

# 地统计学分析方法在水生态系统研究中的应用

赵斌 蔡庆华

(中国科学院水生生物研究所;淡水生态与生物技术国家重点实验室,武汉 430072)

**摘要:**简要介绍了地统计学分析方法的发展和在生态学中的应用情况,提出了将这种方法应用于水生态系统空间格局分析的可能性,然后用梅子垭水库水体藻类叶绿素a含量为例对其浓度场的水平2维空间格局进行了分析,并以此确定了叶绿素a在整个空间和不同方位上空间格局的变化情况。

**关键词:**地统计学;半方差;梅子垭水库;叶绿素a

**中图分类号:**Q178.1    **文献标识码:**A    **文章编号:**1000-3207(2000)05-0514-07

地统计学或地质统计学(Geostatistics)是近30多年来创立起来的一门新兴边缘学科,它是统计学的一个新分支,系在地学(采矿学、地质学)中发展和应用起来的,最初的目的在于解决矿脉和预测等实际问题,后来该学说被地质学家广泛发展与应用,在矿产储量计算方面起到了巨大的作用。其中利用半方差函数(Simevariogram)对空间结构进行分析,以及克立格(Kriging)插值又受到了地学家和生态学家的注意。

与经典的方法不同,地统计学考虑样点的方向、位置和彼此间的距离,直接测定和分析空间依赖性,用于研究有一定随机性和有一定结构性的各种变量的空间分布规律。近年来,地统计学被应用于昆虫学、植物保护学研究中,为研究病虫的空间迁移扩散和分布关系等提供了强有力的工具<sup>[1]</sup>。地统计学仅在近几年里才被应用到生态学研究中,然而它已显示出很大的潜力<sup>[2-4]</sup>。

## 1 地统计学的基本原理

应用经典统计学研究生态学中与地理位置相关的问题时一般明显默认了以下假设:1. 样品不存在空间相关性;2. 研究的对象是纯随机变量;3. 研究的变量可以无限次重复实验或大量观测;4. 样品间具有独立性。事实上这4点对于生态学变量而言均不能满足:生态学变量具有空间分布特征,既有随机性又有结构性,取样后不可能再次取到同样的样品,样品间具有空间相关性等。将地统计学应用于生态学的研究,在克服以上缺点的同时,还具有如下优点:充分利用了各种信息,包括空间位置信息。如果能恰当地将地统计学运用于生态格局的研究,相信一定会在生态格局的定量研究中找到一个新的起点和

收稿日期:1999-10-15;修订日期:2000-05-30

基金项目:国家自然科学基金项目(39670150);国家“九五”科技攻关专题(96-920-04-12)

作者简介:赵斌(1969—),男,湖北钟祥人,主要从事淡水生态学研究和网络信息系统的开发工作

一种新的思维方式。

研究空间分布特征的方法,如地统计学、空间自相关以及分形等,能够对变量空间分布的结构性特征(如空间相关性、方向性、复杂程度等)进行定量描述。同时,地统计学能够提供最优的空间插值手段,从而可制作含量分布图。这种含量分布图是基于对空间结构特征充分认识的基础上的,它优于传统的制图方法和其他插值方法(如几何插值、趋势面分析、普通滑动平均等)<sup>[5,6]</sup>。

所谓区域化变量(Regionalized variable)就是一种数值空间函数,它在空间的每一个点 $X_i$ 都取一个值。例如水体中理化因子的分布就可以看成是一个区域化变量。这种函数值在空间的变化不规则,以致不能用任何直接的数学方法研究它。区域化随机变量与普通随机变量有很大的不同:普通随机变量的取值按某种概率分布而变化,而区域化随机变量则根据其在一个区域内的位置取不同的值。区域化随机变量具有如下的特点:(1)它与普通随机变量相比,只要求松弛了的假定;半方差分析甚至在一些已松弛假定不满足的条件下仍可应用;(2)它为空间格局分析提供从抽样设计到误差分析的综合理论;(3)它可定量地定义生态学上抽样和预测的“代表性”<sup>[7]</sup>。

