, 分别使用 IDW、Spline 和 OK 3 种方法进行插值, 并利用交叉验证法对比不同插值方法的精度, 结果见表 2。表中显示, 对于 1 月平均气温, 3 种插值方法得到的 MAE 分别为 0.23℃、0.22℃和 0.21℃, 可见 OK 插值方法明显优于其他两种方法; 对于 7 月平均气温, 3 种插值方法得到的 MAE 分别为 0.19℃、0.18℃和 0.18℃, OK 插值方法同样优于其他两种方法; 对于 1 月和 7 月平均气温, 3 种插值方法得到的气温的 RMSE 结果也明显表明 OK 优于其他两种插值方法; 年平均气温资料插值的误差较小, 3 种方法的均方根误差都约为 0.23℃, 差异不十分明显。表中还显示, IDW 比 Spline 和 OK 的插值精度低许多, 除年平均气温外, 其误差相差都在 0.01℃

表 2 IDW、Spline 和 OK 三种插值方法的精度 (℃)
Tab. 2 Precision of IDW, Spline and OK methods (℃)

	1 月平均气温		7 月平均气温		年平均气温	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
IDW	0.23	0.28	0.19	0.24	0.18	0.23
Spline	0.22	0.28	0.18	0.23	0.18	0.23
OK	0.21	0.27	0.18	0.21	0.18	0.23

表 3 平均气温与经度、纬度、距海岸距离和高程之间的相关系数

Tab. 3 Correlation coefficients of mean air temperature with latitude, longitude, distance from the coast and elevation

月份	纬度	经度	距海岸距离	高程
1 月	-0.97 ^a	0.59 ^a	-0.03	-0.13
2 月	-0.97 ^a	0.54 ^a	0.03	-0.10
3 月	-0.87 ^a	0.20	0.37 ^a	0.11
4 月	-0.62 ^a	-0.10	0.48 ^a	0.26 ^b
5 月	-0.51 ^a	-0.33 ^a	0.71 ^a	0.38 ^a
6 月	-0.42 ^a	-0.39 ^a	0.74 ^a	0.41 ^a
7 月	-0.94 ^a	0.37 ^a	0.20	0.05
8 月	-0.96 ^a	0.60 ^a	-0.06	-0.14
9 月	-0.96 ^a	0.65 ^a	-0.09	-0.21
10 月	-0.88 ^a	0.73 ^a	-0.29 ^b	-0.30 ^b
11 月	-0.92 ^a	0.82 ^a	-0.33 ^a	-0.34 ^a
12 月	-0.96 ^a	0.70 ^a	-0.16	-0.25 ^b
年平均	-0.95 ^a	0.46 ^a	0.10	-0.06

a: 置信度为 0.01; b: 置信度为 0.05;

之上,而 OK 虽比 Spline 的精度高,但就 1 月、7 月和年平均气温资料而言插值精度的差距不如 IDW 和 OK 与 Spline 之间的大。通过以上分析可知,针对 1 月平均气温、7 月平均气温和年平均气温 3 种情况,无论是 MAE 判据还是 RMSE 都显示出相同的规律,即 OK 优于 Spline,而 Spline 又优于 IDW。

4.2 CK 与 OK 的精度比较

4.2.1 辅助变量的选择

(1)气温与经度、纬度、距海岸距离和高程之间的相关性

运用协同克里格法的前提是选择与目标变量具有相关性的协变量。纬度和高程对气温的空间分布有直接影响,由于江苏省特殊的地理位置,受到海陆分布和东亚季风的影响,气温的空间分布还与距海岸的距离有关,在本研究中,选择了纬度、经度、距海岸线距离和高程共 4 个变量,分别计算它们与各月平均气温的相关系数。由表 3 可见,与平均气温相关性最好的是纬度,除个别月份外相关系数均能达到 0.85 以上,且所有月份的相关系数都通过了置信度 0.01 的检验;相关性次好的是经度,除 3 月和 4 月外,其余月份的相关系数都能通过置信度为 0.01 的检验;3、4、5、6、10 和 11 月的平均气温与距海岸距离呈现出相关性,除 10 月份以外,二

