

资料插值的进展^①

蔡秀华 曹鸿兴

(中国气象科学研究院, 北京 100081)

提 要

对气候学中资料插值、数值天气预报中的客观分析和气候模拟中的降尺度技术作了统一论述,重点是气候资料插值的原理和进展。由于应用领域不同,这些技术也是很不相同的,但它们的动力学和统计学原理是相近或一致的。近十年来三种插值技术,尤其在建立计算机自动化插值系统方面,都取得了长足进展。

关键词: 资料插值 数据同化 克立格法

引 言

在大气科学中资料插值一直受到重视,这是因为许多情况下都需要运用这一技术。(1)插补因灾难、战争等造成的观测缺失;(2)重建、延长代用资料,如利用清宫晴雨录重建北京历史降水资料;(3)把不规则测站上的记录插到网格点上或反之;(4)把粗网格上的模拟值插到细网格上,即所谓降尺度下插(downscaling)。无论在天气分析和预报中,还是在气候诊断、模拟和预测中乃至应用气候、大气环境中,都使用着资料插值技术。大气科学中资料插值从两个不同方向发展着:即在气候学中的资料插值和天气预报中的客观分析。

早期的气候资料插值都采用比较简单的方法,如比例法、加权平均、线性回归等^[1],1980年代以后,考虑气候场的信息以及气候场的插补,现代统计技术应用到资料插值中,如经验正交函数、傅里叶谐波函数、小波函数、样条函数、车贝雪夫多项式等。考虑观测资料在时间上连续的信息,时间序列分析也被应用到资料插值中,考虑气象序列由周期

分量迭加而成的均生函数模型(MGF)也被研究过,结果表明用MGF来进行资料插值有相当好的效果。

过去的天气图是手工绘制的,随着数值天气预报的发展,提出并发展了客观分析技术。其实质就是把不规则测站上的记录插到网格点上以及当作出数值预报后把预报值插回站点上。有名的Cressman逐步订正法是一种以距离作为权重的插值法,Gandin最优内插法则以随机函数理论为基础的,后来加进了动力约束,使之更符合气象场的要求。变分同化最早由Sasaki提出^[2],到1990年代进入数值预报业务系统,现在在先进的数值天气预报模式和气候预测模式中毫无例外地都使用资料同化技术,其中包括变分同化和卡尔曼滤波。变分同化是给定一个目标函数,使之在一定约束条件下达到最小(最大)。变分同化除了把不同时次、不同仪器的不规则网格点上观测值插值到规则网格点上以外,还需使初始值与所用模式相协调,即所谓初始化,其计算量与用模式作预报达同一量级,耗费计算机资源很大。

^① 由国家自然科学基金(40375025)和北京自然科学基金(8042012)资助。

NCEP/NCAR 再分析资料的出现标志着资料插值的上述两个方向的合流,而降尺度技术的研究又进一步推动了资料插值的发展^[3,4]。资料再分析的原理是在全球天气预报模式的资料同化中使用实际观测值以再建一个动力相容的大气历史分析,通过同化处理资料中的错误和误差被消除和订正了^[5]。资料再分析是资料处理技术的一个重大进展。

针对不同变量、不同应用目的,研制了各种资料插值的计算机自动化系统,如针对海表温度(SST)的自动插值系统、与气候模拟相匹配的天气随机发生器、降尺度下插技术的软件等。澳大利亚根据地面观测资料建立了一个有降水、温度、蒸发和日照等要素的综合气候资料库,降水资料始自 1890 年,其他要素始自 1957 年,网格为 $0.05^\circ \times 0.05^\circ$ 。日和月降水量用随机过程方法内插,其他日变量用样条函数内插。该系统与因特网相联,通过其网站可获得该库中的资料。

现在的资料插值同时考虑观测资料的空间信息、时间信息以及模式提供的信息,使得观测资料尽可能逼近真实。本文重点叙述气候资料插值的若干原理和进展。

1 空间信息

1.1 距离权重法

大气状态在空间上是连续分布的,各种不同尺度的天气系统有序地配置在空间上,因此,首先要利用的就是气象要素的空间信息。距离权重法是常用的插值法。例如距离权重反比法,其表达式为

$$z = \frac{\sum_{i=1}^n (z_i \cdot W_i)}{\sum_{i=1}^n W_i} \quad (1)$$

式中, n 为用于插值的气象观测站点的数目, z 为估计的格点的气象要素值, z_i 为气象要素在第 i 个站点的实测值, W_i 为第 i 个站点的权重。

距离权重反比法以插值点与样本点间的距离为权重进行加权平均,离插值点越近的

样本点赋予的权重越大。若以格点到观测点的大圆半径的平方反比作为权重系数,称为距离平方反比法。对象降水这样局地性很强的要素,可在距离权重的基础上,进一步考虑气象要素随海拔高度的变化。

