

文章编号: 1007-6301 (2003) 01-0087-07

基于 GIS 的定位观测数据空间化

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: 属性数据空间化是当前 GIS 领域的前沿问题之一。在对中国 624 个气象站多年平均气温数据空间化过程中, 通过使用 30 秒分辨率的数字高程数据, 把气温分解为受经纬度、海拔高度影响的规律性成分和受其它因素影响的非规律性成分两部分, 并分别用多元回归和反距离权重内插方法对二者实施空间化, 最后将空间化结果进行合成得到基于栅格的中国多年平均气温分布数据。该数据既能反映气温在空间上的宏观变化, 又能反映气温在局部地区的微观变化。该方法可供其它类型观测数据空间化、特别是在观测站点稀疏的情况下参考和借鉴。

关 键 词: GIS; 观测数据; 空间化

中图分类号: N945.1

引言

空间数据和属性数据是地学研究中非常重要的两类数据, 属性数据空间化是目前 GIS 研究领域的前沿问题之一。定位观测数据和社会经济统计数据是 GIS 研究中最常用的属性数据, 属性数据空间化的需求源于数据综合分析的需要, 一方面, 目前遥感已经成为地学研究中的重要数据源, 遥感数据采用栅格结构, 为了与遥感数据进行各种分析, 其它相关的地理信息也必须是栅格结构, 因此需要进行从矢量到栅格的格式转换。另一方面, 非遥感数据之间进行分析, 也需要进行空间化。当进行不同观测数据之间的分析时, 例如水文数据和气象数据, 由于两类数据的观测站点往往不一致, 直接使用两类数据几乎不可能进行空间分析。只有当把每类观测数据从有限的点扩展到面, 并具有相同的地面分辨率时, 才可能进行分析, 这种把观测数据从点到面扩展的过程实质上就是观测数据的空间化。此外, 属性数据空间化还有利于数据质量的控制, 属性数据在空间化之前有时很难发现问题或错误, 而空间化以后, 一些问题或错误根据经验和常识就很容易被发现。

最简单最直接的格式转换方法是空间数据内插和矢量数据的栅格化。但简单的内插和栅格化并不是真正意义上的空间化, 因为它并没有考虑影响统计数据或观测数据的地理环境要素, 其结果的可靠性往往受到质疑。

收稿日期: 2002-07; 修订日期: 2003-01

基金项目: 本研究受国家科技基础性工作专项资金课题 (2001DEA30027-9) 和中国科学院知识创新工程项目资助 (INF105-SDB-1-18)

作者简介: 廖顺宝 (1966-), 男, 博士, 副研究员, 主要从事遥感与地理信息系统应用研究。E-mail: liaosb@igsrr.ac.cn

气象数据是典型的定位观测数据, 数据同化技术已成为气象数据的基本方法。在气象学上, 数据同化是指分析处理随空间、时间分布的观测数据, 从而为数值预报提供初始场的过程, 它与通常所说的数据空间化有类似的含义。目前, 经典的数据同化方法如多元回归方法、逐步插值法等, 在气象业务预报中仍占主导地位。多元回归和插值是两种不同类型的数据空间化方法, 各有特点。周锁铨等 (1994) 利用逐步插值和多元回归的方法计算贵州 1 月、7 月 10 分 \times 10 分网格的平均气温, 发现逐步插值法计算结果的剩余误差仅为多元回归法的 $1/5 \sim 1/10$ 。孙和平等 (1998) 利用反距离权重法用 1200 个台站的数据, 计算出中国及邻近地区 1 度 \times 1 度网格的气压, 结果与真实值几乎完全一致。李新等 (2000) 在研究和比较各种空间内插方法后得出结论认为, 没有绝对最优的空间内插方法, 而只能根据具体数据选择最合适的方法, 同时应对内插结果进行检验。王劲峰 (1995) 在研究空间信息的结构自适应模型时把空间地学信息先分解为全局线性趋势、局域时空线性趋势、周期信号和非线性随机信号等几个部分, 然后再对各部分进行整合。插值又分为点插值和面插值, 但面插值主要于社会经济统计数据 (潘志强等, 2002) 的空间化。