半方差函数是地统计学分析中的关键概念,它可用于研究区域化随机变量的差异和内在联系:通过测定区域化变量分隔等距离的样点间的差异来研究(区域化)变量的空间相关性和空间结构。空间相关分析是用来检验空间变量的取值是否与相邻空间上该变量的取值大小有关<sup>[7]</sup>。进行空间相关分析的变量必须满足正态分布,并由随机抽样而获得<sup>[8]</sup>。

在以距离 $h$ 相隔的两点 $x$ 和 $x+h$ 处的两个区域化变量值 $Z(x)$ 和 $Z(x+h)$ 之间的变异可以用其增量 $[z(x) - z(x+h)]$ 平方的数学期望(即区域化变量增量的方差)来表示:

$$\gamma(h) = \frac{1}{2} E[Z(x) - Z(x+h)]^2$$

式中,  $\gamma(h)$  称为半方差函数,  $h$  为距离滞后(Distance lag),  $E$  表示数学期望,  $Z(x)$  为在位置 $x$ 处的变量值,  $Z(x+h)$  为在与位置 $x$ 偏离 $h$ 处的变量值。随着距离段的变化,可计算出一系列的半方差函数值。以 $h$ 为横坐标,  $\gamma(h)$ 为纵坐标作图,便得到了半方差函数曲线图(Semivariogram)。从计算公式可见,半方差函数实际上是一个协方差函数,是同一个变量在一定相隔距离上差值平方的期望值<sup>[6]</sup>。差值越小,说明在此距离段上该变量值的相关性越好;差值越大,则在此距离段上该变量值的相关性越差。

上述是较为严格的数学定义,适用于空间上连续分布的变量。但在实际工作中,采样点常常是离散的,对于观测的数据系列 $Z_i, i=1, 2, 3, \dots, n$ ,样本半方差函数值的计算如下:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} = 1'Z(x_i) - Z(x_i + h)1^2.$$

式中,  $\gamma(h)$  为半方差,  $N(h)$  为距离等于 $h$ 的点对数,  $Z(x_i)$  为处于点 $x_i$ 处变量的实测值,  $Z(x_i + h)$  为与点 $x_i$ 偏离 $h$ 处变量的实测值。此半方差函数被称为实验半方差函数。根据实验半方差函数就可以作出半方差曲线(如图 1)。

当存在空间自相关时,随着距离  $h$  的增大,半方差函数值  $\gamma(h)$  也增大,当  $h$  超过某一称之为“变程”(Range) $\alpha$  的距离后,  $\gamma(h)$  往往不再增大,并或多或少稳定在一个极限值附近,该值  $\gamma(\infty)$  称为“基台值”(Sill)。在此范围内,两个点  $Z(x_i + h)$  和  $Z(x_i)$  间存在某种程度的相关关系,而  $h > \alpha$  时,它们就不再相关,因此,变程  $\alpha$  可以看作区域化变量的影响范围。半方差函数曲线在  $y$  轴上的截距称为块金系数(Nugget)。

当被研究对象在不同的方向上呈现出不同性质时就称为各向异性。通过对各个方向  $\alpha$  上的  $\gamma(h)$  的研究,就可以确定是否有异向性存在。异向性一般通过线性变换使之转换为各向同性,并可以综合成一个各向异性模型。如果被研究对象在各个方向上没有显著差异,可以不考虑其方向性,而在全方向上计算半方差函数,研究其性质。

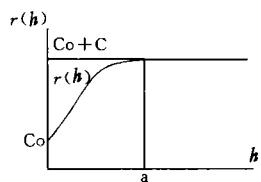


图 1 一个典型的半方差函数图  
( $C_0$ :块金;  $\alpha$ :变程;  $C$ :基台值)

Fig. 1 A typical semivariogram  
( $C_0$ :nugget;  $\alpha$ :range;  $C$ :sill)

几种常用的理论模型如球形(Spherical)、指数(Exponential)、高斯(Gaussian)和线性(Linear)函数都可以拟合实验半方差函数。非水平状的线性模型,可以认为是球形或指数模型的简化形式。

