

文章编号: 1000-1190(2007)03-0477-03

克立格估值方法中权系数空间相关性研究

贺华中^{1,2}, 代侦勇^{2*}

(1. 贵州民族学院 化学与环境学院, 贵阳 550025; 2. 武汉大学 资源与环境学院, 武汉 430079)

摘 要: 论文从 4 个方面论述了克立格估值中权系数的规格化与区域化变量的空间相关性; 克立格方程组的无偏估计的约束条件; 空间系数缺失情况下的权系数规格化方法; 数据组合中的约束规则; 权系数为负值时的规格化方法. 重点结合大气环境监测站点的优化设计从理论和实际上进行了论证, 给出了权系数为负值时的一种规格化方法.

关键词: 克立格; 权系数; 空间相关性; 规格化

中图分类号: P208

文献标识码: A

克立格估计方法得以在地质、地理、环境、生态等方面广泛应用的原因, 在于它是一种线性、无偏估计, 而且可以给出估计方差. 克立格方法的这一优点是通过克立格空间变异函数和克立格方程组得以体现的, 其中重要的一点就是克立格方程组中的权系数 λ_i , 其总和为 1 ($\sum \lambda_i = 1$), 不仅是保证克立格估计是无偏最优估计的前提条件, 其大小和空间分布也很好地反映了区域化变量的空间相关性[1]. 对于克立格方法中区域化变量建立的变异函数理论模型及所反映的区域化现象, 如权系数的大小及其空间分布等, 只有在专业背景下才能得到最充分的解释 (Rossi 等 1992), 克立格法的特色也就在于此. 本文从克立格方程出发, 结合克立格方法在大气环境监测站点优化中的权系数的空间相关性展开论述, 重点讨论了权系数为负值的现象.

1 普通克立格方程组中权系数的空间相关性

普通克里格法进行估计所使用的线性估计量为:

$$Z_v^* = \sum_{i=1}^n \lambda_i Z_i.$$

其中, $Z_i (i=1, 2, \dots, n)$ 为研究区域的实测值, Z_v^* 为估计值, λ_i 为测点 i 的权值.

上式是 n 个样品值的线性组合, 通过 Lagrange 乘数法求解和相应变换, 可以得到一个包括 $n+1$ 个未知数的线性方程组 (具体推求过程参

见文献[2]).

$$[K] \cdot [\lambda] = [M].$$

上式称之为克立格方程组. 其中 $[K]$ 和 $[M]$ 分别为测点之间和待估点与测点之间的相关矩阵. 上述方程组中 $[\lambda]$ 的求解表达式为:

$$[\lambda] = [K]^{-1} [M].$$

克立格方程的推导和求解从两个方面反映了区域化变量的空间相关性: (1) 使克立格估计成为无偏估计的规格化条件 $\sum \lambda_i = 1$; (2) 在 $[\lambda] = [K]^{-1} [M]$ 的右端表达式中, $[K]^{-1}$ 和 $[M]$ 是与方差函数或半变异函数模型有关的一个加权表达式. 这种加权类似于距离反比法, 但不是简单的几何距离加权或统计距离加权^[1], 表达的是空间测点之间与距离有关的空间相关性大小, 既反映了待估点与实测点之间的相关性, 也反映了待估点之间的相关性. 并且这种相关性的矩阵运算, 始终保证 $\sum \lambda_i = 1$, 这正是克立格法有别于其他统计法的特点之一. 对这种相关性权且称之为克立格空间相关性.

2 两种情形下的权系数规格化方法反映了区域化变量的空间相关性

2.1 数据缺失情况下的权系数规格化方法

最早的地质统计要求数据的空间构形规则, 数据完整, 但实际勘测中, 总会有个别网格的数据缺失. 对缺失数据的处理方法之一是用邻域内数据按算术平均法填补. 这种用邻域数据填补的方法计算

收稿日期: 2007-01-27.

基金项目: 国家自然科学基金项目 (40676094).

* 通讯联系人. E-mail: daiwhu@263.net.

