

我国地面气温格点化数据集的研制

张 强 熊安元 阮 新
(国家气象信息中心,北京 100081)

摘 要:在大气科学研究领域,特别是气候变化检测和数值模拟研究中,与高精度、高时空分辨率的格点资料相比,离散的站点观测资料具有明显的优势。随着国家气象科学数据共享工作的纵深开展,站点资料已不能完全满足用户的需求,通过空间插值技术进行格点化气象要素数据集的研制和服务日益迫切。本文结合国家气象信息中心最新研制的中国地面气温格点数据集,对数据集研制过程中的几个关键技术问题及数据集验证情况进行了阐述。

关键词:气温;空间插值;格点数据集;数据共享

中图分类号: P423.7 **文献标识码:** A **DOI:** 10.3772/j.issn.1674-1544.2008.03.003

Research and Development of the Grid - based Surface Air Temperature in China

Zhang Qiang , Xiong Anyuan , Ruan Xin
(National Meteorological Centre, Beijing 100081)

Abstract: In comparison with the observational station data, the high - precision and high - resolution grid data have apparent advantage in the research field of the atmospheric science, especially in the climate change detection and numerical simulation. Due to the wide sharing of the national meteorological scientific data, the station data can not fully meet users' demands, which makes it more and more urgent to develop the grid data set of meteorological elements through exploiting the spatial interpolation skill. Thus, the present study is motivated to illustrate several key issues on the research and development of the grid - based surface air temperature data set in China developed by National Meteorological Information Center. The validation of the data set is assessed as well.

Keywords: surface air temperature, spatial interpolation, grid data set, scientific data sharing

1 引 言

气象资料是气候系统状态及其演变的客观记载,是国家基础信息资源的重要组成部分。2001年底,科技部启动了“气象资料共享系统建设”项目,支持中国气象局率先开展科学数据共享试点

工作。随着项目的深入开展,越来越多的气象资料面向社会共享,发挥了重要的作用,并产生了良好的社会和经济效应。然而,这些资料中绝大部分是基于定点持续观测积累的气象要素序列,由于站点分布不均、序列长短不一、观测台站环境变迁等问题,在气候分析和研究中,站点资料不能完全有效地代表研究区域的气候变化特征,在实际应用

第一作者简介:张强(1977-),男,重庆巫溪人,工程师,研究方向是气候资料质量控制与评估、气候变化等。

基金项目:国家科技基础条件平台建设项目“气象科学数据共享中心建设”(2005DKA31700-02)。

收稿日期:2008年5月1日。

中受到诸多限制。利用空间插值技术将离散的站点资料转换成规则的网格点序列,是一种“浓缩”气象要素场信息的有效方法,从而大大提高了序列在对应网格范围的气候代表性。

国外科研机构非常重视站点资料的网格化,对其进行了大量的研究工作^[1-9],根据不同的研究方向和使用目的,建立了不同版本的全球或区域格点数据集。在这些数据集中较有影响的是:在降水网格化研究中最具代表性的 CMAP 数据集,目前最高水平的全球高分辨率(0.5°×0.5°) CRU TS 2.0 数据集,COADS 全球格点数据集,Dai 降水数据集,Hulme(CRU)降水数据集,Jones(CRU)全球气温格点数据集,Hanson 数据集等。这些数据集作为认识气候变化的基础数据已经被政府间气候变化专门委员会(IPCC)作为主要的参考依据^[5]。但是,对于中国区域而言,这些格点化数据集所包含的站点观测信息有限(通常只有约 200 个台站的观测资料),没有对历史资料的均一性进行检验,其资料的质量及其代表性在中国区域存在的问题。我国在这方面的研究目前还比较零散^[10-18],并没有一个权威的格点气象资料数据集。

针对以上现状,在科技部“气象科学数据共享中心”项目支持下,国家气象信息中心以最新研制的 731 个台站的“中国近 50 年均一化日气温数据集(1951-2004 年)”、“中国近 50 年均一化月平均气温数据集(1951-2004 年)”、“中国地面气候资料日值数据集”、“中国地面气候资料月值数据集”为主要数据源,参考国家气候数据中心(NCDC)研制的全球历史气候网数据集(GHCN_V2.0),利用 GTOPO30 数据生成的中国范围 1°×1°的数字高程模型数据,采用改进的 Kriging 插值方法分别对中国 1951-2007 年的气温资料进行逐日、逐月、逐年和累年值的空间插值处理,获得了“中国地面气温格点数据集(V1.1)”。该套数据集网格密度为 1°×1°,采用 ARCGIS 和 TXT 文本两种格式保存。