如果把上述半方差函数中的  $h$  和  $\gamma(h)$  在双对数坐标中表示出来,就得到双对数半方差图。该图能够明确地提供两个信息,一是被研究对象的格局的尺度及层次性;二是被研究对象的空间异质性在不同尺度间的相互关系。对于这种空间相关关系的表征是采用分形维数的形式,它的获得借助于公式:

$$D = (4 - m)/2$$

其中  $D$  为分形维数,  $m$  为双对数半方差图的斜率<sup>[9,10]</sup>。因为  $D$  也是基于半方差函数而分析的,对半方差函数分析产生影响的参数同样也会影响这种分维分析。

在分形几何中,最初分形维数是被用来描述物体结构的复杂性和空间占据能力的<sup>[11,12]</sup>。而在这里,  $D$  的几何意义可以这样理解:水体生态因子场是一分形体<sup>[13]</sup>,它的结构复杂程度可以用异质性(即半方差)来描述,故由半方差得到的分形维数是对生态因子场结构复杂性的表征。根据上述公式,当  $D = 2$  时,  $m = 0$ ,此时双对数半方差图是水平的。它表明在统计学意义上所有的尺度间隔样点间的差异性都相同,即在所有的研究区域内是同质的。所以,  $D$  越小于 2,即  $m$  越大,双对数半方差图的直线越陡,说明不同尺度上生态因子场的差异性越大。

## 2 地统计学分析方法与步骤

(1) 将野外和实验室的各项测定数据汇总,输入计算机,并按照 Surfer 软件的要求转化成相应的数据格式(包括采样点的坐标  $X$ ,  $Y$  和测定值,为了便于计算和分析,计算中

半方差函数的形状反映了空间分布的结构或空间相关类型,同时还能给出这种空间相关的范围。实验半方差函数的形状有很大变化,如果不存在相关性,半方差函数会立即达到最大值,表明该现象是完全随机的;半方差函数如果是水平直线,则表现为“纯块金”效应,这是由于微型结构所致,并且常常附加有其他结构的变异。半方差函数要么显示“块金效应”,要么显示一定范围的空间相关结构,当在半方差函数中有识别不出的相关结构时,那么其中就存在纯的“块金”<sup>[9]</sup>。

几种常用的理论模型如球形(Spherical)、指数(Exponential)、

没有采用大地经纬度坐标,而是采用一个相对的虚拟坐标)。

- (2) 数据的正态分布检验(用 STATISTICA for windows 5.0 软件进行)。
- (3) 确定分析的步长(distance lag)、最大分析尺度。
- (4) 利用地统计学分析软件 GS+ for Windows (Gamma Design Software 公司, 下载于 <HTTP://WWW.GAMMADESIGN.COM>) 对各个指标分别进行分析。

(5) 输出结果,包括:各向异性的半方差函数图及模型、块金系数  $C_0$ 、结构系数  $C_1$ 、基台值  $C_0 + C_1$ 、变程  $a$ 、分形维数  $D$  及各自的决定系数  $R^2$ ;各向同性的半方差函数图及模型、块金系数  $C_0$ 、结构系数  $C_1$ 、基台值  $C_0 + C_1$ 、变程  $a$ 、分形维数  $D$  及各自的决定系数  $R^2$ 。

- (6) 对结果进行比较分析,并结合生态学规律进行解释。

### 3 结果分析

#### 3.1 数据的正态分布检验

数据的非正态分布会使方差函数产生比例效应,从而使实验方差函数产生畸变,抬高基台值和块金值,增加估计误差,方差函数点的波动大,甚至会掩盖其固有的结构,因此应该消除比例效应<sup>[6,14]</sup>。为了检验这些实验数据是否符合半方差函数分析的条件,用 Kolmogorov-Smirnov(K-S)正态分布检验概率( $P_{K-S}$ )对它们进行了检验,结果见表 1。由于同一批数据在不同的分组条件下,正态分布的检验结果可能有较大的差别<sup>[15]</sup>,表 1 中所列的数据是在分组数限定为 10~21 范围内,对不同的分组分别计算,取  $P$  为最大时的结果。检验时取显著性水平  $\alpha=0.05$ 。若  $P_{K-S} \geq 0.05$ ,则认为数据服从正态分布。