1 方法选择

利用多元回归和插值进行数据的空间化, 其机理是不同的。多元回归是通过有限站点的观测数据建立观测值与环境要素 (经度、纬度和海拔高度) 之间的关系, 从而利用该关系式计算非观测站点所在地区的要素值。多元回归虽然在一定程度上说是从“机理上”进行观测数据的空间化, 它既能反映数据在区域尺度上的宏观变化趋势, 又能反映数据在局部地区的微观变化。但由于自然环境的复杂性, 人们很难找出影响观测要素的全部因素, 因此计算结果与实际数值总会存在差异, 即所谓残差。

内插方法则忽略环境要素, 它仅根据观测值和一定的规则 (如反距离权重) 来计算和内插非观测站所在地区的要素值, 原理和方法都较简单。显然, 该方法比较适合于环境条件变化不大或变化较为规律的地区。

本文以中国 624 个气象台站的 30 年平均气温数据为例, 首先通过多元回归方法找出气温与环境要素 (经度、纬度和海拔高度) 之间的关系并根据该关系式计算出所有栅格的气温值, 形成结果 1, 同时计算出各观测站气温残差, 通过反距离权重方法内插出所有栅格的残差, 形成结果 2, 最后对结果 1 和结果 2 进行求和得到最终结果。其基本思想是, 认为气温由随环境 (经度、纬度和海拔高度) 变化而变化的规律性成分和受局部非规律性成分两部分组成, 回归计算结果代表气温随环境变化的规律性成分, 空间化后的残差代表气温变化的非规律性成分。

2 数据源准备与预处理

2.1 数据源

中国 624 个气象站 30 年平均气温数据;

中国国界图。

2. 2 数据预处理

数据预处理主要包括以下内容:

- 在全球 30 秒分辨率的数字高程图中, 中国跨越 4 个图幅, 编号分别为 E060N40、E060N90、E100N40 和 E100N90, 每个图幅解压后包括 8 个文件, 将每个图幅的 .dem 和 .hdr 文件拷贝至 ARC/INFO 当前工作目录。

- 将 .dem 文件的后缀改为 .bil。

- 在 ARC/INFO 的 GRID 模块中用 IMAGEGRID 命令将 .bil 文件转换为 GRID 文件。

- 利用 MOSAIC 命令将 4 个图幅拼接在一起。

- 由于 IMAGEGRID 命令在将 IMAGE 转换为 GRID 时, 不支持负整数的转换, 因此需要通过运算 $out_grid = con(in_grid \geq 32768, in_grid - 65536, in_grid)$ 对转换后的 GRID 进行修正。

- 拼接后的图象范围为 E60-E140; N90-S10, 而中国的主要面积仅在 E72-E136 N54-N18 的范围内, 为去除无关的数据, 在 GRID 中利用 GRIDCLIP 命令对整个区域进行切割。

- 利用 CON 函数将 DEM 中的 -9999 设置为 NODATA。

- 通过自己编写程序生成经纬度网格数据。数据的空间化要使用经纬度数据, 而在 ARC/INFO 中, 尽管 DEM 数据隐含有每个栅格的经纬度坐标, 但不能取出来进行运算, 因此需要重新构建每个栅格的经度和纬度数据。

GRIDASCII 命令能将 GRID 转换为 ASCII 文件, 而 ASCIIGRID 命令能把相应的 ASCII 文件转换为 GRID, 因此, 利用简单的程序可以生成符合要求的经度 ASCII 文件和纬度 ASCII 文件, 然后再分别将它们转换为经度 GRID 和纬度 GRID。

3 数据分析

3. 1 模型构建

研究表明, 经纬度 (特别是纬度) 和海拔高度是影响一个地区气温变化的主要因素。在 EXCELL 中, 以各观测站的多年平均气温为输出变量, 以各观测站的经度、纬度、和海拔高度为输入变量进行多元回归分析, 结果表明, 各观测站的多年平均气温与经度、纬度和高度明显相关, 其关系式为:

$$Y = 55.1695 - 0.1466Long - 0.7069Lati - 0.00437Alti \quad (1)$$

式中, Long、Lati 和 Alti 分别表示某地的经度、纬度和海拔高度, 单位分别为度、度和米。线性相关系数 $R = 0.98$, 标准误差为 $SD = 1.25$ 。图 1 表示了各观测站实际观测值与通过公式 (1) 计算值的线性相关关系, 图 2 表示了各观测站气温观测值与计算值的误差情况。