2 格点数据集研制的几个关键问题

2.1 插值方法的选择

由于各种空间分析方法基于不同的地理假设和不同的数据建模方法,对于众多的空间数据分

析方法而言,没有绝对最优的空间插值方法,只有特定条件下的最优方法^[19]。研究证明,并不是所有的方法都适应气象要素场的要求。本文从常用于气象要素的空间插值方法中,重点选取反距离加权插值法(Inverse Distance to a Power)、普通克里金插值法(Ordinary Kriging,简称 OK)、最小曲率法(Minimum Curvature)、线性插值三角网法(Triangulation with Linear Interpolation)以及多元回归法(Polynomial Regression)5种插值方法进行温度场插值效果检验。

2.1.1 检验标准

为了估算以上 5 种不同插值方法对实际温度场的可能分析误差,采用交叉验证法验证插值的效果。交叉验证法首先假设每一站点的气象要素值未知,用周围站点的值进行估算,然后根据所有站点实际观测值与估算值的误差大小评判插值方法的优劣。具体采用平均误差(ME)、平均绝对误差(MAE)、误差方差(EV)、均方根预测误差(RMSPE)作为评估几种插值方法的插值效果的标准。其中,ME 总体反映估计误差的大小,MAE 估量估计值可能存在的误差范围,RMSPE 表达式为:

$$RMSPE = \sqrt{\frac{1}{n_i} \sum_{k=1}^{n_i} [Z_i^*(X_k) - Z_i(X_k)]^2} \quad (1)$$

均方根预测误差越小,预测值越接近真实值。插值方法的均方根越小,插值方法越好。

2.1.2 检验结果

以 2001 年 7 月全国平均温度场为例,分别使用 5 种空间数据分析方法进行空间插值,插值效果如表 1 所示。

以上交叉验证结果显示,普通克里金法相对于其他几种方法的各项误差检验标准偏低,对温度场的插值具有明显的优势。

2.2 克里金插值方法的改进

2.2.1 普通克里金法

克里金插值法是以空间结构分析为基础进行估值的。它首先考虑的是空间属性在空间位置上的变异分布,确定对一个插值点值有影响的距离范围,然后用此范围内的采样点来估计待插点的

表1 5种插值方法的交叉验证结果

插值方法	反距离加权	普通克里金	最小曲率法	线性/三角网	多元回归
ME	-0.0255	0.0125	-0.0338	-0.0351	-0.0020
MAE	0.6323	0.5559	0.9370	0.6636	0.9686
EV	0.8368	0.6329	1.6078	0.7203	1.4766
RMSPE	0.9145	0.7950	1.2623	0.8878	1.2143

属性值。在插值过程中,该方法反映了空间场的各向异性,充分利用了数据空间场的性质和数据点之间的空间相关性。

设在一个区域内采样位置坐标为 X_i , 变量观测值为 $Z(X_i)$, $i=1, 2, \dots, n$, 则预测点的估计值 $Z^*(X_0)$ 可用这 n 个样本点的线性组合来表示:

$$Z^*(X_0) = \sum_{i=1}^n \lambda_i Z(X_i) \quad (2)$$

式中, $Z^*(X_0)$ 是 (X_0) 位置的估计值, $Z(X_i)$ 是站点 X_i 位置的测量值, λ_i 是分配给站点 X_i 位置的权重。

普通 Kriging 方程为:

$$\begin{cases} \sum_{i=1}^n \lambda_i \gamma(X_i, X_j) + \mu = \gamma(X_j, X_0), & i, j = 1, \dots, n \quad (3) \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad (4)$$

$\gamma(X_i, X_j)$ 是观测点 X_i 与 X_j 之间的变异函数值; $\gamma(X_i, X_0)$ 是变异函数模型计算出的样本点 X_i 与内插点 X_0 之间的变异函数值。

2.2.2 普通克里金法的改进

普通克里金法在诸多气象要素场的空间分析方法中具有一定的优势,但中国地形复杂,特别是青藏高原大地形作为中国区域特有的地理因素,对温度场的分布影响巨大^[14]。如何在复杂地形条件下利用有限的观测数据插值得到气温的合理空间分布始终是一个难题。普遍认为,为了减小误差就必须把高程对气温的影响作为一个重要的因素来考虑。本文在普通克里金法的基础上,利用高程资料对其进行了改进,以适应中国独特地形的要求。

假设空间变量由背景值和分析值增量组成,那么,某一网格上的气温可以被表示为:

$$T(X_0) = T^*(X_0) + \Delta T(X_0) \quad (5)$$

式中, $T^*(X_0)$ 是气温背景值,由下式计算:

$$\begin{cases} T^*(X_0) = \sum_{i=1}^n \lambda_i T^*(X_i) \\ T^*(X_i) = T(X_i) - \gamma[H(X_i) - H^*] \end{cases} \quad (6) \quad (7)$$