的估计方差 σ^2 并没有增大,但可能会降低估值的精度^[3];另一种方法是先建立 kriging 方程组求解权系数 $[\lambda]$,若其中第 k 个估计点的数据缺失,则把第 k 个估计点及其权系数 λ_k 一同去掉,对剩下的权系数重新归一获得一个新的系数矩阵 $[\lambda']$,用这个新的系数矩阵代替 $[\lambda]$,进行各种计算.据研究,这种经过规格化的“随机”数据构形求得的估计方差与普通的数据构形的估计方差两者相差不到 1%^[3],仍满足估值的精度要求,且可节省计算时间.

2.2 数据组合中的权系数的约束

为了约简克立格方程组的维数,可将有效数据按一定原则加以组合,如按角度组和距离组进行数据组合^[3-4].

此外,可以将待估块段(点)在结构上对称的数据组合在一起,而这样组合的理论依据就是,以某一直线为全对称的块段中,若待估块段(点)在此直线上,则两两对称的点具有相同的权系数 $\lambda_i = \lambda_j$ ($i \neq j$),这就是克立格估值中的“域对称权相等”的原理.

但要注意,以上几种情形中,都必须满足 $\sum \lambda_i = 1$. 因此从这个意义上说, $\sum \lambda_i = 1$ 也是一种规格化约束,只是其被习惯地称之为归一化.

3 大气环境监测站点中克立格权系数为负值的规格化方法

权系数的规格化处理方法虽然较好地处理了数据的缺失和数据组合问题,但对权系数为负的问题仍无法解决.

对于克立格线性方程组,一般总有解^[2],但从线性代数角度求解出的 $[\lambda]$,时常会出现负值.对此文献^[3]认为,这样的点主要出现在待估域的最外层,是由于内层数据的屏蔽作用,使之变小甚至小于 0. 解决的方法是将其摒弃,进一步约简克立格方程,仍可保持其精度.文献^[1]也认为,估计值更多地依赖于最靠近的样本,负值主要出现于边缘地区.但又指出,负权值可能导致负的估计值,只能对权值为负的点估计值视为 0. 这两种处理方法虽然仍可保证估计精度,但没有从理论和实际上给以证明.

4 实例研究

前已述及,区域化变量的模型及其反映的区域化现象,必须结合专业背景才能得到最充分的解释.结合克立格方法在湖北省大气环境监测站点的优化计算中出现的这一现象予以阐述.

湖北省大气环境监测站点最多时是上世纪 90

年代初,共 36 个(其中武汉市 5 个站点合并为一个),见图 1. 在对不同点的求解克立格方程组过程中,总有权系数为负,最多为 8 个.为了消除负权值的影响,提出了一个改进的权系数规格化方法,这就是将权系数为负的 λ_k 全部去掉,再对剩下的权系数进行归一化处理,即 $\lambda'_i = \lambda_i / \sum \lambda_i$, i 为权值大于 0 的测点号. 这样不仅消除了权系数为负的影响,同时解决了邻近估计点的权系数偏大现象. 这样处理可从理论和实践上予以证明. 证明如下:①根据介值性定理,若一个函数 f 在 $[a, b]$ 上连续,且 $f(a)$ 与 $f(b)$ 异号(即 $f(a)f(b) < 0$),则在 (a, b) 内至少存在一点 $f(x_0)$,使得 $f(x_0) = 0$;②区域化变量是按几何支撑(support)定义的,即区域化变量是限制在一定的空间范围内,这一空间范围称为区域化变量的几何域.在几何域或空间范围内,变量的属性(如相关性、连续性)最为明显,而在几何域或空间范围之外,变量的属性表现减弱或表现为 0(Rossi 等,1992; Dutilleul,1993),这一特性就是区域化变量的空间异质性(spatial heterogeneity)(Journel 等,1978).因此,无论是地质上的矿化作用还是大气污染物的扩散作用,其作用范围都是有界的,即待估点对其邻域内监测站点的影响或邻域内监测站点对待估点的影响只在一定范围内发生作用,超出这一范围就谈不上其“相关性”^[4-5].

