

粒度分布数据的计算机光滑插值绘线方法

黄建东

(厦门大学环境科学研究中心, 厦门 361005)

提 要 以一种单调光滑的分段插值函数为基础, 在计算机上实现了光滑的粒度分布累积频率曲线和频率分布曲线的绘制。这一方法为编制软件在计算机上实现粒度分析数据高效高质量的系统处理提供了基础, 为粒度分布特征的普遍表述向更精确、更全面的方向深化提供了途径。

关键词 累积频率曲线 频率分布曲线 插值绘线 单调光滑 计算机 方法

第一作者简介 黄建东 男 36岁 助理研究员(硕士) 从事海岸动力沉积环境与动力地貌研究

粒度分析的直接结果表现为各粒级的百分含量。为进一步了解粒度组成特征, 需将数据整理成连续、光滑的累积频率曲线和频率分布曲线, 并插值求算所需的累积频率和百分位值, 计算各种参数。传统的手工处理工作量极大; 而在计算机处理方面, 除计算参数外, 目前的许多通用绘图软件并未达到与使用曲线尺光滑连接数据点绘制曲线相同的效果, 因其所绘曲线或不满足累积频率曲线的单调性, 或不能满足必要的光滑性。本文提供了一种可在计算机上实现的, 以特定的插值函数的构造为基础的插值、绘线方法。通过合理地解决曲线的单调、光滑问题, 在计算机上实现了光滑度较优的累积频率曲线和相应的频率分布曲线的绘制, 且与使用曲线尺绘线的过程有很好的相似性。由插值函数, 也易进行数据区间内任意给定点上的插值和百分位值计算。

1 数学处理及其在计算机上的实现

累积频率曲线是概率分布函数图形的近似, 频率分布曲线则是相应的概率密度函数图形的近似。概率分布函数为以 0 和 1 作为渐近线的连续函数, 且有单调上升的性质。在计算机上实现连接各数据点绘制光滑曲线, 需寻找较为简单并满足上述性质的函数作为插值函数, 达到插值、绘线的目的。

三次样条插值函数是一类低次分段多项式函数, 具有结构简单、稳定并处处具有二阶光滑度的特点⁽¹⁾, 对于逼近分布函数有较好的适用性。研究成果⁽²⁾指出了以三次样条插值函数 $S(x)$ 逼近正态分布和罗辛分布函数的可行性和应用的限制。即 $S(x)$ 在概率密度曲线下占 90% 以上面积的区间中可以满足上述二种概率分布函数的单调性。由于沉积物的粒度(ϕ 粒级)服从正态分布, 一部分服从罗辛分布⁽³⁾, 故而用三次样条插值函数 $S(x)$ 作为沉积物粒度数据主要部分的插值函数是合理与可行的。而且, $S(x)$ 对各数据点的连接过程与使用曲线尺光滑连接数据点的过程具有相同的力学背景⁽¹⁾, 故以 $S(x)$ 为插值绘线模型与以手工用曲线尺连接数据点的绘线模型在性质上是类似的。从而, 除了所绘曲线相一致外, 由曲线求得的

参数也会是相近的。这就使计算机所绘曲线与以往手工处理的资料具有一定的可比性。而这一点对数据处理方法变更和升级时不同时期数据的衔接对比也具有重要意义。

作为本文工作的基础的插值函数,其提法和构造过程可概述如下:

给定数据点列 $\{(x_i, y_i)\}$, $(i=0, 1, \dots, n)$, $a=x_0 < x_1 < \dots < x_n=b$, $y_i < y_{i+1}$ ($i=0, 1, \dots, n-1$)。需在 $[a, b]$ 上构造一连续的插值函数 $F(x)$, 满足单调性和较好的光滑性。

