

# 随机信号测量的数据处理\*

厦门大学 黄衍镇

## 摘 要

当仪器所观测的物理量由传感器转换为电信号是随机信号时,则测量的给出值是随机的。本文从随机信号的特性和随机误差的规律性出发,结合具体实验,讨论随机测量值的数据处理。

## 一、前 言

气象水文海洋仪器所观测的信号,有一类是随机的。作为例子,水中悬浮泥沙的浓度  $M$  和其声反向散射信号  $S(t)$  的强度之间有一定的关系。因此可通过测量  $S(t)$  的强度,即  $S(t)$  的幅度包络  $e(t)$ ,来反演求得  $M$ 。当通过换能器向水中辐射一个矩形包络正弦载频声脉冲后,水中许许多多悬沙粒子反向散射,并为换能器所接收的信号  $S(t)$ ,近似为许许多多具有随机振幅和相位的同频正弦信号的迭加<sup>[1]</sup>,因而  $S(t)$  的幅度包络  $e(t)$  是随机的。由收发合置的换能器面作为参考点的距离选通门信号  $S_p(T_j, \tau)$  开启采样门,对  $S(t)$  采样得到  $S_j(t)$ ,则  $S_j(t)$  是对应于换能器轴向上不同距离  $Z_j$  的采样体积  $j$  中的悬沙  $M_j$  的反向散射信号(如图 1 所示),即

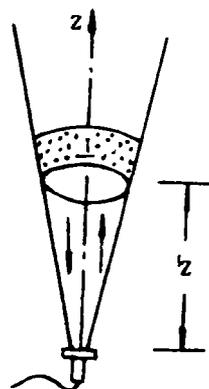


图1 距离选通、采样示意图

$$Z_j = (T_j C + l) / 2 \quad \text{和} \quad l = \tau C / 2 \quad (1)$$

$$S_j(t) = S_p(T_j, \tau) \cdot S(t) = e_j(t) \sum_{i=1}^n \cos[\omega_i t + \varphi_i(t)] \quad (2)$$

式中  $S_p(T_j, \tau)$  为重复周期  $T_j$  脉宽为  $\tau$  的矩形采样脉冲列,  $C$  为水中声速。  $S_j(t)$  中的幅度包络  $e_j(t)$  含有悬沙浓度  $M_j(t)$  的信息,为此要提取  $e_j(t)$ 。

为了遥测水中悬沙含量,笔者研制了声反向散射实验样机,并在实验室进行了模拟实验。参考文献[2]中介绍了从  $S_j(t)$  线性提取  $e_j(t)$  以及对  $e_j(t)$  进行积分、保持、A/D 变换等数据预处理,得到了  $e_j(t)$  的数字量  $x_j$ ,这仅仅是硬件的部分电路。由于  $S(t)$  是随机的,所以  $S_j(t)$ 、 $e_j(t)$  乃至线性 A/D 变换后的测量给出值  $x_j$  也是随机的。测量数据带有随机性,一次测量的数据是不准不可靠的,必须通过  $N$  次测量建立随机样本(总集)—— $\{x_i\}_j$ ,

\* 省科学基金研究内容。

( $i=1, 2, \dots, N$ ), 并对  $\{x_i\}_j$  进行统计平均。本文着重讨论如何对  $\{x_i\}_j$  的数据进行处理, 以期通过反演求得  $M_j$ 。

## 二、随机变量及其误差的统计特性

所谓随机信号是指其波形或参数具有随机性的信号。对随机信号的测量, 其随机性包括两方面特性: 一是事先不能确切预知其值或形态; 二是重复进行观测或试验时, 不能肯定重复出现其原值或原形态。即测量值是一随机变量。对某一次测量而言, 随机误差使测量数据偏离数学期望的大小和方向是没有规律的, 但多次测量的随机误差和测量数据的分布是服从一定的统计规律的。当对随机被测量  $X$  在相同的条件下进行无穷多次 ( $n \rightarrow \infty$ ) 测量, 并认为每次测量是独立的, 满足概率论中的中心极限定理时, 则测量中的随机误差分布及随机误差影响下测量数据的分布大多是接近于正态分布<sup>[3]</sup>, 即测量值  $x_i$  对称地分布在被测量  $X$  的数学期望  $M(X)$  两侧。被测量  $X$  的数学期望  $M(X)$ 、方差  $\sigma^2(X)$  和标准偏差  $\sigma(X)$  分别为

$$M(X) = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{当 } n \rightarrow \infty)$$
$$\sigma^2(X) = \frac{1}{n} \sum_{i=1}^n [x_i - M(X)]^2 \quad (\text{当 } n \rightarrow \infty)$$
$$\text{和 } \sigma(X) = \sqrt{\sigma^2(X)}$$