式中, $T^*(X_i)$ 、 $T(X_i)$ 和 $T(X_i)$ 分别是第 i 个气象站的气温背景值、气温观测值和数字高程; H^* 是进行气温订正的高程; γ 是气温直减率。当气温被订正到同一高程后,可以被认为是一个随机变量,因此采用普通克里金插值法对它作估值,见公式(6)。

$\Delta T(X_0)$ 是分析增量,由下式计算:

$$\Delta T(X_0) = \gamma[H(X_0) - H^*] \quad (8)$$

式中, $H(X_0)$ 是待计算网格的高程,取值于 DEM。

2.2.3 订正高程确定

在参加空间插值的全国 731 个台站中,最低海拔 2 米,位于江苏射阳站(区站号:58150、经度:120.25°E、纬度 33.77°N),最高海拔 4800 米,位于西藏安多站(区站号:55294、经度:91.10°E、纬度 32.35°N),平均海拔 955.635 米。在利用改进的克里金法时,公式(4)中的 H^* 取值设为 1000 米。

2.2.4 气温直减率确定

气温直减率 γ 一般随温度、小地形(迎风坡、高原、盆地等)因素而发生变化,干空气温度直减率 $\gamma = -0.976^\circ\text{C}/(100\text{m})$, 湿空气温度直减率 $(800 \sim 1000\text{hPa}) \gamma = -0.32 \sim -0.94^\circ\text{C}/(100\text{m})$ 。如在英国,报道的 γ 范围为 $-0.59 \sim -0.74^\circ\text{C}/(100\text{m})$ ^[20]; 在加拿大北部,报道的 γ 范围为 $-0.5 \sim -1.0^\circ\text{C}/(100\text{m})$ ^[21]。本文中 γ 取值为 $0.65^\circ\text{C}/(100\text{m})$ 。

2.3 温度场空间结构分析

利用克里金法进行中国区域温度场的插值,

其插值精度取决于对温度场空间结构的准确分析,以及选取适合的变异函数模型去拟合这种空间结构。对中国区域 1951 年 1 月至 2007 年 12 月的温度场的空间结构分析表明,中国区域地面气温变异函数的试验曲线基本可以被拟合为幂函数模型或是高斯模型。图 1 显示了这种拟合结果。

2.4 估值领域优化

一般情况下,估值领域越大,克里金估值精度越高。但是,当估值领域大到一定程度时,即使空间变异函数较理想,相距较远的样本点之间、样本点与待估点之间也不具备相关性,领域内的样本点对估值点没有任何贡献,甚至克里金估值的平稳假设都会受到影响。

已有研究结果表明,气象要素的空间相关性

是有限的,对不同的气象要素相关的临界范围是不同的,对于月值的气象要素,温度超过相关信度距离为 1200 公里,而降水只有 450 公里。在本文中,变异函数的变程尽量控制在 20~22 格网距(约 1000 公里左右)之间。

2.5 高程对插值效果的影响评估

表 2 显示了对 1980 年 7 月中国温度场进行高度订正前后 Kriging 插值的交叉验证结果,从结果的平均误差(ME)、平均绝对误差(MAE),误差方差(EV)、均方根预测误差(RMSPE)等 4 种插值效果检验标准可以明显看出,高程因子在空间插值中起着重要的作用,考虑了高程因子的 Kriging 插值方法,相比普通 Kriging 插值方法在插值效果上有明显的改进。