表 1 原始数据的正态分布检验

Tab. 1 Normal distribution test of original data

指标	1997.5		1997.6		1997.7	
	分组	$P$	$P_{K-S}$	分组	$P$	$P_{K-S}$
Chla	17	0.078	>0.05	21	0.267	>0.05
	16	0.223	>0.05			

#### 3.2 步长和最大分析尺度的确定

所设定的样点,基本能保证地统计学所要求的均匀网格数据,从表 2 各样点的相对坐标可以更清楚地看到这些数据的分布情况:两相邻样点间的距离小于 1 个单位,样点间的最大距离约为 20 个单位。

实验半方差函数图在较大的尺度时,可靠性下降。这时因为样点间的最大长度是给定的,随着间隔距离  $h$  的增大,点对的数量  $N(h)$  逐渐变小,并且某些依赖变量不再对  $\gamma(h)$  有贡献。因此作为统计参量,  $\gamma(h)$  应该有一定的范围限定。一般认为当  $h$  取样线长度的  $1/3$  或一半时,  $\gamma(h)$  才具有统计意义上的代表性;同时,步长不能小于相邻两点间的最小距离。故本研究在应用这个方法进行分析时,取最大样点距离的一半,即 10 个距离单位作为最大分析尺度,取步长为 1 个距离单位,这样就可划分为 9 个距离段。因为实际得到的实验数据不可能取得严格遵守  $h, 2h, 3h, \dots$ , 这些距离的数据对,所以实际计算中所取  $h$  值为各距离段( $nh-(n+1)h$ )中所有数据的算术平均值。鉴于数据基本上是

服从正态分布的,因此这里取算术平均值是比较可靠的,它能代表数据的集中趋势。

为研究半方差函数在不同方向上的特点,即各向异性,对不同方向的半方差函数也进行计算。计算时将全方位平均分为四个角度:S-N、NE-SW、E-W、NW-SE。而各向同性的计算是不考虑方向性的全方向上的半方差函数。

### 3.3 实验半方差函数的参数说明

在上述4个理论模型中,每个模型的基本参数,例如块金值、基台值和变程在不同的模型中所表达的含义都有一定的差异。除此之外,还有三个附加的参数对分析很有用:

空间结构比率(Proportion of spatial structure),即  $C/(C_0 + C)$ ,它提供了一个说明空间结构系数在样点变异中所占比率的测度。如果这个值较大,相应块金效应较小,说明在小尺度空间中被研究对象变化较小,亦说明当前的采样密度对于所进行的研究是足够的。

确定系数  $R^2$ ,用于说明模型拟合程度的优劣,但它没有下面的 RSS 敏感和稳健,所以在判定模型拟合程度的时候它一般用于参考。

剩余平方和 RSS(Reduced sums of squares),可提供模型拟合优劣的精确测度。一般 RSS 取值最小的模型为拟合程度最好的。以下所讨论的所有模型均是取 RSS 值最低的模型。

上述三个参数在所有模型中表达的含义基本相同。

### 3.4 叶绿素 a 空间格局的地统计学分析与讨论

表2和3分别列出了对叶绿素 a 在三个不同采样时间所得的各向同性和各向异性实验半方差函数。

由实测样点构制实验半方差函数的特征曲线解释了变量空间结构的类型,空间结构解释了相隔  $h$  的任意两点的相互作用程度,即空间依赖性。在本研究中,半方差函数图的形状反映了叶绿素 a 空间分布的结构类型及空间相关类型,同时还解释了这种空间相关的范围大小。球形、指数型和高斯型的半方差函数表明数据是聚集的,但数据的空间结构确认存在着很大的变异。