Cressman 逐步订正法也是一种距离权重法,其权重采用圆型权重函数,设 R 为观测站到网格点之间的距离, R_1 为搜索半径,则权重函数为

$$w_i = (R_1^2 - R^2)/(R_1^2 + R^2) \quad (2)$$

距离权重法的优点是简单直观,缺点是对要素的固有特征缺少考虑。

1.2 克里格(Kriging)法

克里格插值法来源于地统计学。地统计学由法国统计学家 G.Matheron 创立^[6],它的应用范围十分广泛。地统计学已在地球物理、地质、生态、土壤等领域应用。气象领域的应用目前还不多见,主要使用 Kriging 法进行降水、温度等要素的最优内插的研究^[7-9]及农业生态方面的研究^[10]。

地统计学研究的变量是区域化变量,所谓区域化变量就是由在某一区域或范围内的空间位置所取的不同数值构成的变量,它是空间坐标为自变量的随机变量及其随机函数。区域化变量不需要多次重复观测的样本,是只需取值一次的变量。地统计学中区域化变量是在空间不同位置上取样,两相邻样本值不一定相互独立,而是具有某种程度的空间相关性。地统计学除了使用平均值、标准差、相关系数等数学特征量外,更重要的是使用变异函数及其参数来研究区域化变量的空间分布特征。

变异函数是描述区域化变量结构性和随机性特有的基本手段。设区域化变量 $z(x_i)$ 和 $z(x_i + h)$ 分别是 $z(x)$ 在空间位置 x_i 和 $x_i + h$ 上的观测值($i = 1, 2, \dots, N(h)$), 变异函数由下式进行估计:

$$r(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (3)$$

这里 $N(h)$ 是分隔距离为 h 的样本量。变异函数是在假设 $z(x)$ 为区域化变量且满足平稳条件和本征假设的前提下定义的。数学上可以证明,变异函数大时,空间相关性减弱,反之亦然。

以 h 为横坐标,以 $r(h)$ 为纵坐标,绘出变异函数曲线图,这些图可以直观地展示区域化变量 $z(x)$ 的空间变异性^[11]。

对任一空间变量点 x 处的克里格插值 z_x^* ,可以通过对该点影响范围内的 n 个有效观测值 $z(x_i)$ 的线性组合得到,即

$$z_x^* = \sum_i^n w_i z(x_i) \quad i = 1, \dots, n \quad (4)$$

式中, w_i 是赋予气象观测值 $z(x_i)$ 的权重,表示各站点气象要素值 $z(x_i)$ 对估计值 z_x^* 的贡献。为达到线性无偏估计,使估计方差最小,权重由普通或简单克里格 (Ordinary/Simple Kriging) 方程组求得。同时,权重取决于变量的空间结构性,而任何变量的空间结构由变异函数表示,研究表明对于日最高、最低温度的逐日插值采用幂函数模型、空洞效应模型较合适;而对于日降水量的插值,可选择幂函数模型^[12]。

Gandin 最优内插法是把气象要素随空间的变化视为一个随机函数,而基于地统计学的克里格插值法是把气象要素分解为结构项和随机项两部分,无疑是对所研究问题的深化;并由此发展出新的资料插值技术。当今,国际上若干天气预报中心和气候研究所正在尝试用克里格法来创建新的资料插值的计算机系统。

气象中许多灾害性天气气候现象很难获取大量重复观测样本,就难以应用基于大量重复观测样本的普通统计学,例如,某一时段某一区域的高温现象、沙尘暴、强暴雨过程等等,就属于这类问题,可以用地统计学进行研究。

庄立伟、王石立^[13] 针对作物生长动力模型区域应用中需要输入逐日格点气象要素值

的问题,以东北地区为例,选用 Kriging 法、以经纬度分布方向为权重的距离权重反比法 (IDW) 及带高度梯度订正的距离权重反比法 (GIDW) 三种插值方法对 4~10 月逐日气象要素插值法进行了对比研究。国内外研究表明对不同气象要素要使用适合的插值法,降水有短的空间相关尺度和大的变率,而温度则有长的空间相关尺度,其变率也较小,相应地,它们的插值法理应不同^[14]。