从残差分布图可以看出, 绝大部分测站的误差绝对值小于 2°C , 少量测站的误差绝对值在 $2\sim 4^{\circ}\text{C}$ 之间, 个别测站的误差绝对值在 $4\sim 6^{\circ}\text{C}$ 之间, 具体情况列于表 1。

误差绝对值在 2°C 以内的观测站超过了 90%, 说明所建立的回归方程基本上反映了多年平均气温随经纬度和海拔高度的变化规律, 可以作为气温数据空间化的重要依据。尽管如此, 仍有将近 10% 的测站误差超过 2°C 。因此, 为了使空间化产生的数据尽可能地反映

气温的实际分布情况，必须对模型计算结果进行修正。

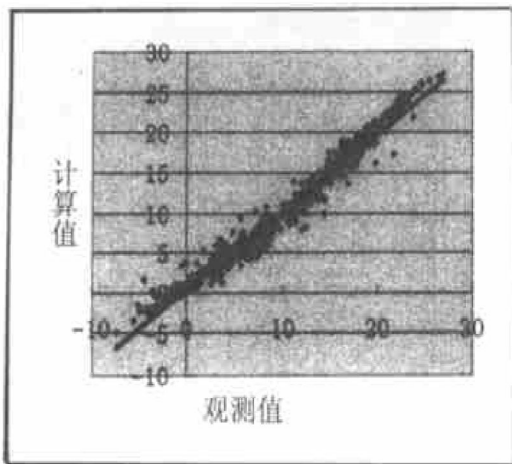


图 1 模型计算结果与实际观测值的关系

Fig.1 Relationship between the results of model calculation and actual measurement

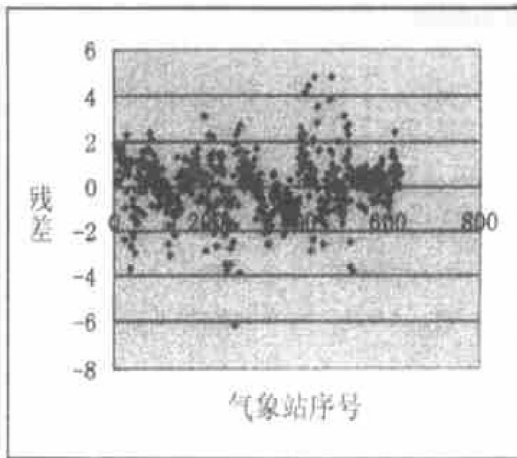


图 2 各站气温残差分布情况

Fig. 2 Temperature residual's distribution of every observation station

表 1 误差绝对值在全部观测站中的分布情况

Tab.1 Distribution of error's in the all observation stations absolute value

误差范围 (°C)	< 1	1~2	2~3	3~4	4~6
测站个数	393	171	42	13	5
所占百分比 (%)	63.0	27.4	6.7	2.1	0.8

上述分析使用的观测站高度为气象资料中提供的高度，当以 DEM 中的观测站高度参与分析时，其结果也几乎完全一样。

3. 2 误差修正

公式 (1) 反映了气温随经纬度和海拔高度变化的情况，但在反映气温受其它局部因素影响方面仍存在不足，因为有 60 个站的气温误差值超过 2℃，其中超过 3℃ 的观测站有 18 个，超过 4℃ 的观测站有 5 个，最大误差达 6℃。

本研究涉及的 624 个气象观测站为反映中国气象、气候变化的主要观测站，同一种气象观测要素在这些观测站之间的变化情况反映了该气象要素在全国的宏观变化趋势，因此，在进行数据的空间化时，首先应当保证空间化获得的数据在这些观测站上与实际观测值一致，这一点可通过模型计算值与残差求和来实现，因为各观测站的残差是已知的。其次，还必须把各观测站残差分布情况推广到广大的非观测站所在地区，即进行残差的空间内插。

本文采用反距离权重法进行残差的空间内插，内插后，各观测站的残差值与原值有些微小的差异，但绝大部分观测站的差异值小于 0.5℃，见图 3、图 4。图 5 为经过内插以后的全国气温残差分布图。