式中方差  $\sigma^2(X)$  或标准偏差  $\sigma(X)$  是用来描述测量值的离散程度或者说用来衡量随机误差的大小, 即随机误差对测量值的影响。

理论上, 求  $M(X)$  和  $\sigma(X)$  都需要无穷多个测量数据, 可是实际的测量中只能进行有限  $N$  次的测量, 因而只能根据  $N$  次测量的数据进行估计  $M(X)$  和  $\sigma(X)$ 。这些  $N$  次测量的数据带有随机性, 称之为随机样本或样本总集  $\{x_i\}$ , ( $i=1, 2, \dots, N$ ), 它是相对于被测量总体而言的。不难推得, 有限  $N$  次测量值的算术平均值  $\bar{x}$  的数学期望  $M(\bar{x})$  和方差  $\sigma^2(\bar{x})$  分别为

$$M(\bar{x}) = M(X) \quad \text{和} \quad \sigma^2(\bar{x}) = \frac{1}{N} \sigma^2(X)$$
$$\text{或} \quad \sigma(\bar{x}) = \sigma(X) / \sqrt{N}$$

上式说明  $N$  次测量值的算术平均值的数学期望  $M(\bar{x})$  就等于被测量  $X$  的数学期望  $M(X)$ , 而平均值的方差  $\sigma^2(\bar{x})$  比总体测量值的方差  $\sigma^2(X)$  小  $N$  倍, 或者说比标准偏差  $\sigma(X)$  小  $\sqrt{N}$  倍。这就是采用统计平均的方法减弱随机误差影响的理论根据。因为平均值的标准偏差  $\sigma(\bar{x})$  是描述大量平均值的离散程度的。

在一定的时、空条件下, 某一被测量的真实  $x_0$  是一个客观存在的确定数值, 但要确切地说出真值  $x_0$  的大小却很困难。这是由于测量仪器、环境、条件、方法以及人为因素等等都会使测量结果与真值不同, 其差别就是测量误差。一般而言, 任何一次测量误差都含有随机误差和系统误差, 有时还会有粗大误差。根据误差的性质的特点, 对 3 种误差的处理方法是不同的: 对含有粗大误差的测量值(坏值)应予首先剔除, 因为它明显地偏离了真

值,严重地歪曲了测量结果;对于随机误差的影响,如上所述,要采用统计平均的方法来消除或减弱,因为随机误差大多接近正态分布,具有对称性和抵偿性的特点;对于系统误差则主要靠测量过程中采用一定的技术措施来削弱或对测量值进行必要的修正(包括修正公式和修正曲线),因为在相同条件下多次测量同一量时,系统误差的绝对值和符号保持恒定,或者在条件改变时按某种确定的规律而变化。多次测量取平均值并不能减少系统误差。通常是把系统误差视为含有恒定部分和变化部分两部分,前者可以通过对测量值进行修正来削弱,后者常与随机误差的变化范围合成在一起,进行减少测量误差的处理。

