

气温数据栅格化中的几个具体问题

廖顺宝 李泽辉

(中国科学院地理科学与资源研究所,北京 100101)

摘要 用地理信息系统软件 ARC/INFO 及 4 种空间插值方法,对中国 617 个气象站 30 年平均气温数据和 1961 年平均气温数据的栅格化试验发现,克立格插值方法的精度最高,反距离权重法次之,样条插值法的精度第三,趋势面插值方法的精度最低。利用多年平均气温数据和年平均气温的距平值进行气温数据的栅格化,虽然可以减少分析和计算量,但栅格化结果的精度比利用年平均气温数据直接进行栅格化的精度要低。气象站的实际高程与气象站经纬度对应的数字高程(DEM)上的高程不完全一致,但气温数据的栅格化完全依赖于 DEM,用气象站实际高程建模进行气温数据栅格化的精度比用与气象站经纬度对应的 DEM 上的高程建模进行气温数据栅格化的精度要高。

关键词 气温 栅格化 内插方法 数字高程 距平值

引言

气温是重要的气候资源,气温数据主要是通过分布在全国范围内的气象观测站的长期连续观测获得。然而,气象站的观测气温值只能代表气象站所在地区(面积非常有限)的气温状况,对于广大非气象站所在地区的气温,只能通过推算求得。

气温数据的栅格化^[1~4]是推算非气象站所在地区气温的有效手段之一。气温数据的栅格化涉及空间插值、数字高程模型的应用以及回归分析^[5~6]等一些最基本的问题。空间插值有不同的方法^[7],哪种插值方法更加适合气温数据的栅格化?数字高程模型上的高程值往往与气象站的实际海拔高度不一致,而气温数据的栅格化又必须依赖于 DEM (Digital Elevation Model),在栅格化分析建模过程中究竟使用哪种高程值更加准确可靠?另外,是否能用多年平均气温数据和(某年的)气温距平值进行年平均气温数据的栅格化?本文将对上述问题进行分析、试验和探讨。

1 数据资料及预处理

1.1 数据资料

(1) 中国 30 年(1951~1980 年)的年平均气温数据,共有 624 个台站的数据。

(2) 中国 1961 年年平均气温数据,共有 650 个气象站的数据记录,每条记录包括气象站的名称、编号、经度、纬度、海拔高度和年平均气温等数据项。

(3) 中国 30'' 分辨率数字高程数据(经纬度坐标)。

1.2 数据预处理

(1) 气象站资料与 DEM 数据复合。由于数据空间坐标的误差,在 ARC/INFO 中用气象站的经纬度数据与 DEM 数据复合时,有 7 个气象站点落在了 DEM 的无值区域(主要是几个位于海岛上的气象站)。本文主要进行气温数据栅格化中的基本问题探讨,因此,在进行数据分析时去掉了这 7 个气象站。

(2) 气温资料的一致性处理。根据气象站编码一一对应的原则,从 30 年平均气温数据和 1961 年年平均气温数据中选择出在二者中均有记录的数据,共 599 条记录。

国家科技基础性工作专项资金课题(2001 DE A30027 - 9)和中国科学院知识创新工程项目(INF105 - SDB - 1 - 18)及中日合作课题“全球变化对中国的影响研究”(AIM/Impact)共同资助

作者简介:廖顺宝,男,1966 生,博士,副研究员,硕士生导师,主要从事遥感和地理信息系统应用以及属性数据空间化等方面的研究,

E-mail: liaosb @igsnr.ac.cn

收稿日期:2003 年 8 月 1 日;定稿日期:2003 年 9 月 30 日

(3) 选择验证气象站。根据均匀分布的原则从全国 617 个气象站中选择 57 个气象站用于气温数据栅格化结果的验证,它们在 1951 ~ 1980 的 30 年平均气温数据和 1961 年的年平均气温数据中均有记录。用于检验的 57 个气象站是:海拉尔市、北安市、齐齐哈尔、依兰、吉木乃、精河、巴仑台、柯坪、民丰、巴里坤哈、安西、山丹、格尔木市、兴海、榆中、苏尼特左、百灵庙、盐池、太原市、安阳市、通榆、林西、九站、宝国图、沈阳市、通化市、张家口市、营口市、唐山市、沂源、拉萨市、玉树、红原、雅安市、德荣、华坪、盘县、临沧、西安市、宝丰、巴东、南充、天门市、桐梓、铜仁、宁冈、桂林市、亳州市、南京市、慈溪市、景德镇市、龙泉、龙岩市、连平、百色市、罗定和琼中。