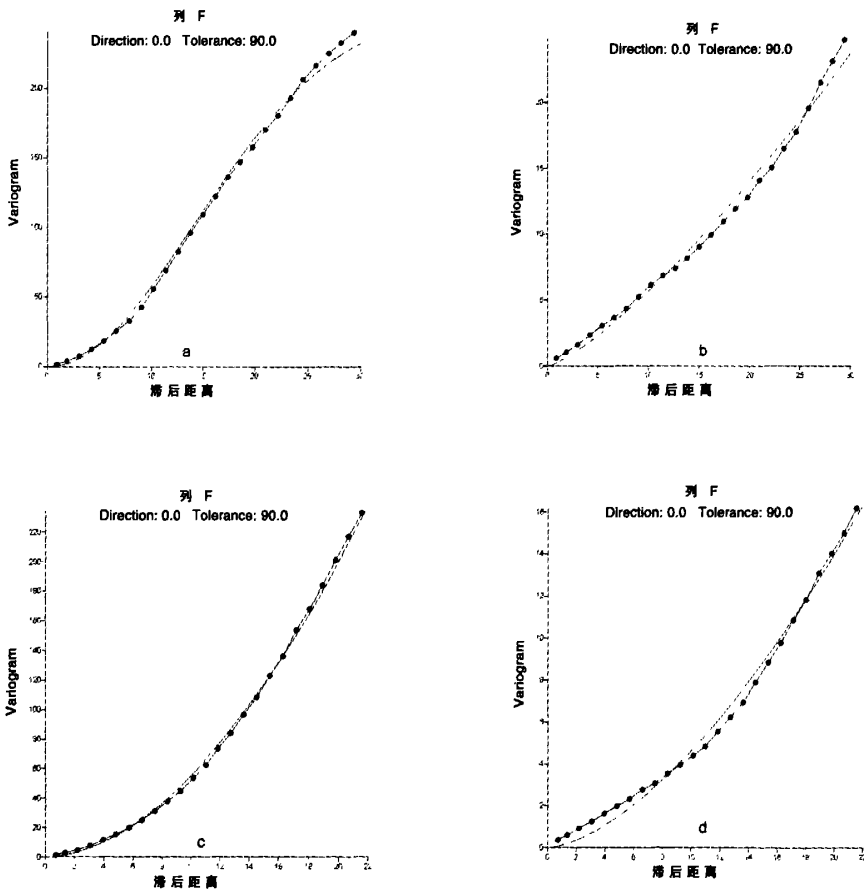


图 1 月平均气温试验变异函数及其拟合效果

(a: 1980 年 1 月 - 高斯模型, b: 1980 年 7 月 - 幂函数模型, c: 2001 年 1 月 - 幂函数模型, d: 2001 年 7 月 - 幂函数模型)

表2 1980年7月高度订正前后 Kriging 插值的交叉验证结果

	ME	MAE	EV	RMSPE
高度订正前	0.0432	1.6579	7.1093	2.6647
高度订正后	0.0072	0.5824	0.7269	0.8520

2.6 边界效应对插值的影响

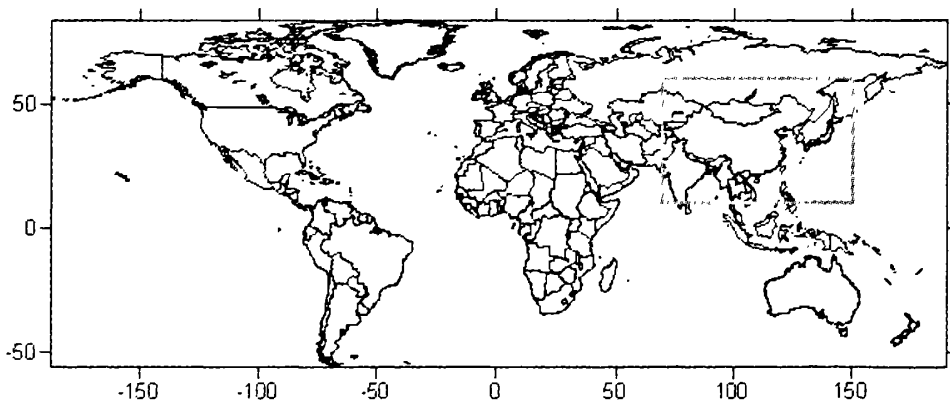
2.6.1 边界效应

对于中国区域,如果只使用国内的观测资料,那么无论使用何种空间数据分析方案,对于边界上的地区而言,由于没有相关的空间结构信息,空间数据分析方法只能依靠边界内部资料确定的空间结构信息进行外推。这种数学上的外推由于缺乏实际空间结构信息的限制,将会对插值结果造成较大的误差。这种误差则被称为由于缺乏边界外的信息而导致的边界效应。

为了消除边界效应,扩大选取资料范围是最直接和最有效的方法。通过引入中国区域以外站点的观测资料,将依靠边界内部资料确定的空间结构信息进行外推,变为使用已知资料和空间结构信息的内插,这将极大地减小由于边界效应导致的空间分析误差。

2.6.2 边界资料选择

图2为资料选择区域示意图,资料选取的范围为 $70^{\circ}-150^{\circ}\text{E}$, $10^{\circ}-60^{\circ}\text{N}$ 。按照IPCC第2次评估报告的全球气候区划,我们所选取的区域涵盖了整个东亚地区和高原地区,包含部分北亚地区、南亚地区和东南亚地区。

图2 国外资料选择区域($70^{\circ}-150^{\circ}\text{E}$, $10^{\circ}-60^{\circ}\text{N}$)示意图

NCDC 研制的全球历史气候网数据集 (GHCN - V2.0) 已经作为世界气象组织 (WMO) 监测和评估当今全球气候变化的核心参考数据。本文采用 GHCN - V2.0 资料作为中国区域以外的站点资料进行空间插值分析。图3是用GHCN - V2.0 资料对中国区域站点补充前后的台站数量对比图。图中显示,经过GHCN - V2.0 资料的补充,上世纪90年代之前参加空间插值的站点平均密度有大幅度的提高,特别是上世纪60-90年代,参加空间插值的站点平均密度由600个左右提高到1000个左右。上世纪90年代以后,由于在GHCN - V2.0 资料中中国区域的站点数较少,站点补充效果不明显。