由上所述,测量数据的处理归结为采用统计平均的方法来减少随机误差,并对测量值进行必要的修正来削弱系统误差。

### 三、随机测量的数据处理

#### (一)建立合适的观测样本 $\{x_i\}_j$

为了探求水中悬沙含量  $M_j$  与其声反向散射信号强度的关系,首先对粒径  $\phi \leq 63 \mu\text{m}$ , 浓度  $M_j$  为  $500 \text{mg/L}$  的高岭土悬浮液进行实验。充分搅拌悬浮液,使浓度分布尽量均匀;但又要待水体平静了才开始取数,以消除气泡的影响。观测系统(实验样机)对示于图 1 的选测距离  $Z_j$  上的水层(水团)1 采集其反向散射信号。每采样一次,得到一个  $S_j(t)$ ;经参考文献[2]的数据预处理,获得一个对应的原始数据  $x_i$ (数字量);采样  $N$  次,就得到  $N$  个测量数据的随机样本  $\{x_i\}_j$ , ( $i=1, 2, \dots, N$ )。  $N$  越大,由这组数据求被测量总体的数学期望和标准偏差的估计值就越接近实际情况。但在实际测量中,  $N$  的取值既要取决于测量精度的要求,还要保证全部  $N$  次测量中的测量条件保持不变,因此  $N$  又不能过大。本实验悬浮液中的高岭土微粒会慢慢自然沉降,为了保持采样水体中  $M_j$  尽量不变,要求在 1 分钟内读取观测样本,所以  $N$  取 40,得如表 1 所列的  $\{x_i\}_j$ 。

表 1  $M_j = 500 \text{mg/L}$  的观测样本  $\{x_i\}_j, i=1, 2, \dots, 40$

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$x_i$	366	269	285	273	255	236	256	260	257	253	257	262	259	252	254	250	255	251	249	253
i	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
$x_i$	243	244	253	249	242	254	250	216	248	251	247	240	245	245	247	245	236	257	246	259

观察表 1 中的数据可知:(1)  $x_i$  的数值是无规律的,其集合  $\{x_i\}_j$  是一组随机变量;(2) 随着采样序数  $i$  的增加,  $x_i$  的数值在总体上有减少的趋势。这是由于水中悬浮液微粒自然沉降,以至越过了采样水体 1,使得 1 中的  $M_j$  降低了,因而  $x_i$  减小。所以观测样本表 1 的数据是符合试验时的实际情况的。

#### (二)对观测样本 $\{x_i\}_j$ 求统计平均

采样水体 1 中如果有气泡,则其反向散射的贡献要比相同粒径的泥沙大得多。因此表 1 中  $x_i$  的几个最大值,应认为是既有许多悬浮粒子,又有微小气泡共同作用因而偏离了真值,要作为粗大误差引起的坏值给予剔除。而这几个数据又含有悬浮粒子反向散射的随机误差,根据随机误差有对称性和抵偿性的特点,去掉几个最大的值,还要相应去掉

几个最小的。所以对表 1 的样本, 去掉 4 个最大的 ( $i=1, 2, 3, 4$  的  $x_i$ ) 和 4 个最小的 ( $i=25, 28, 32, 37$  的  $x_i$ ), 即相当于 10 个数据中去掉一大一小, 只取 32 个数据来求统计平均, 即

$$\bar{x} = \frac{1}{N-8} \sum_{i=1}^{N-8} x_i = \frac{1}{32} \sum_{i=1}^{32} x_i = 251.94 = \langle X \rangle_j \quad (3)$$

$$\text{和 } \hat{\sigma}(X) = 5.14 \quad (4)$$

式中  $\bar{x}$  和  $\hat{\sigma}(X)$  分别是这组数据的数学期望和标准偏差的估计值。 $\hat{\sigma}(X)$  越小, 估计值  $\bar{x}$  越接近于实际情况。

如果取表 1 的全部数据参与统计平均, 则

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{40} \sum_{i=1}^{40} x_i = 254.20 \quad \text{和} \quad \hat{\sigma}(X) = 21.17$$

虽然两种方法所得的  $\bar{x}$  值相差无几, 但是前者的  $\hat{\sigma}(X)$  值远小于后者, 所以应取前者的  $\bar{x} = 251.94$  为观测的估计值。

用上述同样的方法, 对  $M_j$  为 750、1000、1250、1500、1750、2000 (mg/L) 的悬浮液分别实验, 建立各自的观测样本, 并对各个样本按式(3)进行统计处理, 得到各个样本的  $\langle X \rangle_j$  列于表 2, 并由微机作图, 示于图 2 中的实线点。

