

水质常规监测采样频率确定方法研究

邓英春

(安徽省水文局,安徽 合肥 230022)

摘要 对安徽省淮河蚌埠闸以上干流 2 个水质监测断面的水质监测资料用偏差分析的方法,研究确定水质常规监测采样频率的方法。对 1998 年以来淮河蚌埠闸以上干流和颍河干流安徽段 7 个监测断面的水质常规监测数据进行了正态性检验,并对 1998 年以来的 12 个监测断面水质常规监测数据用统计学的方法,研究确定水质常规监测采样频率的方法。

关键词 水质监测;水质站;监测断面;采样频率

中图分类号:X832 文献标识码:A 文章编号:1004-693X(2005)04-0037-05

Method for determination of sampling frequency of routine monitoring of water quality

DENG Ying-chun

(Hydrology of Bureau Anhui Province, Hefei 230022, China)

Abstract Field data from two water quality inspection sections upstream of the Bengbu Gate on the Huaihe River of Anhui Province were analyzed by the deviation method. The method to determine the sampling frequency of routine monitoring of water quality was discussed. Through the analyses on routine monitoring data of water quality since 1998 of seven sections in mainstreams above of the Bengbu Gate of the Huaihe River and in Anhui section of the Yinghe River by the normality test method and data of 12 sections by statistic method, the technique to determine the sampling frequency of routine monitoring of water quality was discussed.

Key words water quality monitoring; water quality station; monitoring section; sampling frequency

本文对 1998 年以来的安徽省淮河蚌埠闸以上干流和颍河干流安徽段 12 个水质常规监测断面水质常规监测数据进行了分析研究,得出了确定水质常规监测采样频率的方法。

1 研究区域概况

本次研究区域为安徽省境内蚌埠闸以上淮河干流和颍河干流,蚌埠闸以上流域面积约 12.1 万 km²。淮河由洪河口进入安徽省境内至蚌埠闸约 288 km,其左岸有颍河、涡河等较大支流汇入,右岸有史河、淝河、东淝河等较大河流汇入,沿岸分布有寿县、凤台县、淮南市、怀远县以及颍县南照集、正阳关等重要城镇,并有在建的临淮岗洪水控制工程。颍河流

域面积 36 728 km²,其中安徽省为 4 010 km²,河道长 207 km,沿岸分布有界首市、太和县、阜阳市、颍上县等重要城镇。安徽省淮河蚌埠闸以上污水排放量为 144.14 万 t/d,其中直接排入淮河干流污水为 56.52 万 t/d,排入颍河干流的污水为 43.57 万 t/d,排入涡河干流的污水为 10.1 万 t/d。研究区的水质监测工作始于 20 世纪 50 年代,其中水污染监测始于 20 世纪 70 年代,水质资料完整,且系列较长。

2 偏差分析方法

为了使水质常规监测的采样频率更科学合理,现以淮河干流石头埠和蚌埠闸上监测断面为例,介绍确定水质常规监测采样频率的方法。通过对水质

历史资料进行分析,选用每个监测断面的代表性项目,剔除颍河、涡河开闸期间大量下泄高浓度污水的监测资料后,选取各种采样频率进行偏差分析,其具体方法是:以淮河干流石头埠和蚌埠闸上监测断面2000~2003年每日监测1次的水质数据作为参考值,分别对每3日、每5日、每10日、每15日、每月、每两月监测1次的月平均水质浓度、通量和年平均水质浓度、通量进行偏差分析(以相对偏差表示),以研究确定水质常规监测的最小采样频率。相对偏差以式(1)计算:

$$e_p = \frac{\bar{c}_d}{\bar{c}_g} \times 100\% \quad (1)$$

式中: e_p 为相对偏差(%); \bar{c}_d 为水质项目低采样频率的月(年)平均浓度(通量); \bar{c}_g 为水质项目高采样频率的月(年)平均浓度(通量)。

根据多年来水质监测实践,河流水体中氨氮和高锰酸盐指数以及河流流量是有代表性的敏感指标。但考虑到淮河和颍河的水文特性,丰枯期流量变化太大,因此,本次研究确定水质常规监测采样频