者之间的相关系数都通过了置信度为 0.01 的检验;高程与平均气温之间的相关性较小,除个别月份能通过相关性检验之外,其余月份没有明显的相关性,这主要是因为江苏省的高程变化小,其他因子对气温分布的影响大于高程的作用。

(2)经度、纬度、距海岸距离和高程之间的相互独立性评价

为了减小协同克里格插值结果的不确定性,需要辅助变量之间相互独立。表 4 给出了 4 个变量之间的相关系数,可以发现经度与纬度、距海岸距离、高程之间都有很强的相关性,均通过了置信度为 0.01 的相关性检验。经度与纬度之间的相关性是由于江苏省呈西北-东南走向的地理分布形态决定的,纬度越大的地区具有相对较小的经度(见图 1)。另外,江苏省沿海岸线分布和江苏省地势总体上东低西高的特征导致了经度与距海岸距离和高程之间都具有明显的相关性;由计算结果还可以看出,纬度只与经度具有一定的相关性,与其他变量均没有相关性。

(3)主成份分析

变量之间相互关联,如果直接作为辅助变量进行协同克里格插值,数据的冗余势必给插值结果带来误差,本文进一步利用经典的统计学软件 SPSS,对纬度、经度、距海岸距离和高程 4 个变量,进行主成份分析(Principal Component Analysis, PCA),变换后的主成份对方差的累积贡献率分别是 59.95%、87.29%、99.43%和 100%,根据累积贡献率达到 85%的原则,前两个主成份即可以基本反映全部的信息。各个变量与 4 个主成份之间的相关系数见表 5,从表中可知第一主成份主要由经度和距海岸距离控制,第二主成份主要反映纬度的信息,而第三主成份主要由距海岸距离和高程联合控制。

(4)气温与各主成份分量的相关性评价

4 个主成份分量与逐月平均气温的相关系数如表 6 所示。在 4 个主成份分量中与各月平均气温相关性最好的是第二主成份,这与上文中平均气温与纬度相关性最好和第二主成份分量主要由纬度控制的结论相一致。除 3 月和 4 月份外,第一、二主成份与各月平均气温的相关系数都通过了置信度为 0.01 或者 0.05 的检验,且第一主成份与各月平均气温主要呈负相关,而第二主成份与各月平均气温都呈正相关。在整年中,3、4、5 和 6 月的气温

表 4 各变量之间的相关系数矩阵
Tab. 4 The correlation coefficient matrix among the four auxiliary variables

	纬度	经度	距海岸距离	高程
纬度	1			
经度	-0.60 ^a	1		
距海岸距离	-0.01	-0.77 ^a	1	
高程	0.11	-0.56 ^a	0.57 ^a	1

a:置信度为 0.01;

表 5 各变量与各主成份分量之间的相关系数
Tab. 5 The correlation coefficients between principal components and auxiliary variables

	PC1	PC2	PC3	PC4
纬度	0.47 ^a	-0.88 ^a	0.08	0.07
经度	-0.96 ^a	0.20	0.18	0.11
距海岸距离	0.83 ^a	0.42 ^a	-0.35 ^a	0.08
高程	0.76 ^a	0.32 ^a	0.57 ^a	0.01

a:置信度为 0.01。

变异规律与其他月份不一致,主要表现在这 4 个月的平均气温与第一主成份呈正相关或者不相关,而与第三和第四主成份的相关性在全年中最大,该现象与表 3 中该 4 个月平均气温与距海岸距离或者高程呈现出明显的相关性是一致的。

通过以上分析,在下文的协同克里格插值的过程中,以选择与逐月平均气温相关性较好的主成份分量作为最终参与插值的辅助变量为原则,故 3 月仅选择 PC2 作为辅助变量,4 月和 7 月选择 PC2 和 PC4 作为辅助变量,其他月份均选择 PC1 和 PC2 作为辅助变量。

表 6 平均气温与各主成份分量之间的相关系数
Tab. 6 The correlation coefficients between principal components and mean air temperatures