图4显示了用GHCN - V2.0 资料补充后,参加实际空间插值的站点分布情况(以1980年1月为例,进行空间插值的站点数由补充前的686个达到补充后的1042个)。图中显示,除中国东南沿海地区观测台站较少外,中国区域其他边界周围的台站补充较为充分,特别是东部沿海地区受日本和朝鲜半岛的外围资料补充十分明显。

2.7 数据集的存储方式

格点数据集的存储不仅要考虑到存储文件的可读性和有效性,同时也要考虑到用户的使用需求。本数据集采用GIS中通用的ARCGIS标准格式和TXT文本两种方式存储。其中,ARCGIS标准格式主要用于绘图和分析,数据实体包括文件头和文件实体两部分。数据集网格覆盖范围包括中国区域共1080个网格点(图5)。

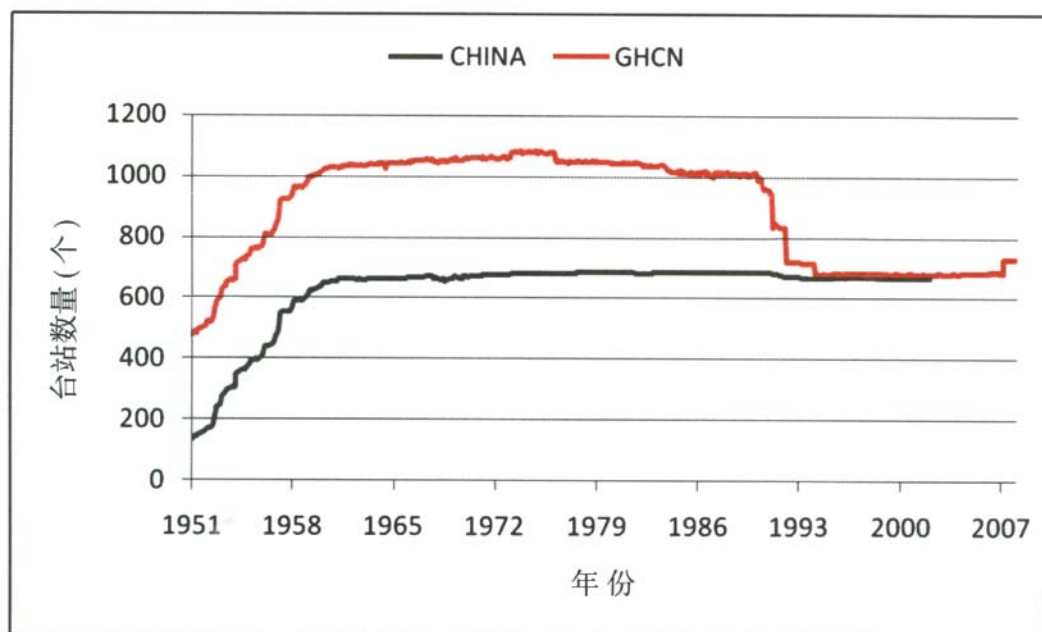


图3 1951年1月-2007年12月GHCN资料进行中国站点补充前后台站数量对比