同时, 对各个样本, 均取

全体数据 ( $N=40$ ) 分别求统计平均, 得到各个样本的  $\langle X \rangle_j$  列于表 3, 并示于图 2 中的虚线点。表 2 和表 3 中同一个样本的  $\langle X \rangle_j$  相差也不多, 但由计算可知, 表 2 各个样本的  $\hat{\sigma}(X)$  均小于表 3 对应的  $\hat{\sigma}(X)$ 。所以取表 2 的  $\langle X \rangle_j$  作为各个观测样本的估计值更为合理。

表 2 对不同  $M_j$  的观样本的统计平均值  $\langle X \rangle_j$  (舍去各样本中 4 个最大、4 个最小值)

mg/L	500	750	1000	1250	1500	1750	2000
$\langle X \rangle_j$	251.94	340.38	481.19	548.66	646.19	785.16	907.50

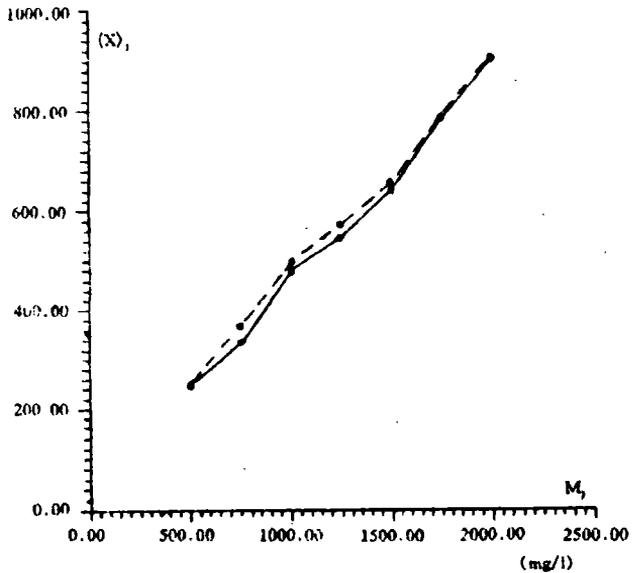


图 2 不同  $M_j$  与其观测样本的统计均值  $\langle X \rangle_j$ , 实点为表 2 中的  $\langle X \rangle_j$ , 虚线点为表 3 中的  $\langle X \rangle_j$

表3 对不同  $M_j$  的观测样本的统计平均值  $\langle X \rangle_j$  (取样本全体数据的统计平均)

mg/L	500	750	1000	1250	1500	1750	2000
$\langle X \rangle_j$	254.20	371.32	497.72	570.90	658.30	791.38	912.05

### (三) 系统误差的实验室校正

由表2和图2中的实线点, 可得在观测范围  $M_j = 500 \sim 2000 \text{mg/L}$  内,  $M_j$  和  $\langle X \rangle_j$  的关系拟合为

$$\langle X \rangle_j = A_0 + KM_j \quad (5)$$

式中  $A_0 = 33$ ,  $K = 0.437 / (\text{mg/L})$ 。或者为

$$\begin{aligned} M_j &= (\langle X \rangle_j - A_0) / K \\ &= k(\langle X \rangle_j - A_0) \end{aligned} \quad (6)$$

式中  $k = 2.29$ 。因此, 电微机处理求得  $\langle X \rangle_j$ , 便可反演得到  $Z_j$  上的水层1中的  $M_j$ 。

必须指出, 式(5)中  $K = 0.437$  仅仅是观测系统(实验样机)在实验室条件下得到的校正因子。实际上  $K$  是反向散射函数  $F(Z, T, P)$ 。它和散射强度、距离  $Z$ 、温度  $T$  及压力  $P$  有关, 因此必须由实验室的和观测现场的大量观测数据才能最后确定。这是有待今后继续实验研究的内容。

### 参 考 文 献

- [1] A. E. Hay: "On the remote acoustic detection of suspended sediment at long wavelenthes", J. G. R. 88(C12 - 15), PP. 7525(83)
- [2] 黄衍镇等: "线性提取幅度信息及其数据预处理", 《气象水文海洋仪器》第3期, 1996年
- [3] 陈元亨等: 《信号处理原理》, 四川大学出版社, 1989年