为此,依给定的数据,先在 $[a, b]$ 上以两端一阶导数等于零为边界条件,求出 $[a, b]$ 上的三次样条插值函数 $S(x)$ ^[2]。在出现 $S'_i(x) < 0$ 的子区间集 I 上,通过子区间两端点的坐标和指定两端点处的一阶导数值 m_i, m_{i+1} ^[4],以两种初等形式的函数作为插值函数,求解 $[x_i, x_{i+1}] \subset I$ 上的 Hermite 插值问题,得到这样的各子区间上的插值函数 $Y_i(x)$ 。而在 $S'_i(x) \geq 0$ 的子区间段,以与 $Y_i(x)$ 所在子区间相邻点的 $Y'_i(x)$ 为边界条件,再次求取 $S(x)$ 。若在这样求得的 $S_i(x)$ 所在子区间上还有 $S'_i(x) < 0$ 的点,则对所有这样的子区间,由以上求解 Hermite 插值问题的方法继续求得 $Y_i(x)$ 。最后,在 $S'_i(x) \geq 0$ 的子区间,令 $F_i(x) = S_i(x)$; 在其余子区间,令 $F_i(x) = Y_i(x)$,便得到 $[a, b]$ 上的分段插值函数 $F(x)$ ^[4]。

由 $F(x)$ 的构成函数的形式^[4],可知在 $[a, b]$ 上总有 $F'(x) \geq 0$,且 $F'(x)$ 在所有点 $x \in [a, b]$ 都是连续的。而在 $F(x) = S(x)$ 的子区间, $F'(x)$ 连续。这样的子区间对于满足正态分布或罗辛分布的数据来说,占有 $[a, b]$ 中的大部分。这就保证了在 $[a, b]$ 的大部分中 $F(x)$ 的单调性和二阶光滑度。

因为频率分布曲线为 $F'(x)$ 的图形, $F(x)$ 的二阶光滑对于频率分布曲线的绘制便具有重要意义。 $F'(x)$ 连续保证了 $F'(x)$ 曲线为一阶光滑,使频率分布曲线有较光顺的性质。

由以上构造的分段插值函数 $F(x)$,可在计算机上绘出单调光滑的累积频率曲线和相应的频率分布曲线。同时,可插值求算所需的累积频率值以及求算各百分位值,计算矩法和图解法粒度参数。还可求出频率分布曲线上的众数位点和各峰位点及峰高值。而且,这样的插值函数对于实际计算中数据点横坐标等距的情形和不等距的情形均可作出效果良好的处理。

2 实例

依上节的原理,在计算机上绘制了光滑的粒度分布累积频率曲线和相应的频率分布曲线。对数百例这类曲线的绘制都获得满意的效果。表 1 和图 1 给出几个沉积物样品的粒度分析数据和在计算机上绘出的曲线。分析样品时所用套筛粒级的 ϕ 值是不等间距的。

表 1 几个沉积物样品的各粒级百分含量(%)

Table 1 Percentage data of the grain size distribution for several sediment samples.

粒级(ϕ)	-2.322	-1.678	-1.000	0.007	0.786	1.490	2.737	3.366	4.158
样品 A	0.0000	0.0000	0.0000	0.6631	5.2455	17.6543	70.4317	6.0054	0.0000
样品 B	0.2517	1.0573	21.7055	31.2775	32.8130	9.0146	2.8420	0.9381	0.1003
样品 C	0.0000	3.1337	15.1595	13.0822	9.1644	7.0827	41.7912	10.0923	0.4880
样品 D	1.6356	7.3771	26.6339	12.2855	18.7671	12.8782	19.8263	0.5161	0.0634

注: 样品为现代海滩沉积物。在杭州大学河口与港湾研究室完成分析;分析方法为筛析法。

图 1-A 中的频率分布曲线为略有些不对称的尖瘦型单峰曲线;图 1-B 则为双峰曲线,其中较细组分的成分稍多,并有一个稍长的尾部;样品 C 的曲线为图 1-C,显示三个峰,中间部分的组分含量略少;图 1-D 为样品 D 的曲线,有两个主要部分,较细的部分又由两个次组分组组成。所有图中曲线都有良好的光滑性并且累积频率曲线均保证了不失其单调性质。