4 气温数据空间化

4. 1 数据空间化

首先利用公式 (1) 计算出全国的多年平均气温分布图，结果如图 6 所示，然后将图 5

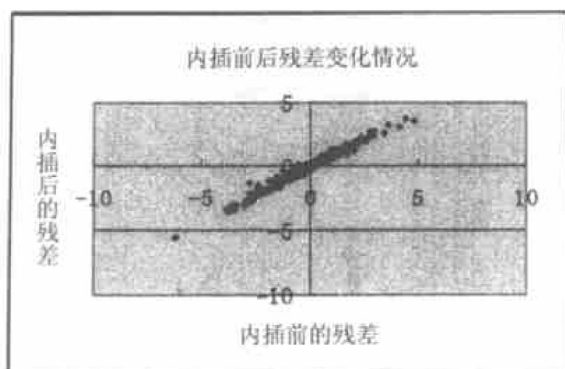


图 3 内插前后各观测站气温残差的关系

Fig.3 Temperature residual's relationship before interpolation and after interpolation

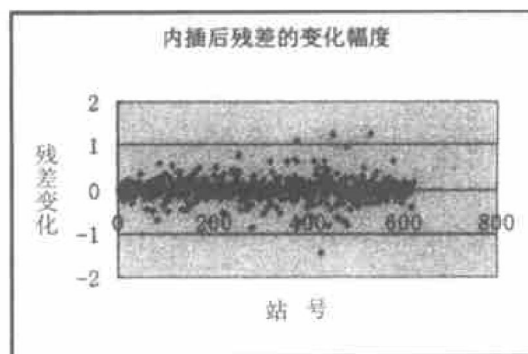


图 4 内插后各观测站残差变化幅度

Fig.4 Residual's variety range of every observation station after interpolation

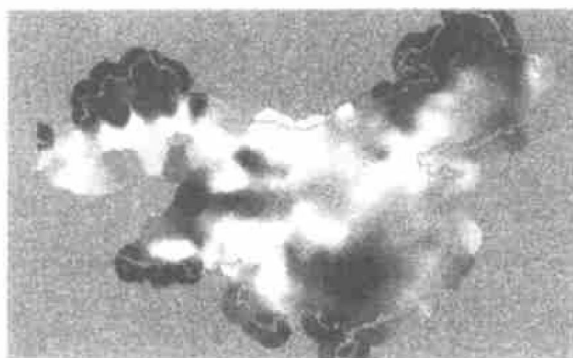


图 5 内插以后的全国气温残差分布图

Fig.5 Distribution figure of China's temperature residual after interpolation



图 6 根据回归模型计算出的全国多年平均气温分布图

Fig.6 Distribution figure of China's average temperature calculated by means of regression model

和图 6 进行求和运算, 即在 ARC/INFO 的 GRID 中进行两个 grid 的求和运算, 最终获得全

本文根据多元回归和反距离权重内插方法进行了中国多年平均气温的空间化, 将空间化结果按每 2°C 一个级差进行可视化并与 1999 年出版的《中华人民共和国国家自然地图

集》中的中国多年平均气温图对比,二者在分布规律和变化模式上非常接近。这说明按照本文的空间化方法所获得的数据比较客观地反映了中国多年平均气温的宏观变化趋势。同时,由于空间化时考虑了影响气温变化的地理要素(经度、纬度和高度),因此它又反映了气温在空间上的微观变化,其结果可用于与空间变化有关的多种研究,如气候生产潜力的空间分布,生物量的空间分布等等。

