

地统计学分析技术及其在气象中的适用性

魏凤英 曹鸿兴

(中国气象科学研究院, 北京 100081)

提 要

首先论述了地统计学的一些特点, 然后从变异函数和 Kriging 插值两个方面介绍了地统计学的分析技术, 最后讨论了地统计学在气象研究中的适用性。

关键词: 地统计学 变异函数 Kriging 插值

引 言

地统计学是以具有空间分布特点的区域化变量理论为基础, 研究自然现象的空间变异与空间结构的一门学科^[1]。它针对像矿产、资源、生物群落、地貌等有着特定的地域分布特征而发展的统计学。由于最先在地学领域应用, 故称为地统计学^[2]。地统计学的主要理论是法国统计学家 G. Matheron 创立的, 经过不断完善和改进, 目前已成为具有坚实理论基础和实用价值的数学工具。地统计学的应用范围十分广泛, 不仅可以研究空间分布数据的结构性与随机性、空间相关性与依赖性、空间格局与变异, 还可以对空间数据进行最优无偏内插, 以及模拟空间数据的离散性及波动性。地统计学由分析空间变异与结构的变异函数及其参数和空间局部估计的 Kriging 插值法两个主要部分组成, 目前已在地球物理、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^[3~5]及气候对农业影响方面的研究^[6]。

本文从变异函数及其参数和 Kriging 空间局部插值两个方面概要介绍地统计学, 讨论地统计学在气象研究中的适用性。

1 地统计学的特点

(1) 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的空间位置所取的不同数值构成的变量, 它是

以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, 只需取值一次的变量。

(2) 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。

(3) 地统计学除了使用平均值、标准差、相关系数等数字特征量外, 更重要的是使用变异函数及其参数研究区域化变量的空间分布特征。

由上述可见, 地统计学提供了有效分析和解释空间数据的变化规律和格局的数学工具^[7]。地统计学的内容主要包括: 定量描述和解释空间变异性和空间相关特性; 建立空间预测模型; 对空间格局的尺度、形状、变异方向进行估计, 并将其与研究要素的过程联系起来, 给予物理解释; 空间局部插值、模拟空间数据的离散性和波动性等等。

2 变异函数及其参数

2.1 区域化变量

设以空间点的三个直角坐标 x_u, x_v, x_w 为自变量的随机场 $Z(x_u, x_v, x_w) = Z(x)$ 是一个区域化变量, 它是普通随机变量在某一区域内确定位置上的特定取值, 是与位置有关的随机函数, 进行一次观测就可以得到区域化变量 $Z(x)$ 。

假设区域化变量 Z 具有随机性和结构

性,那么,在点 x 处的观测值 $Z(x)$ 可以表示为:

$$Z(x) = \mu + \epsilon'(x) + \epsilon \quad (1)$$

μ 是区域化变量 Z 的平均值; $\epsilon'(x)$ 是空间相关误差,平均值为 0; ϵ 是通常统计学中随机误差。式(1)表示区域化变量的空间结构,其变异由 $\epsilon'(x)$ 和 ϵ 两部分构成。变量 Z 的空间变异性可以分解为两部分,即

$$SH(Z) = SH_C + SH_R \quad (2)$$

其中 SH_C 表示 $\epsilon'(x)$ 的相关变异; SH_R 表示 ϵ 的随机变异。 SH_C 和 SH_R 可以通过变异函数来分解并量化。

2.2 变异函数

变异函数是描述区域化变量随机性和结构性特有的基本手段。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。

对于二维区域化变量,根据网格点数据,用式(3)分别计算南北、东西、西北—东南及东北—西南 4 个方向或更多方向上的变异函数。在计算西北—东南及东北—西南方向上的变异函数时,分隔距离取对角线的距离 $\sqrt{2}h$ 。图 1 是计算 4 个主要方向变异函数样本数据对的构成示意图。

2.3 变异函数的参数

二维区域化变量的变异函数不仅与分隔距离 h 有关,也与方向有关。设 $r(h, \theta_1)$ 代表区域化变量一个方向的变异函数, $r(h, \theta_2)$ 代表该区域化变量另一个方向的变异函

数,两者的比值

$$k(h) = r(h, \theta_1) / r(h, \theta_2) \quad (4)$$

等于 1 或接近 1 时则表明空间变异性为各向同性,否则为各向异性。比值越低各向异性越高,表明空间变异性程度越大。

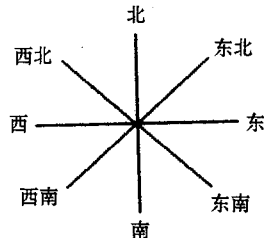


图 1 4 个主要方向样本数据对构成示意图

通过变异函数还可以得到以下几个表征空间变异特征的重要参数:

I. 基台值(still): 变异函数 $r(h)$ 随 h 增大, 从非零值达到一个相对稳定值的常数时, 这个常数称基台值 $C_0 + C$ 。

II. 块金值(nugget): 当 $h = 0$ 时, $r(0) = C_0$ 时, C_0 称为块金值。它表示区域化变量在小于观测的尺度时的非连续变异。

III. 变程(range): 变异函数达到基台值时的分隔距离 a 称为变程。

IV. 分维数(fractal dimension): 它由变异函数 $r(h)$ 和 h 确定

$$2r(h) = h^{(4-2D)}$$

分维数 D 值大小表示变异函数曲线的曲率, 是随机变率的量度。

3 Kriging 空间局部插值

Kriging 空间局部插值是地统计学的重要内容之一, 它是建立在变异函数理论分析基础上的, 是对有限区域内的区域化变量取值进行无偏最优估计的一种方法。这种方法与传统插值方法不同的是, 在估计无观测样本数值时, 不仅考虑待插值站点与邻近有观测数据站点的空间位置, 还考虑了各邻近站点之间位置的关系, 在除了考虑空间位置, 还利用已有观测值的空间分布的结构特点, 使其估计比传统方法更精确, 更符合实际, 并可以有效避免系统误差产生的“屏蔽效应”。经过统计学家们的大量研究, 发展了多种空间局部插值方法, 下面是常用的两种。

3.1 普通 Kriging 方法

普通(ordinary)Kriging 是地统计学中最常用的插值方法。对于任意待估计站点的估计值 $Z^*(x_0)$, 均可以通过待估计站范围内的 n 个观测样本值 $Z(x_i)$ ($i = 1, 2, \dots, n$) 的线性组合得到, 即

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (5)$$

式中的 λ_i 是权重系数, 其和等于 1, $Z(x_i)$ 为测站观测值, 它们位于区域内 x_i 位置。为了达到线性无偏估计, 使估计方差最小, 权重系数由 Kriging 方程组得到。普通 Kriging 方程组为:

$$\begin{cases} \sum_{i=1}^n \lambda_i C(x_i, x_j) - u = C(x_i, x_0) \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad (6)$$

式中 $C(x_i, x_j)$ 为样本间协方差函数, $C(x_i, x_0)$ 为样本点与待插值点间的协方差函数, u 为极小化处理时的拉格朗日乘子。

Kriging 插值的权重取决于变量的空间结构, 它可以由变异函数来表征。

3.2 协同 Kriging 方法

协同 Kriging (Cokriging) 插值原理与普通 Kriging 相同, 但是它考虑了所插值的要素与其它变量之间的关系。例如, 在进行气温、降水变量插值时, 考虑海拔高度的作用。

协同 Kriging 插值过程比普通 Kriging 插值要复杂些, 需要引入一个新的假定, 即两个变量之间差值的方差最小。另外, 还要引入交叉变异函数, 即两个不同变量之间的相关随距离变化的函数。

4 地统计学在气象研究中的适用性

早在 20 世纪 60 年代, 前苏联学者就用普通统计研究空间气象场, 提出了结构函数的概念和理论。在此基础上, Gandin 发展了著名的最优内插法, 这一方法至今仍在数值天气预报中使用。利用地统计学 Kriging 方法进行空间插值, 考虑了结构性与随机性的特征, 因此, 移植地统计学的理论与方法将为气象场空间分布型态和插值研究提供一个新的基础。Prudhomme 等人^[3,5] 利用地统计学 Kriging 方法对山区极端降水指数和年最大

日降水进行了插值, 并绘制了图。Courault 等人则利用地统计学技术按照大气环流型对法国东南部的气温做了空间插值, 取得了比较理想的效果。

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, 而无法使用基于大量重复观测样本的普通统计学分析, 例如, 某一时段某一区域的高温天气、沙尘暴、强暴雨过程等等, 就属于这类问题, 可以尝试用地统计学进行研究。魏凤英等利用地统计学中的变异函数及其参数研究了 1998 年 7 月下旬长江流域强暴雨过程的空间变异特征^[8]。

大气环流遥相关型(如 PNA, WP 等)研究揭示了气象要素大范围的空间结构特征, 实际上这种特征与地统计学中结构性是一致的, 但遥相关型是用计算点相关系数来获得的, 因此, 可以预期, 如果运用地统计学可以揭露出新的大气空间场结构特征。

地统计学的显著特点是, 区域化变量在空间不同位置上取样, 要求具有某种程度的空间相关性。气象中研究的大多数变量场一般都具有一定程度上的相关。在地统计学中变量的空间变异可以分解为相关变异和随机变异两部分, 这两部分通过变异函数来分解并定量化。通过对两部分的对比分析, 了解气象要素场的空间相关性与随机性的相互关系, 从而获得空间结构方面的信息, 认识天气气候系统的潜在空间特征。

参考文献

- 1 Matheron G., Principles of geostatistics, Economic Geology, 1963, 58:1246—1266.
- 2 Joarnel, A. G. and Huijbregts, C. J., Mining Geostatistics, Academic Press, London, 1978.
- 3 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.
- 4 Courault, D. and Monestiez, P., Spatial interpolation of air temperature according to atmospheric circulation patterns in southeast France, J. Climatol., 1999, 19:365—378.
- 5 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.

(下转第 23 页)

(上接第 5 页)

- 6 James W. Hansen and James W. Jones, Scaling-up crop model for climate prediction applications, Climatic Prediction and Agriculture, International START Secretariat, Washiton, DC, USA ,2000.
- 7 王政权. 地统计学及在生态学中的应用, 北京: 科学出版社, 1999, 1~195.
- 8 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 年待发表.

Geostatistics and Its Application to Meteorological Studies

Wei Fengying Cao Hongxing

(Chinese Academy of Meteorological Sciences, Beijing 100081)

Abstract

The characteristics of geostatistics are described. The variograms technique and Kriging method for spatial interpolation and their application to meteorological studies are discussed.

Key Words: Geostatistics Variograms Kriging spatial interpolation