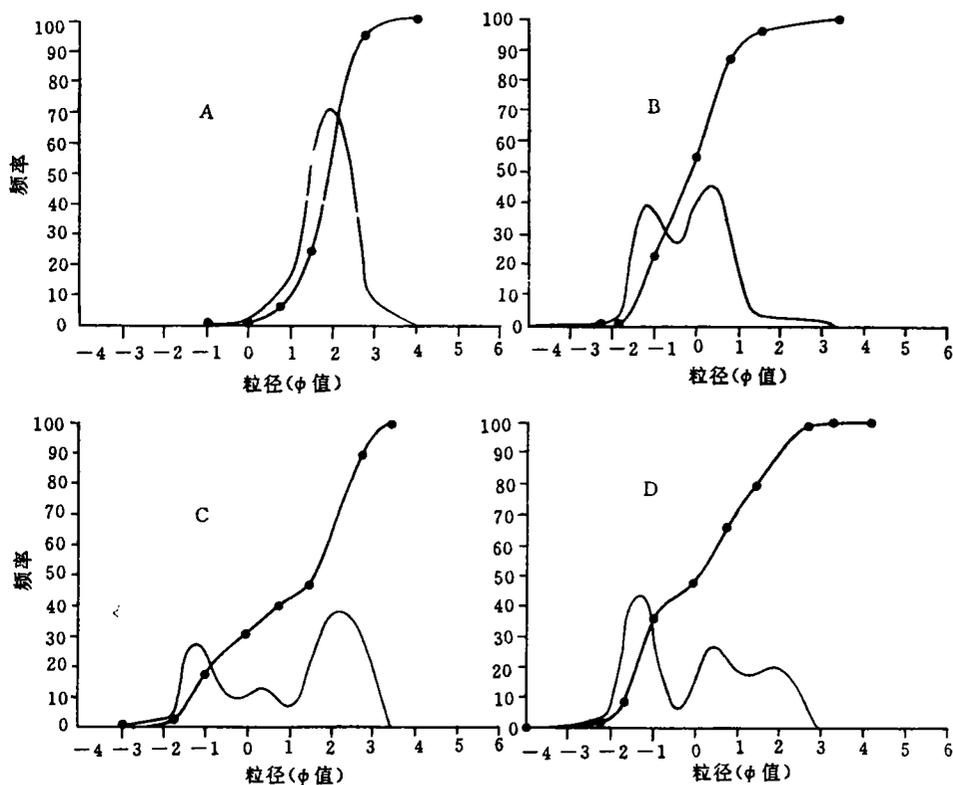


图 A,B,C,D 分别对应于表 1 中的样品 A,B,C,D。

图 1 由本文的方法在计算机上绘出的几组累积频率曲线和频率分布曲线。

Fig. 1 Four groups of the cumulative curves and the frequency curves, drawn in the computer by the method introduced in this paper. Fig. 1-A, 1-B, 1-C, and 1-D correspond to the data Group A, B, C, D in Tab. 1.

图 1 的各组曲线,特别是其中的频率分布曲线,对粒度组成的反映是详细而清晰的。通过众数组(曲线中的峰)的个数,各峰位,峰高,各峰间的关系,各组分的百分含量等,可对沉积动力和沉积环境,沉积物来源及搬运过程等作进一步的探讨。而频率分布曲线的这种作用是粒度特征的其它表述方法所无法替代的。

3. 关于方法的意义

(1) 以三次样条插值函数和另二类初等形式的函数构造出既保单调又有较优光滑度的分段插值函数,在其一阶导数于所有数据点处都可保持连续的基础上,在许多子区间段上

都有二阶光滑的性质。据此,在计算机上实现了对累积频率曲线和频率分布曲线的绘制。通过插值函数,可方便地求取所需的累积频率值和百分位值,求得各种参数。从而,使粒度分布数据的绘线与计算实现了在满足特定要求下的高效高质量处理,为开发粒度分析数据处理的基础软件准备了良好的基础。在此基础上作者编制了这种处理软件,获得了良好的应用效果。

(2) 在大部分子区间上,累积频率曲线的绘制与使用曲线尺光滑连接数据点的过程有相同的力学背景。从而,在计算机上依本文的插值函数模型绘出的曲线与使用曲线尺绘制的曲线在性质上相一致,计算的参数结果也具可比性。因此,绘线方法更新前后的曲线和参数都可相互参照,实现资料的衔接。

(3) 除对于正态分布和罗辛分布的数据有良好效果外,对于来自多个总体,反映多组分组成的样品数据,所构造的插值函数亦能较好地满足需要。对不满足正态分布或罗辛分布的数据,在许多子区间端点上仍可达二阶光滑。对此,图 1 的多峰曲线是很好的说明。