率的方法,只采用氨氮和高锰酸盐指数两个水质项目,计算各种采样频率下,氨氮、高锰酸盐指数相对每日采样1次(作为高采样频率)浓度和通量月平均以及年平均相对偏差范围的数据所占的比例,计算结果详见表1和表2。

2.1 月平均相对偏差分析

从表1中可以看出,与每日1次监测结果比较,其月平均相对偏差小于20%的氨氮、高锰酸盐指数浓度和通量相对偏差范围的数据所占比例:

a. 蚌埠闸上。每5日采样1次的分别为90%、100%和91%、100%;每10日采样1次的分别为78%、98%和80%、98%;每15日采样1次的月分别为70%、98%和70%、96%;每月采样1次的分别为61%、93%和61%、95%;每两月采样1次的分别为41%、68%和41%、68%。

b. 石头埠。每5日采样1次的分别为92%、100%和91%、100%;每10日采样1次的分别为85%、95%和86%、96%;每15日采样1次的月分别为76%、92%和78%、91%;每月采样1次的分别为

表1 各种采样频率下相对每日采样1次浓度与通量月平均相对偏差范围所占比例统计 %

段面	采样频率	相对偏差 ≤ 5%				相对偏差 ≤ 10%				相对偏差 ≤ 15%				相对偏差 ≤ 20%			
		NH ₃ -N	COD	NH ₃ -N 通量	COD 通量	NH ₃ -N	COD	NH ₃ -N 通量	COD 通量	NH ₃ -N	COD	NH ₃ -N 通量	COD 通量	NH ₃ -N	COD	NH ₃ -N 通量	COD 通量
蚌埠 闸上	3日1次	50	88	51	87	77	100	76	100	85	100	88	100	95	100	93	100
	5日1次	48	73	47	71	72	98	73	98	80	100	79	100	90	100	91	100
	10日1次	35	68	38	64	55	85	58	88	70	96	72	96	78	98	80	98
	15日1次	30	61	31	60	48	77	52	82	62	92	64	93	70	98	70	96
	每月1次	25	51	28	48	42	68	45	72	55	88	54	90	60	93	61	95
	两月1次	12	25	14	28	18	50	20	51	25	60	26	61	41	68	41	68
石头 埠	3日1次	51	91	50	90	88	99	89	99	93	100	92	100	97	100	96	100
	5日1次	42	82	40	83	83	96	81	95	89	98	87	98	92	100	91	100
	10日1次	35	70	32	68	78	88	75	89	84	91	82	91	85	95	86	96
	15日1次	25	60	26	61	61	82	63	80	65	84	71	82	76	92	78	91
	每月1次	23	55	22	55	51	74	50	76	56	82	55	84	60	88	62	90
	两月1次	17	20	17	20	20	52	20	51	27	58	27	58	45	67	45	68

注:表中相对偏差值为绝对值。

表2 各种采样频率下相对每日采样1次浓度与通量年平均相对偏差范围所占比例统计 %

段面	采样频率	相对偏差 ≤ 5%				相对偏差 ≤ 10%				相对偏差 ≤ 15%				相对偏差 ≤ 20%			
		NH ₃ -N	COD	NH ₃ -N 通量	COD 通量	NH ₃ -N	COD	NH ₃ -N 通量	COD 通量	NH ₃ -N	COD	NH ₃ -N 通量	COD 通量	NH ₃ -N	COD	NH ₃ -N 通量	COD 通量
蚌 埠 闸 上	3日1次	75	100	75	100	100	100	100	100	100	100	100	100	100	100	100	100
	5日1次	25	100	25	100	75	100	75	100	75	100	75	100	100	100	100	100
	10日1次	25	100	25	100	25	100	25	100	75	100	75	100	75	100	75	100
	15日1次	25	100	25	100	50	100	50	100	75	100	75	100	75	100	75	100
	每月1次	50	50	50	50	50	75	50	75	75	100	75	100	100	100	100	100
	两月1次	25	50	25	50	50	75	50	75	50	75	50	75	50	75	50	75
石 头 埠	3日1次	60	100	60	100	100	100	100	100	100	100	100	100	100	100	100	100
	5日1次	60	100	60	100	80	100	80	100	100	100	100	100	100	100	100	100
	10日1次	60	60	60	60	80	100	80	100	100	100	100	100	100	100	100	100
	15日1次	20	80	20	80	40	100	40	100	100	100	100	100	100	100	100	100
	每月1次	40	20	40	20	60	40	60	40	60	80	60	80	80	100	80	100
	两月1次	0	40	0	40	0	40	0	40	20	40	20	40	40	80	40	80