2 几个具体问题分析

2.1 几种常用空间插值方法的比较

ARC/INFO 是最专业的地理信息系统软件之一,它提供的常用的空间插值方法包括反距离权重(IDW)插值、克里格(Kriging)插值、样条(Spline)插值和趋势面(Trend)插值等。对于气温数据的栅格化,究竟哪种插值方法最好?下面以全国 1951 ~ 1980 年间的 30 年多年平均气温数据为例进行试验。共有 617 个气象站的数据,用其中 560 个站的数据进行气温数据栅格化的建模和计算,用其余 57 个站的数据对栅格化结果进行检验,从而判断哪种插值方法最好。

关于气温数据栅格化的方法,相关文献^[4]已有研究,其结果是“多元回归 + 内插残差”的方法要明显好于“气温数据直接内插”的方法。为了说明 4 种空间插值方法在气温数据栅格化中的优劣,将 4 种插值方法同时应用于两种栅格化方法中,并对结果进行分析和比较。

直接内插法是根据全国 560 个气象站的经度、纬度和气温值,在 ARC/INFO 中直接使用 4 种空间插值函数内插生成全国 30" × 30" 分辨率的栅格气温数据。用 57 个验证站(它们没有参加插值运算)的实测数据进行验证,结果见表 1。为增加判断的可靠性,除了用计算气温值与实测气温值之间的相关系数和标准误差作为判断参数外,再增加平均误差作为判断参数之一。平均误差的定义为:验证站计算气温值与实测气温值之差的绝对值再取平均值(下同)。

表 1 4 种空间插值方法直接用于气温数据内插的比较

判断参数	反距离权重	克里格插值	样条插值	趋势面插值
相关系数	0.944	0.937	0.910	0.857
标准误差/℃	2.21	2.34	2.89	3.45
平均误差/℃	1.44	1.41	1.76	2.68

所谓“多元回归 + 内插残差”的气温数据栅格化方法,是首先以气象站(560 个)的气温值为输出变量,以气象站的经度、纬度和海拔高度为输入变量进行多元回归分析并建立回归方程,同时得到每个气象站的气温残差。将气温残差进行空间插值后再加上回归方程的计算结果即得到栅格化的气温数据。4 种空间插值方法在 57 个验证站上的检验结果见表 2。

从表 1 和表 2 可以看出,4 种插值方法不论是直接用于气温数据的直接内插还是对回归分析的气温残差进行内插,反距离权重内插法和克里格插值法都具有较高的精度,样条插值法的精度次之,趋势面插值的精度最低。从气温数据栅格化的实际(通常用“多元回归 + 内插残差”的方法而不是“气温数据直接内插”方法)出发,4 种插值方法的优劣顺序为:克里格插值法、反距离权重法、样条插值法和趋势面插值法。

表 2 4 种空间插值方法用于气温残差内插的比较

	反距离权重	克里格插值	样条插值	趋势面插值
相关系数	0.988	0.990	0.985	0.984
标准误差/℃	1.03	0.96	1.17	1.18
平均误差/℃	0.62	0.62	0.72	3.20

2.2 距平值在气温数据栅格化中的应用探讨

相对于年气温而言,多年平均气温在空间上的变化更加趋于平稳。年平均气温与多年平均气温的差称为气温的距平,气温距平值在气候长期变化研究中具有重要作用。

根据“多元回归 + 内插残差”的方法进行气温数据的栅格化,需要对每年(或每月)的数据进行回归分析和建立相应的回归方程,因为每年(或每月)的情况都是不一样的。这样,在进行栅格化气温数据产品的生产时就显得比较繁琐,那么是否能够采取一种比较简单的方法,即只对多年平均气温数据建立回归方程和对残差进行内插,其它任何年都使用该回归方程和内插后的残差,同时计算该年年平均气温与多年平均气温的距平值并在空间上进行内插,最后再加上栅格化的多年平均气温就得到该年