(4) 本方法中频率分布曲线的绘制,以分布函数的导函数为基础,数学背景清楚,意义明确。频率分布函数及其曲线对粒度分布特征的表述是清晰直观的,较直方图更为细致,较粒度参数更为全面和直接。频率分布曲线的作用是包括概率值累积频率曲线在内的其它粒度特征表述方法所无法替代的。概率值累积频率曲线表现出各个正态次总体,但各个次总体的实际的粒径范围(表示某一组分的线段向两侧延伸的程度)并未能从曲线上得到反映,从而各个组分间的混掺情况实际上也就难以完全反映出来。所以,频率分布曲线中各组分的分布状况与概率值累积频率曲线中各组分之间并不一定存在一一对应的关系。前者的曲线上“谷”的位置并不一定对应于后者曲线中截点的位置。以表 1 中的数据换算成累积值数据绘制概率值累积频率曲线,便可见各截点位置与图 1 中各“谷”的位置并未一一对应。以样品 C 为例,在概率值累积频率曲线上,其粗、细截点分别在约 -1ϕ 和 2.2ϕ 处,但在频率分布曲线上,这两个点的近旁却是众数位的位置。实际上,这类现象在同时绘出上述两种曲线时是经常会遇到的,如文献^[5]中即可见这样的例子,虽然其作者对此并未提及。这种现象的存在,从一个方面说明频率分布曲线在粒度特征的表述中具有独到的作用。而且,将概率值累积频率曲线与频率分布曲线结合起来考察分布性状,才能对分布特征获得更全面的了解,并在此基础上对相应的沉积动力与环境进行深入的探讨。本方法的意义之一便是使这一点得以方便地实现。

当今计算机技术的发展和普及已为高效高质量的数据处理提供了可能性。本文的方法则为使沉积物粒度资料实现这种处理提供了基础。该方法在提高沉积物粒度数据处理的效率和质量,更精确更全面地揭示粒度分布特征方面的意义是明显的。

收稿日期:1995 年 3 月 17 日

参 考 文 献

- [1] H. J. 阿尔伯格等,赵根榕等译,样条理论及其应用,北京:科学出版社,1981。
- [2] 黄建东,两种概率分布函数的样条逼近,厦门大学学报(自然科学版),1988,27(2):234~236。
- [3] 戴东林主编,沉积物(岩)粒度分析及其应用,北京:地质出版社,1978,37~44。
- [4] 黄建东,单调光滑分段插值函数的一种构造及其在统计类数据处理中的应用,厦门大学学报(自然科学版),待刊。

(5) 张君元,台湾海峡及其附近地形和沉积特征的初步研究,海洋科学集刊,1989,(30):1~17。

A Computer—Based Method for Drawing a Smooth and Monotone Curve of the Grain Size Distribution Data

Huang Jiandong

(Research Center of Environmental Sciences, Xiamen University, Xiamen 361005)

Abstract

To draw a computer plotted cumulative curve by an interpolation function, it is essential to possess the characters of monotonicity and smoothness. Based on a method to construct a sort of monotone and smooth interpolation function, the computer drawing of the cumulative curve of the grain size data has been realized with the monotonicity and the first or more than first order of smoothness of the cumulative curve. By the derivative function of the interpolation function, the homologous frequency curve has been plotted in the computer as well with its smoothness corresponding to that of the cumulative curve. Because the grain size distribution (in phi scale) of sediments follows the normal distribution and Rosin's distribution, while as an approximating function, the cubic spline interpolation function can satisfy (with some restriction) the monotonicity of the probability distribution functions of these two sorts of distribution in an interval, in which there is more than 90% of the area under the probability density curve, so in the method introduced in this paper, the cubic spline function is taken to be the interpolation function in the main part of the data interval $[a, b]$. In those subintervals in which there exist points of $s'(x) < 0$, two classes of the elementary functions are assigned to be the interpolation functions with the character of monotonicity and at least first order of smoothness at all knots. By this interpolation function, besides plotting the monotone and smooth cumulative curve and corresponding frequency curve, grain size parameters of moment and graphic measures can be calculated through the interpolation and percentiles computation. From the information revealed by the frequency curve, such as number of the peaks, location of each peak, the peak values, correlation among the peaks, and so on, the further studies could be taken out on the characteristics of the dynamics, environment, source and transportation of the sediment. The method has provided a basis for developing a software of the grain size analysis data processing to treat the data effectively and with high quality. And it is of benefit to remedy the defect of other means in expressing grain size distribution characters and to make it more exactly and more comprehensively.