注:表中相对偏差值为绝对值。

60%、88%和62%、90% ;每两月采样1次的分别为45%、67%和45%、68%。

2.2 年平均相对偏差分析

从表2中可以看出,与每日1次监测结果比较,其年平均相对偏差小于20%的氨氮、高锰酸盐指数浓度和通量相对偏差范围的数据所占比例:

a. 蚌埠闸上。每5日采样1次的分别为100%、100%和100%、100% ;每10日采样1次的分别为75%、100%和75%、100% ;每15日采样1次的月分别为75%、100%和75%、100% ;每月采样1次的分别为75%、100%和75%、100% ;每两月采样1次的分别为50%、75%和50%、75%。

b. 石头埠。每5日采样1次的分别为100%、100%和100%、100% ;每10日采样1次的分别为100%、100%和100%、100% ;每15日采样1次的月分别为60%、70%和61%、68% ;每月采样1次的分别为80%、100%和80%、100% ;每两月采样1次的分别为40%、80%和40%、80%。

通过对表1、表2中的相对偏差范围的数据所占比例进行综合分析可知,每月采样3次、6次与每日采样1次的水质指标浓度和通量月平均、年平均都最为接近,每月采样2次与每日采样1次的水质指标浓度和通量月平均的相对偏差稍大一些,年平均的水质项目浓度和通量较为一致,每月采样1次的水质项目浓度和通量月平均、年平均的相对偏差都稍微大一些,每两月采样1次的水质项目浓度和通量月平均、年平均的相对偏差都较大,代表性较差。

一般情况下,将月平均浓度和通量的相对偏差控制在20%以内,水质指标的浓度和通量相对偏差范围的数据所占比例达到70%以上时,基本上能满足评价水质季节性变化或年际变化的实际需要。因此,若要考虑掌握水质季节性变化或年际变化情况,常规水质监测的采样频率可确定为每月1~2次,并辅之在颍河、涡河大量下泄重污染水时进行的水污染紧急状态下动态变化监测,既能达到目的,还可以节省大量的投入,这是较为现实的采样频率。若以每两月采样1次的采样频率进行监测则达不到水质监测的目的和要求。

3 统计学方法

用偏差分析的方法确定水质常规监测的采样频率虽然较为直观和合理,但要在每个采样断面较长时间地采集高采样频率的水质资料是不现实的。水质是天然的现象和人类活动影响的函数,一般影响水质变化的这两种因素在某种程度上受机遇法则的支配,这一点对由天然水文循环对水质变化带来的

影响尤为可信。一个流域或区域在一定的时期内,工农业生产和人口增长是随机的,正常情况下人类活动对水质的影响也是随机的,因此,可以利用有限的历史水质监测资料,用统计学方法来估算水质常规监测的采样频率。

3.1 水质常规监测数据正态性检验

根据GB4882—85《数据的统计和处理正态性检验》方法,对淮河流域和长江流域典型监测断面1998~2003年的水质常规监测数据进行正态性检验,以判断水质监测断面处常规水质监测情况下代表性水质指标数据是否来自于正态总体。

正态性检验是一种特殊的假设检验,其原假设为: H_0 总体为正态分布。

3.1.1 检验方法的选择

在确认总体是否为正态分布的正态性检验中,分有方向的检验和无方向的检验。在进行确认总体是否为正态分布的正态性检验时,应使用无方向检验方法。无方向检验有W检验(SHAPIRO-WILK检验, $30 \leq n \leq 50$)和D检验(D'AGOSTINO检验, $50 \leq n \leq 1000$)两种。