2 时间信息

气象要素随时间的演变具有连续性,因此需从时间方向上抽取信息进行插值。

2.1 时间序列分析

时间序列分析认为观测序列是系统运动的轨迹,时间序列中蕴藏着丰富的信息,它既刻划了系统本身的状态变化,又是外环境影响的结果。时间序列分析常用的有 AR(自回归)、MA(滑动平均)和 ARMA(自回归滑动平均)模型等。用在资料插值中主要是 AR 模型。如缺测 t_i 时次的 x_i 值,可以先给定一个初估值,然后建 AR 模型,由模型算得 x_i 值,再建模……,直到前后两次插值之差小于给定的允许误差。AR、ARMA 模型的缺点是所插之值易偏向均值。

2.2 人工神经网络

1990 年代以来人工神经网络 (ANN) 在气象中受到愈益广泛的应用,在美国、日本、欧洲,人工神经网络应用在数值天气预报的释用中,即根据数值天气预报产品运用人工神经网络来制作专业预报,即用户需要的预报。ANN 模型中的参数是通过观测序列的训练而获得的,因此无疑可以应用 ANN 来作资料插值和气象预报。至今,在气象中主要应用回传网络 (BP), 由于它可以设置中间层,模型的可塑性大,因此 BP 网络往往对历史资料拟合和试报的效果 (x_f) 很好^[15,16], 但当用于实际预报时效果 (x_r) 往往会明显下降,两者之差记为

$$\Delta x = x_f - x_r$$

这个差通常比其他预报方法的差要大,我们不妨称此为 ANN 问题。至今尚无人对 ANN 问题进行研究。除 BP 网络外,象模糊(fuzzy)神经网络、径向基神经网络等也应用到气象中了,人工神经网络气象应用的前景是广阔的,但待解决的问题也不少。

2.3 动态数据自忆模型

引进记忆函数 $\beta(r, t)$, 对大气运动方程进行积分,运用微积分中的分部积分和中值定理可求得一个差分-积分方程^[17,18]

$$\beta(r, t)x(t) - \beta(r, t-p)x(t-p) - \sum_{i=-p}^0 (\beta(r, t_{i+1}) - \beta(r, t_i))x^m(t_i) - \int_{t-p}^t \beta(r, \tau)F(x, \lambda, \tau)d\tau = 0 \quad (6)$$

式中 x 为物理量, λ 为参数, F 为源函数或空间项, p 为回溯阶,初始时次为 t_0 。称式(6)为描述系统运动的自忆性方程。一旦用历史观测资料确定了记忆函数 β ,就可用式(6)来进行积分。称以差分-积分方程(6)为基础构建的模式为自记忆模式。取等间隔采样 Δt , 即 $t_i = t_0 + i\Delta t$, $i = 1, 0, -1, -2, -p$, t_0 为初值; $t_0 + \Delta t$ 为预报时次。式(6)中的中值取为前后时次值的平均,积分由求和替代,则待求时次 t_i 的插值为

$$x(t_i) = x(t_1)/\alpha_i - [\alpha_{-p}x(t_{-p}) + \dots + \alpha_0 x(t_0) + \sum_{i=-p}^0 \theta_i F(t_0 + i\Delta t)\Delta t]/\alpha_i \quad (7)$$

式中源函数可取自系统运动的微分方程,也可由量测序列反导出一个微分方程^[19]。当取为大气运动方程时,由此可发展出一种新的四维同化格式。动态数据自忆模型是一种有潜力的插值技术。

3 时空信息

同时提取观测值的时空信息,无疑会优化插值,尽管这需要更多计算量。通常是把气象场用正交函数如 EOF(经验正交函数)或谐波函数展开,再对其分量进行插值^[20]。

4 模式信息

在数值天气预报的变分同化中^[20],目标函数中除含观测场项外还包含初估场项,后者即是由模式提供的信息。模式信息对于资料稀少地区如高山、极地、海上等是极为重要的,是对观测资料的一个补充。在降尺度技术中,则同时需要关于某一区域的模拟值和观测值,而全球模式的模拟值是工作的前提。当今,气候模式已能相当好地模拟出现在气候,因此模式信息是不可忽视的用于插值的源信息。动力降尺度技术是把区域气候模式嵌套在全球模式中,以细化模拟值。这时在运行模式中需细网格上的观测值和模拟值,故必须使用资料插值技术。而在统计降尺度中则需建立粗网格和细网格之间的关联模型,实际上是一种统计插值技术。常用的统计方法有多元分析、典型相关分析等^[21]。

5 检验评估

对于插值效果的检验,一般采用交叉验证(cross-validation)来检验其插值的效果。假设每一站点的气象要素值未知,用周围站点的值来估算,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的优劣。可采用平均绝对误差(MAE)和均方根误差(RMSE)来度量插值效果。另一种检验方法是使用样式识别(pattern recognition)技术,即考察由观测值绘制的等值线图与插值绘制的图形一致的程度^[22]。

在大气科学中,多个领域需要进行资料插值,它们的目的、前提和要求各不相同,使用的数理原理也不同,从近似计算、微分方程到概率统计,从人工智能、模糊集理论到控制论;因此发展针对问题的插值法、比较其优劣是今后的这些领域的重要研究课题。

参考文献

- 1 么枕生,丁裕国.气候统计.北京:气象出版社,1993:779~825.
- 2 张玉玲,吴辉旋,王晓林.数值天气预报.北京:科学出版社,1987:380~420.