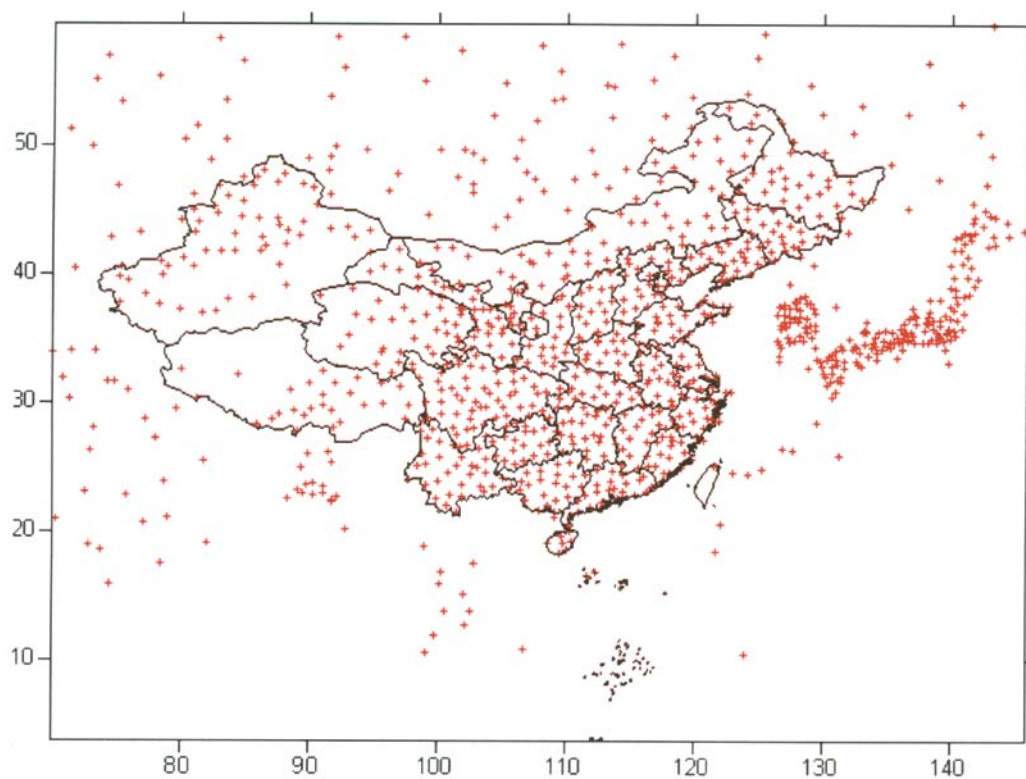


图4 1980年1月GHCN-V2.0资料补充后的站点分布图

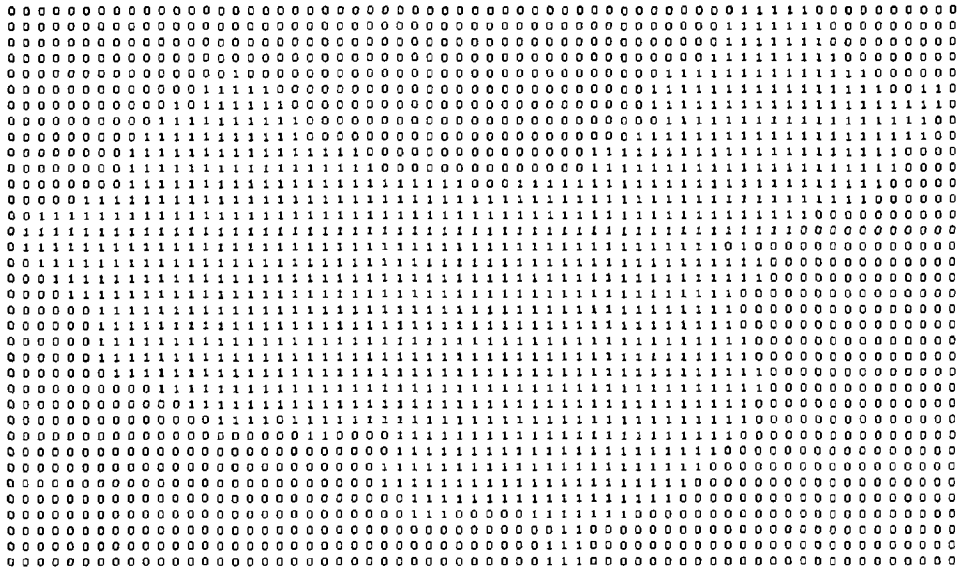


图5 中国地面气温格点数据集网格覆盖分布

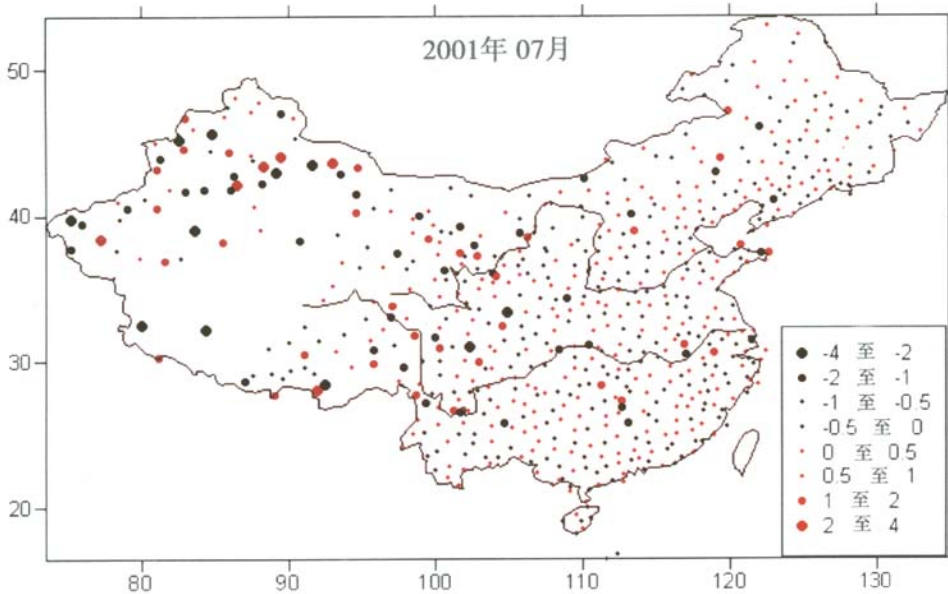


图6 2001年7月温度格点场的残差空间分布

插值精度检验。图6是2001年7月温度格点场的残差空间分布图。

3 数据集检验及误差分析

3.1 数据集检验

3.1.1 数据集天气学验证

利用交叉验证法对数据集从天气学角度进行

从图6可以看出,对于全国区域而言,使用国外资料消除了边界效应,在边界上的残差总体减小,主要绝对残差值较大的站点集中分布于新疆北部地区。该地区由于受天山山脉地形的影响,温度空间分布复杂,导致插值效果较差。青藏高原大部分地区插值效果比较理想,个别边界站点仍然