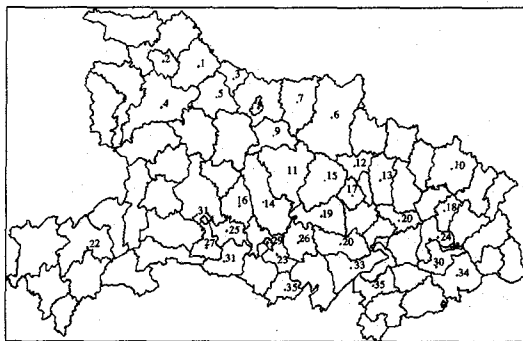


图 1 监测站点示意图

Fig. 1 Sampling points

第①条从理论上证明了对待估点影响程度随着距离增大而逐渐减小,至权系数为 0 处,影响程度为 0,则在此点之外的监测站点(即权系数为负的监测站点)可舍去.其实际意义是在不同方向上权系数为 0 的点的连线的内封闭区域确定了该待估点对周围的影响范围,或影响待估点的邻域范围.从环境污染理论分析,可权且称之为待估点的“作用区域”.有些统计学书上称之为“影响半径”.

第②条则从实际上说明“作用区域”的存在不仅能将研究范围限制在一个合理的区域之内,而且

从环境污染分析的角度,确定某一污染源或污染物的“作用区域”,是有效地实施监测、评价和预测的基础。因此,这一改进是合理可行的。

用 SO_2 浓度计算结果表明^[6],除去 21(宜昌)异常值测点(高出平均值 238.0%)和 35(赤壁)浓度最低的测点(比平均值低 69.27%)之外,所有测点的重新估值与规格化之前的估值相比,其误差绝对值由 5.2%降为 3.47%。虽然改进前后估值结果都在误差许可范围之内,但估值精度有了明显提高。

5 结论

克立格方程组中 $\sum \lambda_i = 1$,不仅是克立格估值为无偏估计的前提条件,也反映了区域化变量之间的一种空间相关性。

在空间构形完整的克立格估值中,对缺失数据的规格化处理办法,以及约简克立格方程组维数时用到的“域对称权相等”的原理等,也很好地反映了权系数的空间相关性。

在权系数为负的情形下,本文结合大气环境监测站点的优化研究,给出了一种新的权系数规格化

方法,并从理论可行性作了相关阐述。实验结果表明,对提高估计精度,节省计算时间,都有改进。

研究区域化变量的现象和处理方法,要充分结合专业背景知识,本文给出的负权值处理方法,仅仅是在大气环境监测站点优化中的一种尝试,是否适用于其他领域,还有待进一步研究。

参考文献:

- [1] 周国法,徐汝梅. 生物地理统计学[M]. 北京:科学出版社,1998.
- [2] 王政权. 地统计学及在生态学中的应用[M]. 北京:科学出版社,1999.
- [3] 田世丰. 数学地质浅析[M]. 北京:地质出版社,1988.
- [4] Eynon B. P. Statistical Analysis of Precipitation Chemistry Measurements over the Eastern United States, Part II: Kriging Analysis of Regional Patterns and Trends[J]. Journal of Applied Meteorology, 1988, 27:1 334-1 343.
- [5] Zekai See. Regional Air Pollution Assessment by Cumulative Simi-variogram Technique[J]. Atmospheric Environment, 1995, 29:543-548.
- [6] 贺华中. 模糊分析理论与方法在环境规划与管理中的应用研究[D]. 武汉:武汉大学资源与环境学院, 2002.

Research on spatial correlation of weighted coefficient in kriging estimation method

HE Huazhong^{1,2}, DAI Zhenyong²

(1. College of Chemistry and Environment, Guizhou University for Ethnic Minorities, Guiyang 550025;
2. College of Resource and Environment, Wuhan University, Wuhan 430079)

Abstract: The paper discusses the spatial correlation between the standardization of weight coefficient of Kriging estimate and the variance at the regional scale, from the following four aspects: 1. Constraint condition of the nonbiased estimate of Kriging equations. 2. Standardization method of weight coefficient under missing spatial coefficient values. 3. Constraint rules in data combination. 4. Standardization method with negative weight coefficient. The paper focuses on and provides a standardization method under a condition of negative weight coefficient. The method is verified by the optimization of monitoring network/points in atmospheric environment from the theory and practice.

Key words: Kriging; weight coefficient; spatial correlation; standardization