的栅格化气温。下面以全国 542 个气象站的 30 年 (1951 ~ 1980 年) 平均气温和 1961 年年平均气温为例进行 1961 年年平均气温数据的栅格化, 以另外 57 个气象站的实测气温值进行验证。

(1) 多年平均栅格化气温 + 当年栅格化气温距平
 全国 30 年年平均气温 (542 站) 与气象站经度、纬度和海拔高度的多元回归方程为:

$$T_{30} = 55.29 - 0.1487 X - 0.7032 Y - 0.0044 Z \quad (1)$$

式中, X 、 Y 、 Z 表示气象站的经度、纬度和海拔高度, 单位分别为度、度和 m (下同)。

用该回归方程计算出全国范围内的栅格气温, 同时将回归残差和 1961 年气温距平值进行空间内插 (本次试验用 IDW 内插法), 三者求和得到 1961 年的栅格化气温数据

(2) 当年平均气温直接栅格化

全国 1961 年平均气温 (542 站) 与气象站经度、纬度和海拔高度的多元回归方程为:

$$T_{1961} = 53.66 - 0.1320 X - 0.6925 Y - 0.0045 Z \quad (2)$$

用该回归方程计算出全国范围内的栅格气温, 同时将回归残差进行空间内插 (IDW 内插法), 二者求和得到 1961 年的栅格化气温数据

将上述两种方法得到的 57 个验证气象站的计算气温分别与实测气温进行比较和分析, 结果见表 3。

表 3 两种栅格化方法在 57 个验证站上的比较

	当年回归计算结果 +	多年回归计算结果 +
	当年残差	多年残差 + 当年距平值
相关系数	0.988	0.987
标准误差 / °C	1.02	1.04
平均误差 / °C	0.62	0.66

从表 3 可以发现, 仅用当年的气温数据进行栅格化比用多年平均气温数据和气温距平值进行栅格化的精度要高。图 1 是通过两种栅格化方法计算得到的 1961 年平均气温与实测平均气温的关系图。

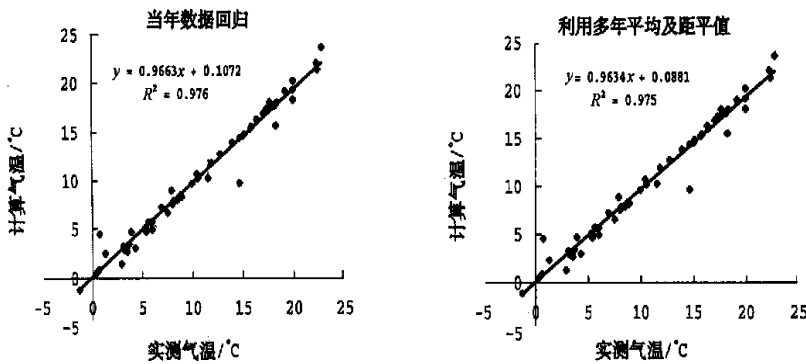


图 1 用两种栅格化方法计算得到的 1961 年年平均气温与实测年平均气温的关系

2.3 DEM 数据误差对栅格化结果的影响

在进行气温数据栅格化的过程中, DEM 数据是非常重要的参数之一, 因为气温随高程的变化而发生规律性的变化。气象部门提供气象数据时也会提供每个气象站的海拔高度, 但由于种种原因, 气象部

门提供的高程与通过气象站经纬度从 DEM 上查询到的高程往往不一致, 表 4 为全国 617 个气象站实际海拔高度与气象站相应经纬度上 DEM 高度的差异情况。

表 4 气象站的实际高度与相应经纬度 DEM 上高度的误差

	误差绝对值范围 / m						
	0 ~ 10	10 ~ 20	20 ~ 50	50 ~ 100	100 ~ 200	200 ~ 500	> 500
气象站个数	241	142	140	47	22	19	6
所占百分比 / %	39.06	23.01	22.69	7.62	3.57	3.08	0.97