	PC1	PC2	PC3	PC4
1 月	-0.47 ^a	0.84 ^a	-0.07	-0.14
2 月	-0.42 ^a	0.86 ^a	-0.10	-0.15
3 月	-0.09	0.91 ^a	-0.20	-0.21
4 月	0.17	0.74 ^a	-0.17	-0.47 ^a
5 月	0.40 ^a	0.73 ^a	-0.28 ^b	-0.36 ^a
6 月	0.46 ^a	0.67 ^a	-0.28 ^b	-0.29 ^b
7 月	-0.25 ^b	0.91 ^a	-0.10	-0.25 ^b
8 月	-0.49 ^a	0.82 ^a	-0.05	-0.18
9 月	-0.54 ^a	0.79 ^a	-0.09	-0.10
10 月	-0.66 ^a	0.64 ^a	-0.02	-0.20
11 月	-0.72 ^a	0.66 ^a	-0.01	-0.03
12 月	-0.60 ^a	0.76 ^a	-0.07	-0.11
年平均	-0.35 ^a	0.87 ^a	-0.12	-0.21

a:置信度为 0.01; b:置信度为 0.05;

4.2.2 CK 和 OK 的精度对比

相关研究表明,在具体的插值过程中如果对去除趋势面之后的残差进行对应的插值可以提高插值精度^[21],这主要是由于克里格插值方法要求变量符合二阶平稳或者固有假定引起的,去除趋势(Trend)或者漂移(Drift)后能更好地满足这一前提。因此,本文在对江苏省 2003 年逐月和年平均气温资料进行协同克里格插值时,去除了气温资料的一阶趋势面。协同克里格法和普通克里格法的交叉验证如表 7 所示。结果表明,除 6 月份协同克里格插值法的 MAE 稍大于普通克里格插值的 MAE 外,其余各月的 MAE 和 RMSE 均表现出 CK 稍小于或等于 OK 的特征,即协同克里格法略优于普通克里格插值法。6 月的异常可能是由于辅助变

量选择所导致的误差,即表 6 中显示 6 月的平均气温与 4 个主成份分量的相关性差别不大,而在协同克里格插值的过程中只选择了相关性最大的两个主成份分量,这一过程可能造成有效信息的丢失从而引起插值误差。

表 7 还显示,相比于普通克里格法,协同克里格法的 MAE 最大可以减小 5.6%,而 RMSE 最大可以提高 4.8%。春季(3、4 和 5 月)的插值误差随着气温的回暖而逐渐增加,冬季(12、1 和 2 月)的插值误差随着气温的降低而逐渐减小,夏季(6、7 和 8 月)插值的误差变化最为剧烈,而秋季(9、10 和 11 月)插值的误差变化范围相对最小。

表 7 协同克里格法和普通克里格法的精度对比(℃)

	CK		OK	
	MAE	RMSE	MAE	RMSE
1 月	0.19	0.26	0.20	0.26
2 月	0.15	0.19	0.15	0.19
3 月	0.18	0.23	0.19	0.23
4 月	0.19	0.23	0.20	0.24
5 月	0.22	0.28	0.22	0.28
6 月	0.27	0.32	0.26	0.32
7 月	0.17	0.20	0.18	0.21
8 月	0.20	0.25	0.21	0.26
9 月	0.25	0.31	0.25	0.31
10 月	0.26	0.32	0.27	0.32
11 月	0.22	0.27	0.23	0.27
12 月	0.22	0.28	0.22	0.28
年平均	0.18	0.23	0.19	0.23

为进一步评价 CK 和 OK 两种插值方法在空间上的差异,生成了 1 月、7 月和年平均气温插值时标准误差的空间分布图(见图 2)。图中显示江苏省范围内 1 月、7 月和年平均气温插值时 CK 误差的最小值均低于 OK 误差的最小值;从等值线的分布情况可以看出,插值误差较小的范围在 1 月、7 月和年平均气温中均是 CK 大于 OK,这在 7 月份更加明显。此外,图 2 中还显示 CK 会在局部地区产生“低误差岛”,相比而言,OK 误差的分布则没有显示出这样的细微特征,这表明 CK 能够产生更加真实的气温分布场。通过以上分析,无论交叉验证结果还是误差空间分布情况都显示在气温插值中 CK 优于