3.1.2 水质代表性项目选择

水质总体信息由抽取的样本信息获得,水质总体是否服从正态分布,也只能从水质样本信息中得到。通过对水质样本中的数据进行正态性检验,即可在一个很大的概率下确认水质样本数据是否来自于正态总体。根据多年来水质监测实践,本次研究的河段中氨氮和高锰酸盐指数具有代表性,因此选用1998~2003年的常规水质监测数据中的氨氮和高锰酸盐指数这两个指标进行正态性检验。

对长江流域青弋江弋江站1998~2003年常规水质监测资料进行正态性检验时,由于该站氨氮和高锰酸盐指数含量极低,氨氮含量大多数在检出限以下,这两个指标已不具有代表性,因此改由综合指标总碱度和总硬度作为代表项目较为合理。

3.1.3 正态性检验

本次用于检验的水质常规监测断面的样本数均大于50,小于1000,在进行确认常规水质监测数据是否来自于正态总体的正态性检验时,选用D检验法进行正态性检验。正态性检验步骤如下:

a. 将观测值按非降次序排列成:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

b. 按公式

$$D = \frac{\sum_{k=1}^n \left(k - \frac{n+1}{2} \right) x_{(k)}}{(\sqrt{n})^2 \sqrt{\sum_{k=1}^n [x_{(k)} - \bar{x}]^2}} \quad (1)$$

和

$$Y = \frac{\sqrt{n}(D - 0.28209479)}{0.02998598} \quad (2)$$

计算统计量 Y 的值。

c. 根据 α 和 n 查 D 检验统计量 Y 的 P 分位数 Z_α 表, 得 $Z_{\frac{\alpha}{2}}$ 和 $Z_{1-\frac{\alpha}{2}}$ 。

d. 作出判断: 若 $Y < Z_{\frac{\alpha}{2}}$ 或 $Y > Z_{1-\frac{\alpha}{2}}$, 则拒绝 H_0 。若 $Z_{\frac{\alpha}{2}} \leq Y \leq Z_{1-\frac{\alpha}{2}}$, 则不拒绝 H_0 。

在 n 较小时, 为便于列表计算, 统计量 D 的分子可以换成与其相等的形式 $\sum_{k=1}^l \left(\frac{n+1}{2} - k \right) [x_{(n+1-k)} - x_{(k)}]$ 其中 n 为奇数时, $l = \frac{n-1}{2}$ 。

本次课题对研究区域内的淮河干流、颍河干流的王家坝、鲁台子、田家庵、蚌埠闸上、界首上、阜阳闸上、颍上闸上水质监测断面, 以及长江流域青弋江弋江站水质监测断面 1998~2003 年的常规水质监测数据进行正态性检验, 其检验结果详见表 3。

表 3 正态性检验结果

河流	断面	n	NH ₃ -N		是否拒绝	COD _{Mn}		是否拒绝
			D	Y	H_0	D	Y	H_0
淮 河	王家坝	81	0.2701	-3.614	不拒绝	0.2824	-0.100	不拒绝
	鲁台子	70	0.2310	-14.261	拒绝	0.2579	-6.710	拒绝
	田家庵	64	0.2862	1.090	不拒绝	0.2796	-0.690	不拒绝
	蚌埠闸上	69	0.2629	-5.322	拒绝	0.2770	-1.425	不拒绝
颍 河	界首上	67	0.2784	-1.003	不拒绝	0.2731	-2.290	不拒绝
	阜阳闸上	68	0.2715	-2.926	不拒绝	0.2736	-2.355	不拒绝
	颍上闸上	69	0.2587	-6.468	拒绝	0.2754	-1.850	不拒绝

从表 3 中可以看出, 本次共对 4 个水质项目 16 组水质数据进行了正态性检验, 其结果是不拒绝原假设 H_0 的有 12 组, 占总检验组数的 75%, 拒绝原假设 H_0 的有 4 组, 占总检验组数的 25%。在被拒绝的总体为正态分布的原假设的 4 组数据中, 鲁台子两组水质数据全被拒绝了正态分布原假设, 根据分析发现, 其主要原因是由于颍河高浓度污水集中下泄的直接影响所至。同样, 颍上闸上监测断面氨氮数据被拒绝总体为正态分布的原假设, 是受阜阳闸下积累的高浓度污水集中推下所至; 蚌埠闸上监测断面氨氮数据被拒绝总体为正态分布的原假设, 是受蒙城闸下积累的高浓度污水集中推下所至。