从理论上讲, 气象部门提供的气象站高程要比 DEM 上的气象站高程准确, 但在气温数据栅格化的过程中使用的高程数据为 DEM 数据, 因为气象部

门提供的气象站的高程相对于整个国土面积来说毕竟是非常有限的点。因此, 在建立回归方程时究竟是使用气象站实际高程还是使用 DEM 上的高程值

得进一步分析。

仍以全国 560 个气象站 30 年(1951 ~ 1980 年)平均气温数据为例,分别以两种高程进行多元回归分析,回归分析所得到的参数列于表 5。

表 5 气象站实际高程和相应经纬度 DEM 高程分析结果

	回归分析所用高程	
	气象站实际高度	气象站所在经纬度的 DEM 高度
相关系数	0.983	0.981
标准误差/℃	1.29	1.38
平均误差/℃	0.97P	1.04

注:平均误差为残差绝对值取均值

相应的回归方程分别为:

$$T_1 = 55.09 - 0.1452 X - 0.7079 Y - 0.0044 Z \quad (3)$$

$$T_2 = 55.23 - 0.1473 X - 0.7074 Y - 0.0043 Z \quad (4)$$

用基于两种高程建立的回归方程和相应残差分别进行气温数据的栅格化,并把 57 个验证站上的栅格化气温值与其实测气温值进行比较和分析,结果见表 6。

表 6 基于气象站实际高程和相应经纬度 DEM 高程的气温栅格化结果与实测气温的比较

	回归分析所用高程	
	气象站实际高度	气象站所在经纬度的 DEM 高度
相关系数	0.9833	0.9829
标准误差/℃	1.22	1.34
平均误差/℃	0.72	0.70

从表 5 可以看出,用气象站的实际高程进行多元回归,不论是相关系数还是误差均比用 DEM 上的高程进行多元回归要好(但相差不大),这也在一定程度上证明了气象部门提供的气象站高度值比相应经纬度上 DEM 的高度值要准确一些。而从表 6 得到的信息是,用基于气象站实际高度所建回归模型进行气温数据栅格化的精度也比用基于相应经纬度 DEM 上的高度所建回归模型进行气温数据栅格化的精度略高,但同样是相差不大。图 2 是基于两种高程模型所计算的多年平均气温与实测多年平均气温的关系。

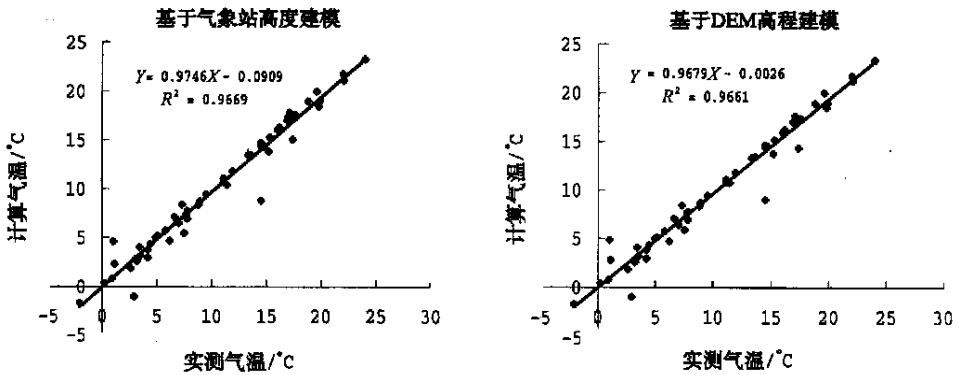


图 2 基于两种高程的模型计算的多年平均气温与实测多年平均气温的关系

3 结论与讨论

本文基于 1951 ~ 1980 年 30 年的平均气温和 1961 年的年平均气温以及 DEM 数据,利用 ARC/INFO 地理信息系统软件和 EXCEL 数据处理分析软件,对气温数据栅格化中的几种空间插值方法、气温距平值在气温数据空间化中的应用以及 DEM 的数据误差对气温数据栅格化结果的影响等问题进行了分析和试验,得到以下结论。