- 3 Murphy J. M. , An Evaluation of Statistical and Dynamical Techniques for Downscaling Local Climate. *J Climate*, 1999,12:2256—2284.
- 4 Huth R. Uncertainty in the Climate Change Estimates by Statistical Downscaling. *International Conference on Earth System Modeling*, 15—19 Sept. 2003, Hamburg, Germany.
- 5 Kalnay E. Kanamitsu M. , Kistler R. et al. , NCEP/NCAR Reanalysis. *Bull. Am. Met. Soc.* , 1996,77:437—471.
- 6 Matheron G. , Principles of Geostatistics, *Economic Geology*, 1963,58:1246—1266.
- 7 Prudhomme, C. and Reed, D. W. , Relationships between Extreme Daily Precipitation and Topography in a Mountainous Region. A Case Study in Scotland, *Int. J. Climatol.* , 1998,18:1439—1453.
- 8 Courault, D. and Monestiez, P. , Spatial Interpolation of Air Temperature According to Atmospheric Circulation Patterns in Southeast France, *J. Climatol.* , 1999, 19: 365—378.
- 9 Prudhomme, C and Reed, D. W. , Mapping Extreme Rainfall in a Mountainous Region using Geostatistical Techniques: a Case Study in Scotland, *Int. J. Climatol.* , 1999,19: 1337—1356.
- 10 李海滨, 林忠辉, 刘苏峡. Kriging 方法在区域土壤水分估值中的应用. *地理研究*, 2001,20(4):446~452.
- 11 魏凤英, 曹鸿兴, 徐祥德. 变异函数在降水场空间特征分析中的应用, *南京气象学院学报*, 2002,23(6): 795~799.
- 12 Kerry K. and Hawick K. Kriging Interpolation on High Performance Computer. *Proc. of High Performance Computing and Networks, HPCN98*, April 1998, DH-PC-35, Australia.
- 13 庄立伟, 王石立. 东北地区逐日气象要素的空间插值方法应用研究. *应用气象学报*, 2003,14(5): 605~615.
- 14 Shen S. S. P. , Dzikowski P. , Guilong Li and Griffith D. Interpolation of 1961-97 daily Temperature and Precipitation Data onto Alberta Polygons of Eco-district and Soil Landscapes of Canada. *J Appl. Meteor.* , 2001,40(12): 2162—2177.
- 15 Xiaodan Wu, Hong-Xing Cao, Andrew Flitman et al. Forecasting Monsoon Precipitation using Artificial Neural Networks. *Adv. Atmos. Sci.* , 2001,18(5):950—958.
- 16 Jin Long, Luo Ying, Lin Zhenshan. Comparison of Long-term Forecasting of June-August Rainfall over Changjiang-haihe Valley. *Adv. Atmos. Sci.* , 1997, 14 (10):87—92.
- 17 Cao Hong-Xing. Self-memorization Equation in Atmospheric Motion. *Science in China, Series B*, 1993, 36 (7):845—855.
- 18 曹鸿兴. 动力系统自忆性原理—预报与计算应用. 北京:地质出版社, 2002:22~31.
- 19 贾晓静, 曹鸿兴, 封国林. 一种动态数据的新建模法及其预报应用. *应用气象学报*, 2002,13(1): 96~101.
- 20 王耀山. 客观分析与四维同化(II). *气象科技*, 2001,29 (3):1~11.
- 21 Houghton J. T. , Ding Y. Griggs D. J. et al. (editors). *Climate Change 2001*. Cambridge University Press, 2001:616~623.
- 22 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. *气象*, 1994,20(10): 26~30.

Advance of Data Interpolation

Cai Xiuhua Cao Hongxing

(Chinese Academy of Meteorological Sciences, Beijing 100081)

Abstract

A unified briefing of data interpolation of climatic data, objective analysis of numerical weather forecast and downscaling technique of climate modeling is presented. However, it is an emphasis on the principle and advance of climatic data interpolation. Because of the distinctness of its application fields, these techniques are also different, but their dynamic and statistical principles are close or nearly the same. In recent ten years, three kinds of data interpolation technique, especially on the automatic computer system of data interpolation, have made progress greatly.

Key Words: data interpolation data assimilation Griging method