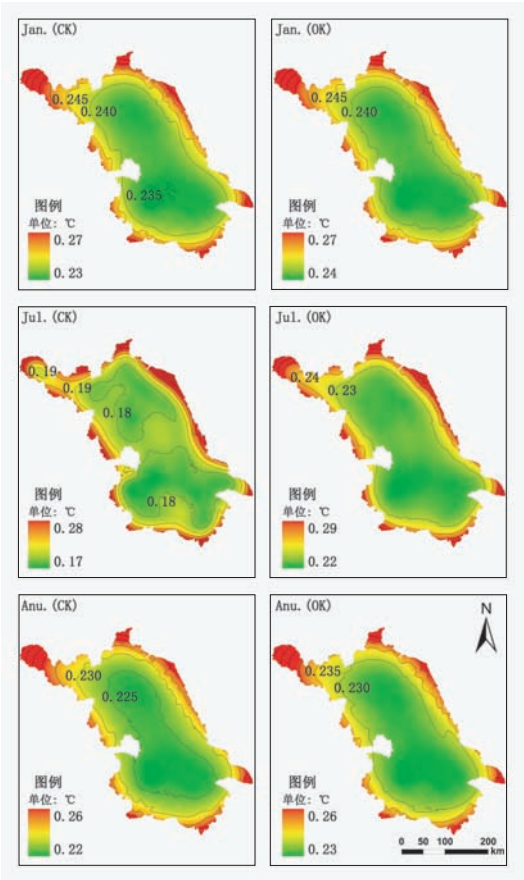


图 2 CK 和 OK 插值标准误差空间分布图
Fig. 2 Maps of spatial distribution of standard errors for CK and OK

OK。值得注意的是,图 2 中显示在江苏省边界地区的插值误差明显高于其他地区,这是由于在插值时只利用了江苏省的气象站资料,但这并不影响对结果的分析。

4.3 月平均气温的协同克里格插值结果

利用协同克里格插值法对江苏省 2003 年逐月的平均气温和年平均气温资料进行栅格化,如图 3 所示。结果表明,江苏省 2003 年年平均气温为 13.40~17.30℃,与 4 月和 10 月的平均气温相当;全年最高气温和最低气温分别出现在 1 月和 7 月,最低温度为-0.62℃,最高温度为 30.32℃;月平均气温的南北差异冬季大于夏季。图 3(a)和(b)分别为江苏省冬季和春季、夏季与秋季的气温栅格化结果,图中显示冬季和秋季的气温的等值线呈西南—东北走向,这与江苏省冬季和秋季的气温分布模式相符。其反映了秋冬季“热海洋”对江苏省气温空间分布模式的影响;与秋冬季相反,春季和夏季的气温等值线呈西北—东南走向,这也与实际的气温分布模式相同,反映了春夏季“冷海洋”对江苏省气温空间分布模式的影响。另外,图 3 中还显示,4 月、5 月和 6 月的等值线倾斜程度在全年中最大,表示气温与距海岸距离的相关性最大,这与表 3 中气温与距海岸距离的相关系数在这三月中最大是一致的。