(1) 反距离权重插值、克里格插值、样条插值和趋势面插值等 4 种空间插值方法应用于气温数据栅

格化时,克里格插值方法的精度最高,反距离权重法次之,样条插值法的精度第 3,趋势面插值方法的精度最低。因此,在具体的气温数据栅格化过程时,建议使用前 2 种插值方法。

(2) 基于多年平均气温,通过内插气温距平值的方法也可以进行年平均气温的栅格化。这种方法的优点是不需要对每年的数据完整地进行栅格化,而只需对多年平均气温数据进行完整的栅格化和对当年的气温距平值进行内插,因此减少了数据处理和计算的工作量,但栅格化结果的精度比完全用当年的数据进行栅格化的精度要低。相反,完全基于当年的数据进行栅格化,虽然工作量要大一些,但结果

的精度要高一些。

多年平均气温的气象站点数往往与某一年气温数据的气象站点数据不一致(一般是要少),因此,基于多年平均气温进行气温数据的栅格化往往要使样本数降低,这也是精度降低的原因之一。同时,一些气象站位置(经度、纬度和海拔高度)的变动(但名称和编码不变)信息也很难在多年平均气温的记录中得到反映,有时,同一气象站在多年平均气温数据和年平均气温数据中的坐标并不完全一致。因此,从方便和实用的角度出发,仅用当年的数据(而不用多年平均气温数据)进行栅格化更具可操作性,而一些重复性的工作可以通过程序让计算机完成。

(3) 尽管气温数据的栅格化完全依赖于 DEM,但在栅格化建模过程中,使用气象站的实际高程比使用与气象站经纬度对应的 DEM 上的高程,其结果的精度要高。用 DEM 的高程值建模进行栅格化,尽管可以把气象站所在点的气温值恢复到与实测值一致,但总体上的精度降低了;相反,用气象站

的实际高程值建模进行栅格化,虽然因气象站实际高程与 DEM 之间的差异而不能把气象站所在点的气温值恢复到与实测值完全一致,但在广大的非气象站所在地区的总体精度提高了。

参考文献

- 1 孙和平,罗少聪. 中国及其邻区地表气象数据预处理和网格化数值结果分析. 地壳形变与地震,1998,18(3):51-56
- 2 周锁全,缪起龙,吴战平. 山区平均气温细网格插值方法的比较. 南京气象学院学报,1994,17(4):488-492
- 3 廖顺宝,李泽辉. 基于 GIS 的定位观测数据空间化. 地理科学进展,2003,22(1):87-93
- 4 廖顺宝,李泽辉,游松财. 气温数据栅格化的方法及其比较. 资源科学,2003,25(6):75-80
- 5 董蕙青. 回归分析中因子问题的探讨. 广东气象,1998,4(增2):74-76
- 6 赵仲莲,王澄海. 回归分析在气象应用中的一些问题和进展. 甘肃科学学报,1996,8(2):94-96
- 7 李新,程国栋,卢玲. 空间内插方法比较. 地球科学进展,2000,15(5):571-575

Some Practical Problems Related to Rasterization of Air Temperature

Liao Shunbao Li Zehui

(Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing 100101)

Abstract: Through an experiment of rasterization for 30-year mean air temperature and 1961's mean air temperature data from 617 meteorological stations in China with ARC/INFO and four interpolation methods, it was found that the Kriging method resulted in the highest precision, IDW method the second, Spline method the third, Trend method the lowest. The perennial mean air temperature and annual temperature anomaly can be used to rasterize annual air temperature data. This method will result in less analysis and computation but lower precision compared with the rasterization using annual mean air temperature data only. There is difference between meteorological stations' actual elevation and the DEM (Digital Elevation Model) values corresponding to the meteorological stations' longitudes and latitudes. Rasterization of air temperature data relies DEM fully. However, the model based on meteorological stations' actual elevation will result in higher precision for rasterization of air temperature data than that based on the DEM values corresponding to meteorological stations' longitudes and latitudes.

Key words: air temperature, rasterization, interpolation, Digital Elevation Model, annual temperature anomaly