球形半方差函数指明的是聚集分布(或团块分布),它的空间结构是当样点间隔距离达到变程之前时,样点的空间依赖性随样点的距离增大而逐渐减低。指数型与球形模型类似但其基台值是渐渐线,非水平状直线型的半方差函数表明数据是中等程度的聚集分布,其空间依赖范围超过研究尺度;如果是随机分布,则  $\gamma(h)$  随距离没有一定规律的变化;完全随机或均匀的数据,  $\gamma(h)$  呈直线或稍有斜率,块金值等于基台值,表明在抽样尺度下没有空间相关性。 $\gamma(h)$  呈水平状态时称为纯块金效应。空间数据的这些变化对于探索理解叶绿素 a 的分布是怎样形成和为什么形成这样的空间分布有着重要意义。半方差函数曲线模型可以评价它们的空间结构类型。

从表2基台值远远大于块金值可以看出,在全方向上叶绿素 a 具有明显的空间结构,1997年的5月份和6月份,空间结构比较一致,从变程9.93和9.43来看,几乎在整个研究尺度(10个单位)范围内均存在空间依赖性,但5月份的空间结构性更强,因为在同样的取样尺度下,5月份的空间结构比率(0.967)更大,即总变异中有41.58%是由空间依赖性产生的。而7月份叶绿素 a 的空间结构不管根据块金值、基台值或变程来看,同前两个月相比均有较大的不同,因为7月份受降雨因子的影响较大,导致这种差异是完全合理

的。这个结论同前面用盒子维进行的分析结果也是一致的。

表 2 叶绿素 a 的各向同性实验半方差函数

Tab. 2 Isotropic experimental semivariogram of Chlorophyll-a

模型 Model	块金 Nugget $C_0$	基台 Sill $C_0 + C$	变程 Range $\alpha$	空间结构比率 Proportion of spatial structure $C/(C_0 + C)$	$R^2$	
1997.05	Exponential	0.67	20.52	9.93	0.967	0.695
1997.06	Linear	6.58	11.51	9.43	0.428	0.518
1997.07	Spherical	0.04	0.52	3.96	0.923	0.830

表 3 叶绿素 a 的各向异性组合实验半方差函数

Tab. 3 Composed anisotropic experimental semivariogram of Chlorophyll-a

模型 Models	块金 Nugget $C_0$	基台 Sill $C_0 + C$	变程 Range $\alpha_1$	变程 Range $\alpha_2$	空间结构比率 Proportion of spatial structure $C/(C_0 + C)$	$R^2$	
1997.05	Exponential	5.87	39.03	31.11	102.36	0.850	0.230
1997.06	Spherical	6.21	27.36	51.40	143.50	0.773	0.360
1997.07	Exponential	0.08	1.66	42.12	42.15	0.950	0.489

表 4 叶绿素 a 在全方向和不同方向上的分形维数

Tab. 4 Isotropic and anisotropic fractal dimension of Chlorophyll-a

时间	全方向 All directions		S-N		NE-SW		E-W		NW-SE	
	D	$R^2$	D	$R^2$	D	$R^2$	D	$R^2$	D	$R^2$
May	1.717	0.790	1.550	0.690	1.586	0.396	1.878	0.036	1.988	0.004
June	1.906	0.452	1.765	0.489	1.997	0.000	1.506	0.579	1.871	0.240
July	1.971	0.588	1.953	0.640	1.722	0.643	1.584	0.928	1.419	0.481

同时这里还可以再考虑一下空间结构比率:前面已经说明,空间结构比率可以反映小尺度空间的变化大小,这个值越大,相应地小尺度空间的变化越小。7月份叶绿素 a 的空间结构比率达到了 0.923,似乎比 6 月份的 0.428 大许多,但是,由于 7 月份的变程 3.96 太小,代表空间结构性(或空间依赖性)的尺度  $3.96 \times 0.923 = 3.66$  还是比 6 月份的  $9.43 \times 0.428 = 4.04$  要小。所以相应地叶绿素 a 空间结构的变化可以描述为:首先在较少的降雨因子影响下,其分布由较大尺度的聚集型转变成较大尺度上的随机型分布,在更强的降雨因子作用下,又转化成小尺度上的聚集型。

而同样根据表 3 各向异性的组合模型可以看出,在各个不同的方向上,得到的结论是相同的。说明其各向同性与各向异性具有相同的空间结构变化规律,在本例中用各向同性来说明整个研究中的空间格局变化规律是合理的、有意义的。