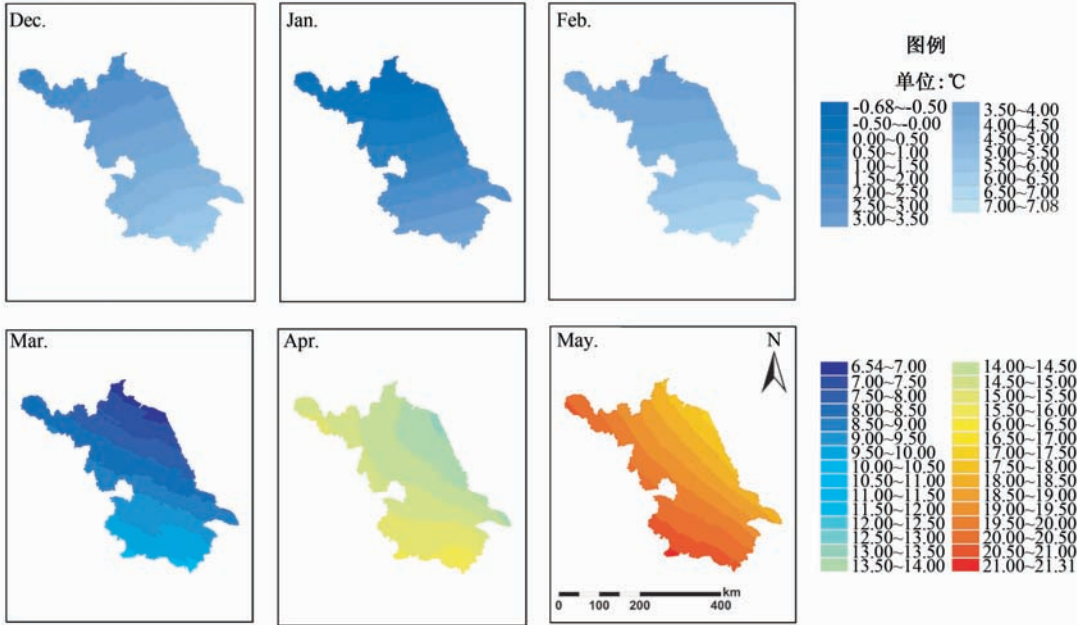


图 3(a) 2003 年江苏省月平均气温的协同克里格插值结果
Fig. 3(a) The Co-Kring interpolation result of monthly mean air temperatures in Jiangsu, winter and spring, 2003

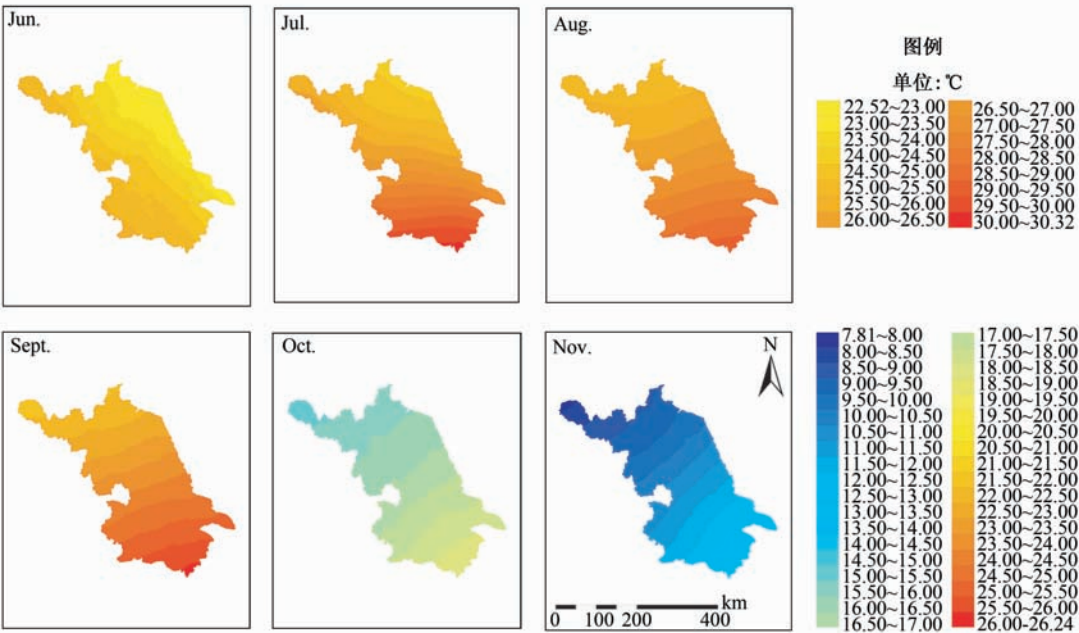


图 3(b) 2003 年江苏省月平均气温的协同克里格插值结果

Fig. 3(b) The Co-Kring interpolation result of monthly mean air temperatures in Jiangsu, summer and autumn, 2003