受站点稀少的影响,导致插值效果较差。总体而言,全国大部分地区插值效果较好,随着插值时间精度的减小,逐日、逐月、逐年的插值效果逐步提高。从统计结果来看,气温资料的格点化数据集从天气学精度基本达到了预期要求。

3.1.2 数据集气候学检验

图 7 比较了格点资料和站点资料对中国东部地区 (103°E 以东) 1951 - 2005 年间年平均气温变化状况的反映。可以看出,中国上世纪 50 年代年平均气温变化差异较大,气温明显偏高。这主要是由于 1951 - 1960 年期间,中国境内基本、基准观测站稀少,且分布不均匀造成的统计误差。1961 年后,随着国内站点的增加,站点资料与格点资料反映的中国年平均气温变化状况的差异也随之减小,格点资料序列与站点资料序列吻合得很好,年际变化的趋势和程度都非常一致。

3.2 数据集误差分析

本文建立的中国地面气温格点化数据集的误差可以归结为以下几个部分。

3.2.1 观测误差

由于测量仪器、方法、环境影响及人为失误等

因素,使得任何气象测量结果都不能做到绝对严密,从而导致测量结果不可避免地存在误差。这些误差根据其性质来源可以分为系统误差和随机误差^[22]。用于插值的原始站点观测气温资料的误差最终会累加到格点化数据集中。另外,数字高程模型 (DEM) 客观上也存在误差,DEM 的精度主要受原始资料的精度和内差精度的影响^[23]。同样,在使用温度直减率时,由于难以确定全国不同区域和地形条件下的准确值,也将影响最终插值结果的精度。

3.2.2 模型误差

在建立数学模型时,由于忽略了一些次要因素而产生的误差,称为模型误差。这种误差不受计算方法的限制和控制。克里金估值是在一定条件下的无偏性和最佳性的线性估值,建立在一定假设条件下的估计误差是客观存在和不可避免的^[24]。另外,克里金空间分析中假设气象要素场在空间是连续的,但对于特殊区域可能存在空间不连续或空间异质,这将造成模型估计误差。

3.2.3 截断误差

截断误差是对参与计算的数学公式做简化处

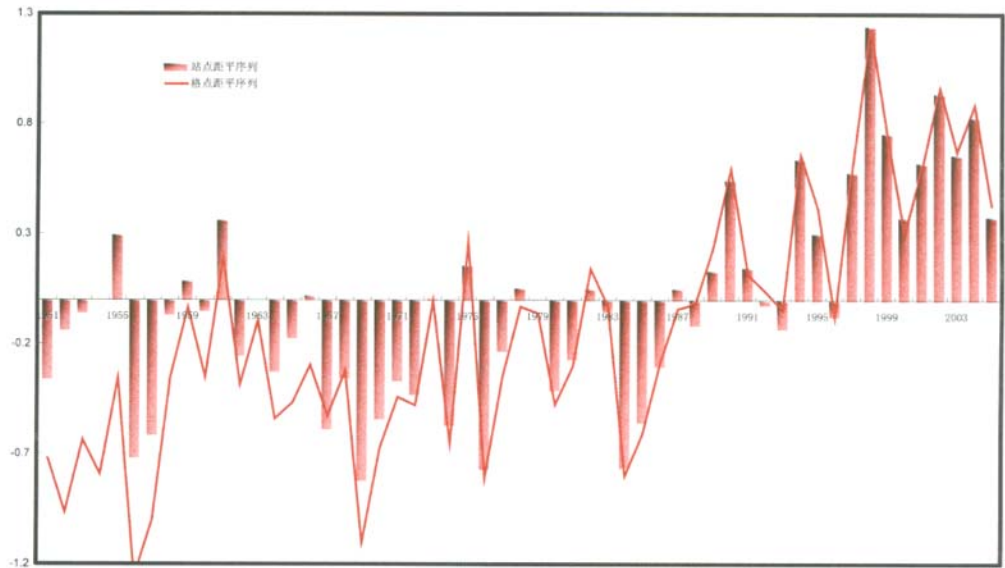


图 7 1951 - 2005 年中国东部年平均气温变化(℃)