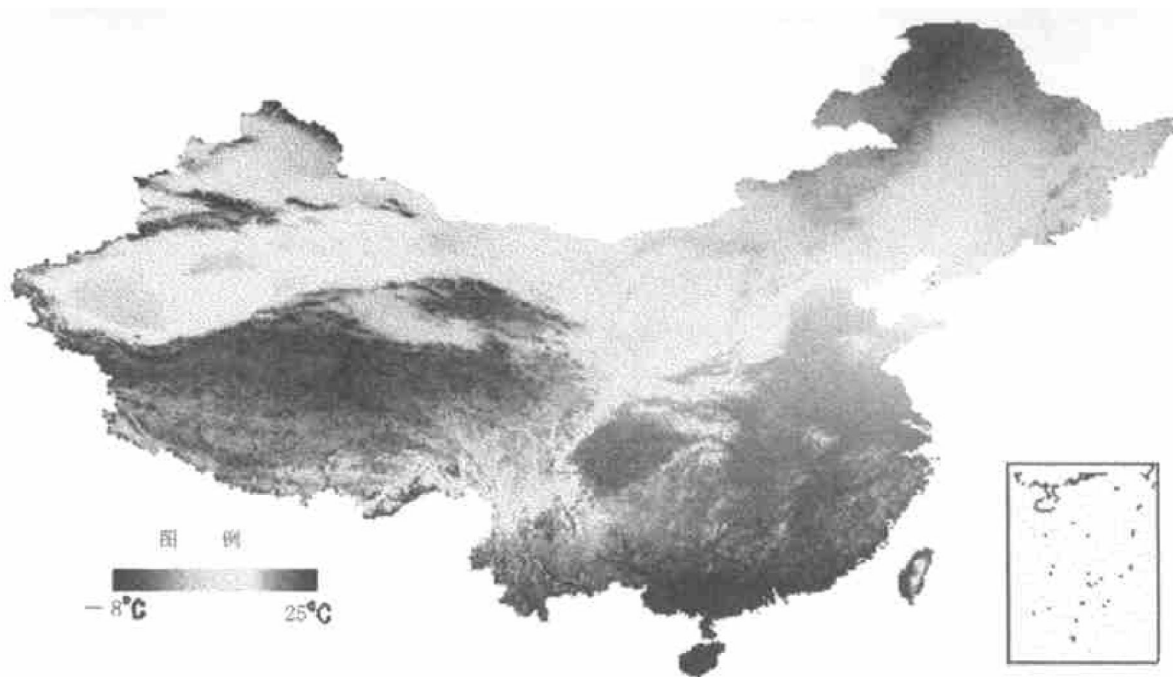


图7 修正以后的全国多年平均气温(1951~1980)分布图(0.1°C)

Fig. 7 Distribution figure of China's average temperature after modified from 1951 to 1980

虽然本文是以中国多年平均气温为例来探讨定位观测数据的空间化,但其原理和方法完全可以用于其它气象要素和其它观测要素的空间化,如年平均气温、月平均气温、旬平均气温、候平均气温以及湿度、气压、降水等要素的空间化。特别是在观测站点稀疏的情况下,该方法具有较高的借鉴和参考价值。

参考文献

- [1] 潘志强,刘高焕.面插值的研究进展.地理科学进展,2002,21(1).
- [2] 李新,程国栋,卢玲.空间内插方法比较.地球科学进展,2000,15(3):260~265.
- [3] 高山红,吴增茂,谢红琴.Kalman滤波在气象数据同化中的发展与应用.地球科学进展,2000,15(5):571~575.
- [4] 国家地图集编撰委员会,中华人民共和国国家自然地图集,北京:中国地图出版社,1999.
- [5] 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.
- [6] 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.
- [7] 王劲峰.空间信息的结构自适应模型.地理学报,1995,50(增刊):51~61.
- [8] 周锁全,缪起龙,吴战平等.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.

参考网站: <http://edcdaac.usgs.gov/gtono30/gtono30.html>.

A Methodology of Spatialization of Observed Data Based on GIS

LIAO Shun-bao, LI Ze-hui

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

Abstract: Spatialization of attribute data is one of forward issues in the field of GIS. While 30-year mean temperature data from 624 meteorological stations in China was spatialized, the temperature was divided into regular component, which is affected by longitude, latitude and altitude, and irregular component affected by other local factors. They were spatialized with multiple variable regression and inverse distance weighted interpolation respectively. There was a correlation ratio of $R = 0.98$ between temperature and geographical factors including longitude, latitude and altitude. Summing the two spatialized components generated grid-based temperature data. It can reflect temperature change both at large scale and at small scale. The method used in this paper can be applied in spatialization of other observed data and it is especially suitable when distribution of observation stations disperses.

Key words: GIS; Observed data; Spatialization