5 讨论和结论

5.1 讨论

本文运用协同克里格插值后精度提高的程度不大,最多也只能提高 5%左右,这是由于协同克里格是在普通克里格的基础上为解决不完全采样的问题(如某些台站的数据难以获取或存在异常时)而提出的一种插值方法,对于完全采样问题(如每一个气象台站均存在有效记录时),协同克里格与普通克里格的插值精度差异不明显。当采样不完全时,协同克里格方法能够引入辅助变量,从而具有较明显的优势。在江苏省的 67 个气象台站中随机选择 10、20、30、40、50 和 60 个台站作为实验样本参与插值,剩余的台站用于检验精度,以进一步分析随着气象台站密度的变化,普通和协同克里格两种插值方法的精度变化规律。结果表明(图 4),在台站数量较少时(如 10 个),无论是针对 1 月和 7 月平均气温还是年平均气温,协同克里格插值的精度明显较普通克里格高,但随着台站密度的增大,两种方法之间的精度差异均逐渐减小。由此可见,在对气象台站密度较低,存在较大观测盲区地区的气温进行栅格化时,协同克里格法的优势将更加明显。图 4 还表明,随着气象台站的加密,江苏省气温的插值精度在 10-20 个台站之间明显降低,但随

着台站密度进一步增大,精度并没有对应地提高,而是在 0.2~0.3 $^{\circ}\text{C}$ 之间保持相对稳定。

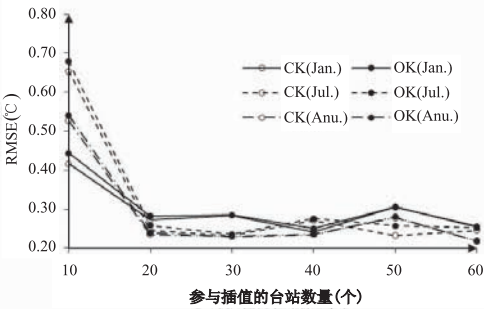


图 4 插值精度随台站密度的变化情况

Fig. 4 The variation of interpolation precision along with the change of station density

5.2 结论

- (1)气温数据插值过程中,普通克里格法优于反距离权重法和张力样条插值法,即考虑了空间自相关性的克里格方法比一般的插值方法具有更高的精度。
- (2)协同克里格法充分利用经纬度、距海岸距离、高程等变量对温度的影响,插值精度稍高于普通克里格法,尤其在站点稀疏的情况下,协同克里格法的优势更加明显。

(3)由于受到海陆分布和东亚季风的影响,江苏省气象要素的空间分布与距海岸的距离有关,在利用协同克里格进行空间插值时,距海岸距离也是一个值得关注的因素。

致谢 江苏省气象局、国家基础地理信息系统(NFGIS)、日本经济贸易产业部(METI)和美国国家航空航天管理局(NASA)为本文的研究提供了数据支持;南京大学国际地球系统科学研究所的田庆久教授和匿名审稿专家提出了修改意见,在此表示诚挚的感谢。

参考文献:

- [1] 廖顺宝,李泽辉. 气温数据栅格化中的几个具体问题[J]. 气象科技, 2004, 32(5): 352 - 356.
- [2] 白爱娟,翟盘茂. 中国近百年气候变化的自然原因探讨[J]. 气象科学, 2007, 27(5): 584 - 589.
- [3] 张燕,张秀丽,唐红昇,等. 江苏夏季气温异常的时空变化特征[J]. 气象科学, 2009, 29(1): 133 - 137.
- [4] 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.
- [5] 谭继强,丁明柱. 空间数据插值方法的评价[J]. 测绘与空间地理信息, 2004, 27(4): 11 - 13.
- [6] 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.
- [7] 夏智宏,刘可群,梁益同,等. 基于 Arc Engine 和 DEM 辅助的气温网格化的设计与实现[J]. 测绘与空间地理信息, 2008, 31(6): 86 - 88.
- [8] 杨凤海,孙彦坤,于太义,等. 近 10 年黑龙江省气温的时空变异分析[J]. 地球信息科学学报, 2009, 11(5): 585 - 596.
- [9] 汤国安,杨昕. ArcGIS 地理信息系统空间分析实验教程[M]. 北京: 科学出版社, 2006, 372 - 388.
- [10] 朱会义,刘述林,贾绍凤. 自然地理要素空间插值的几个问题[J]. 地理研究, 2004, 23(4): 425 - 432.
- [11] 王劲峰,等. 空间分析[M]. 北京: 科学出版社, 2006, 17 - 29.
- [12] Chang Kang-tsung. Introduction to Geographic Information Systems (third edition, Chinese translation by Chen Jianfei *et al.*) [M]. Beijing: Tsinghua University Press, 2009, 310 - 312.
- [13] Johnston K, ver Hoef J M, Krivoruchko K, *et al.* Using Arc GIS Spatial Analyst[M]. CA, USA: Environmental Systems Research Institute Inc., 2001, 136 - 148.
- [14] 李军,游松财,黄敬峰. 中国 1961 - 2000 年月平均气温空间插值方法和空间分布[J]. 生态环境, 2006, 15(1): 109 - 114.
- [15] 张仁铎. 空间变异理论及其应用[M]. 北京: 科学出版社, 2005, 13 - 88.
- [16] 刘贤赵,张安定,李嘉竹. 地理学数学方法[M]. 北京: 科学出版社, 2009, 127 - 138.
- [17] 徐建华. 计量地理学[M]. 北京: 高等教育出版社, 2006, 131 - 151.
- [18] Myers D E. Matrix Formulation of Co - Kriging[J]. Mathematical Geology, 1982, 14(3): 249 - 257.
- [19] Royle J A, Berliner L M. A Hierarchical Approach to Multivariate Spatial Modeling and Prediction[J]. Journal of Agricultural, Biological, and Environmental Statistics, 1999, 4(1): 29 - 56.
- [20] ver Hoef J M, Cressie N. Multivariate Spatial Prediction[J]. Mathematical Geology, 1993, 25(2): 219 - 240.
- [21] 关宏强,蔡福,王阳,等. 短时间序列气温要素空间插值方法精度的比较研究[J]. 气象与环境学报, 2007, 23(5): 13 - 16.

Suitability Assessment of Different Interpolation Methods in the Gridding Process of Station Collected Air Temperature: A Case Study in Jiangsu Province, China

PENG Bin¹, ZHOU Yanlian^{1, 2}, GAO Ping³, JU Weimin²

(1. School of Geographic and Oceanographic Sciences in Nanjing University, Nanjing 210093, China;

2. The International Institute of Earth System Science in Nanjing University, Nanjing 210093, China;

3. Jiangsu Meteorological Bureau, Nanjing 210008, China)

Abstract: Air temperature is an important parameter observed in metrological stations, and there are many ways to improve the precision of air temperature interpolation result. In this paper by using the air tem-

perature data at 67 meteorological stations in Jiangsu Province in the year of 2003 and digital elevation model (DEM) data with spatial resolution of $30\text{m} \times 30\text{m}$, four common interpolation methods, including Inverse Distance Weighting (IDW), Spline with tension (Spline), Ordinary Kriging (OK) and Co-Kriging (CK), were used to interpolate the monthly and yearly mean air temperature and the precision of those four methods was compared by using cross validation method. The results showed that OK has a much higher precision than IDW and Spline, indicating the method accounting for spatial self-correlation is more accurate than others. Four auxiliary variables, including latitude, longitude, distance from the coast and elevation, were selected for CK, and correlation analysis showed that the monthly mean air temperatures are best correlated with latitude, and the three other variables followed. As the four variables are correlated with each other, principal component analysis (PCA) was conducted in this paper. The first principal component mainly representing longitude and distance from the coast and the second one mainly representing latitude were utilized as the optimized auxiliary variables for Co-Kriging interpolation in most months except March whose input is only the second one, April and July whose inputs are the second and the fourth principal components. The results indicated that the precision of CK which makes good use of related auxiliary factors is slightly higher than that of OK; while it is obviously better than OK where there are fewer stations and is a potential ideal method for air temperature interpolation. The results of this paper also showed that distance from the coast is a critical factor to the spatial pattern of air temperature in Jiangsu, China, which should be an auxiliary variable for CK.

Key words: Co-Kriging; Ordinary Kriging; gridding process; air temperature