理后产生的误差。本文中,截断误差主要来自克里金计算。

4 应用前景

在对要素时空分布的描述中,格点资料相比站点资料具有明显的优势。随着气象科学数据共享项目的纵深开展,常规气象要素的站点资料已不能满足社会和广大科研工作者的需求。特别是近几年来,随着我国自主大气、气候模式的开发和业务化运行,模式初始化及验证对高精度、高时空分辨率的格点数据的需求也日益凸显。深入开展气象要素场的空间分析,研制具有理想精度的中国区域网格气象要素数据集,是满足当前用户需求、提高气象资料共享服务能力和技术含量的重要途径。本数据集的建立可以很好地揭示和展现气温的时空变化规律,为更精细的气候变化研究提供参考,并为各种气候模式的运行提供数据支持。

参考文献

- [1] Jones P D, Raper R, Bradley R S, et al. Northern hemisphere surface air temperature variations: 1851–1984[J]. J Climate Appl Meteor, 1986, 25(2): 161–179.
- [2] Hanson J and Lebedeff S. Global surface air temperature: update through 1987[J]. Geophys Res Lett, 1988, 15(4): 323–326.
- [3] Vose R S, Schmoyer R L, Steurer P M, et al. The Global Historical Climatology Network: Long-term monthly temperature, Precipitation, sea-level pressure, and station pressure data[EB/OL]. [2008-06-20]. <http://cdiac.esd.ornl.gov/epubs/ndp/ndp041/ndp041.html>.
- [4] Peterson T C, Vose R S. An Overview of the global historical climatology network temperature database[J]. Bulletin of the American Meteorological Society, 1997, 78(12): 2837–2849.
- [5] Jones P D. Hemispheric surface air temperature variations: A reanalysis and an update to 1993[J]. J Climate, 1994, 7(11): 1794–1802.
- [6] Hulme M. A 1951–80 global land precipitation climatology for the evaluation of general Circulation Models[J]. Climate Dyn, 1992, 7(1): 57–72.
- [7] Huffman G J, Adler R F, Rudolf B R, et al. The Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset[J]. Bull Amer Meteor Soc, 1997, 78(1): 5–20.
- [8] Xie P and Arkin P A. Global Precipitation: A 17-year Monthly Analysis Based on Gauge Observations, Satellite Estimates and Numerical Model Outputs[J]. Bull Amer Meteor Soc, 1997, 78(11): 2539–2558.
- [9] Adler R F, Kidd C, Petty G, et al. Intercomparison of Global Precipitation Products: The Third Precipitation Intercomparison Project (PIP-3) [J]. Bull Amer Meteor Soc, 2001, 82(7): 1377–1396.
- [10] 林忠辉, 莫兴国, 李宏轩, 等. 中国陆地地区气象要素的空间插值[J]. 地理学报, 2002, 57(1): 47–56.
- [11] 冯锦明, 赵天保, 张英娟. 基于台站降水资料对不同空间内插方法的比较[J]. 气候与环境研究, 2004, 9(2): 261–277.
- [12] 刘峰. 应用 Kriging 算法实现气象资料空间内插[J]. 气象科技, 2004, 32(2): 110–115.
- [13] 苏志侠, 程麟生. 两种客观分析方法的比较: 逐步订正和最优内插[J]. 高原气象, 1994, 13(2): 194–205.
- [14] 李新, 等. 青藏高原原气温分布的空间插值方法比较[J]. 高原气象, 2003, 22(6): 565–573.
- [15] 封志明, 杨艳昭, 丁晓强, 等. 气象要素空间插值方法优化[J]. 地理研究, 2004, 23(3): 357–364.
- [16] 朱芮芮, 李兰, 王浩, 等. 降水量的空间变异性和空间插值方法的比较研究[J]. 中国农村水利水电, 2004(7): 25–28.
- [17] 庄立伟, 王石立. 东北地区逐日气温要素的空间插值方法应用研究[J]. 应用气象学报, 2003, 14(5): 605–615.
- [18] 魏凤英, 曹鸿兴. 我国月降水和气温网格点资料的处理和分析[J]. 气象, 1994, 20(10): 26–30.
- [19] 李新, 程国栋, 卢玲. 空间内插方法比较[J]. 地球科学进展, 2000, 15(3): 260–265.
- [20] Lennon J J, Turner J R. Predicting the spatial distribution of climate: temperature in great Britain[J]. J Anim Ecol, 1995, 64: 370–392.
- [21] Nalder I A, Wein R W. Spatial interpolation of climate normals: test of a new method in the Canadian boreal forest[J]. Agric For Meteorol, 1998, 92: 211–225.
- [22] 中国气象局监测网络司译. 气象仪器和观测方法指南(第六版)[M]. 北京: 气象出版社, 2005.
- [23] 李志林, 朱庆. 数字高程模型[M]. 武汉: 武汉大学出版社, 2003.
- [24] 张仁铎. 空间变异理论及应用[M]. 北京: 科学出版社, 2005.