下面继续探讨通过双对数半方差图所计算的分形维数  $D$  和确定系数  $R^2$ (表 4)来比较不同方向上叶绿素 a 的空间分布格局。 $D$  在这里的几何意义是空间相关性(空间异质性)程度的度量,  $R^2$  是用来说明这个模型对被研究对象的解释效率。 $D$  越小,被研究对象的空间相关性程度越高。从表 4 可以看出,在不同时间,各个方向上的空间分布格局是有较大差别的。5 月份 S-N 方向空间相关性较强,而降雨影响后,变成了 E-W 方向空间格局的优势( $R^2 = 0.928$ ,是各方向  $R^2$  值中的最大者)。也就是说,叶绿素 a 的分布受

降雨水流的影响,空间相关性程度逐渐下降,由聚集型分布转变成均匀性分布。

从叶绿素 a 的分析来看,S-N 方向由于空间走向的优势,对整个水库生态因子场空间格局的变化起主导作用,特别是在降雨作用后,水体中的叶绿素 a 由于受雨水的稀释和水流的冲刷,叶绿素 a 全方向和 S-N 方向上的变化情况非常一致。

### 参考文献:

- [1] 陈晓峰. 生态环境研究与可持续发展[M]. 北京:中国环境科学出版社,1997,415—422
- [2] Robertson G P. Geostatistics on ecology: interpolating with known variance[J]. *Ecology*, 1987, **68**(3):744—748
- [3] Legendre P, et al. Spatial pattern and ecological analysis[J]. *Vegetatio*, 1989, **80**:107—138
- [4] Li H, et al. A geostatistical analysis of spatial patterns of alternative forest cutting patterns: a simulation approach[J]. *Ecological applications*, 1991, **1**:344—382
- [5] 张朝生,等. 长江水系河流沉积物重金属元素含量的计算方法研究[J]. 环境科学学报,1995, **15**(3):257—264
- [6] 张朝生,等. 长江水系沉积物重金属含量空间分布特征研究—地统计学方法[J]. 地理学报,1997, **52**(2):184—192.
- [7] 刘建国. 当代生态学博论[M]. 北京:中国科学技术出版社,1992. 209—233
- [8] Sokal R R, et al. Spatial autocorrelation in bology: 1. Methodology[J]. *Biological Journal of Linnean Society*, 1978, **10**:199—228
- [9] 蔡庆华,等. 芦苇生长格局分形特征的初步研究[J]. 水生生物学报,1998, **22**(2):123—127
- [10] Burrough P A. Fractal dimensions of landscapes and other environmental data[J]. *Nature*, 1981, **194**:240—242
- [11] Burrough P A. Principles of geographical information systems for land resource assessment[M]. Clarendon Press. 1986
- [12] Mandelbrot BB. The fractal geometry of nature[M]. W. H. Freeman. 1982
- [13] Falconer KJ. Fractal geometry:mathematical foundations and applications[M]. John Wiley & Sons. 1990
- [14] 王学仁,等. Outier 和影响分析及其在地质中的应用[A]. 中国地质学会数学地质专业委员会. 中国数学地质(6)[M]. 北京:地质出版社,1995. 27—35
- [15] 姜炳麟,等. 数理统计疑问解析[M]. 武汉. 华中理工大学出版社,1993. 182—197

## AN APPLICATION OF GEOSTATISTICAL ANALYSIS IN FRESHWATER ECOSYSTEM

ZHAO Bin and CAI Qing-hua

(Institute of Hydrobiology, The Chinese Academy of Sciences; State Key Laboratory of Freshwater Ecology and Biotechnology, Wuhan 430072)

**Abstract:** The development of geostatistical analysis and its application in ecology is reviewed in the paper, following with the possibility analysis of spatial pattern in freshwater ecosystem. As a preliminary study, this paper makes use of geostatistical method to analyze horizontal two-dimension spatial pattern of concentration field of algea chlorophyll-a in Meiziy Reservoir, by which we can determine the variation over all study range(isotropic) and in different direction(anisotropic).

**Key words:** Geostatistics; Semivariance; Meiziy Reservoir; Chlorophyll-a