通过以上检验和分析可以看出, 如果水质监测断面不是受到人为活动的剧烈影响, 可以认为水质样本数据来自于正态总体。也就是说, 在正常情况下, 水质总体服从正态分布, 即水质常规监测的水质数据来自于正态总体。

3.2 统计计算基本方法

根据以上分析, 正常情况下, 水质总体服从正态

分布。由于水质常规监测是随机抽样过程, 其水质采样频率可根据监测断面处代表性水质项目的含量大小, 用统计学的方法加以确定。水质常规监测数据均值的置信区间宽度可由下式给出:

$$[\bar{x} - Z_{\frac{\alpha}{2}} \text{va}(\bar{x}), \bar{x} + Z_{\frac{\alpha}{2}} \text{va}(\bar{x})] \quad (2)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

$$\text{va}(\bar{x}) = \frac{\sigma^2}{n} \quad (4)$$

式中 \bar{x} 是样本的均值; x_i 为第 i 个测定值; σ^2 为总体方差; n 为样本数。

3.2.1 单站单水质项目

确定水质采样频率的最简单情况, 应是选择在指定的水质站对一个指定的水质项目的年均值(或几何均值)期望置信区间宽度中产生的采样频率。如果所采集的样本能够假定是独立的, 那么样本均值的方差表达式由式(4)给出。一般说来, 总体方差 σ^2 是未知的, 但是能根据现有资料估算方差 s^2 。作为真实均值 μ (期望值) 的范围, 可由公式(5)表示:

$$\bar{x} - Z_{\frac{\alpha}{2}} \text{va}(\bar{x})^{\frac{1}{2}} \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}} \text{va}(\bar{x})^{\frac{1}{2}} \quad (5)$$

用 σ^2/N 来代替 $\text{va}(\bar{x})$ (这里 N 是用于计算 σ^2 的样本数), 并重新整理各项。由于真实总体均值 μ 实际上难以得到, 而总体均值 μ 落在一定置信水平下的置信区间范围内, 因此, 均值的未来估计值所需要的样本数 n , 可以利用已确定的置信水平 p 的置信区间宽度 ΔL , 用式(6)计算得到:

$$n \geq \left[\frac{t_{\frac{\alpha}{2}} s}{\Delta L} \right]^2 \quad (6)$$

式(6)中 ΔL 则根据下式确定:

$$\Delta L = \frac{2t_{\frac{\alpha}{2}} s}{\sqrt{N}} \quad (7)$$

式(7)中 $t_{\frac{\alpha}{2}}$ 值由自由度(与水质样本数有关)和给定的置信水平 p 确定; 标准差 s 可由贝塞尔公式得到:

$$s(x_k) = \sqrt{\frac{1}{N-1} \sum_{k=1}^n (x_k - \bar{x})^2} \quad (8)$$

在式(6)中, 确定 n 需要 3 个参数: 合适的置信水平 p (用于确定统计量 t 值), 置信区间宽度 ΔL 和总体方差 σ^2 的估计值 s^2 。 s^2 可由已有的水质资料计算得出, 此时, n 的计算则成为迭代问题。由于不知道 t 值, 就不能确定 n 值, 因此要解决这个问题, 就必须先要估计一个 n 值, 以便确定自由度, 再确定一个合适的置信区间宽度 ΔL 的目标值, 使用式(6)计算 n 的第 1 次近似值。在初始估计值和第 1 次近似值之间选择新的 n 值, 但要较接近于新的 n 值, 然后再计算第二次近似值。重复此步骤到几乎

相等为止。一般进行 3 次试算就足够了。

本次水质采样频率研究中的基础资料,采用国务院要求淮河流域 1997 年底工业污染源达标排放后 1998 年至 2003 年的水质监测资料,计算设计采样频率,其水质资料质量可靠,可比性强,数量充足,各河段采样断面的水质样本数都大大超过了 50 个,因此本次研究水质常规监测采样频率可以直接用式(6)计算确定。

3.2.2 单站多水质项目

确定单站多水质项目采样频率的一个简单方法是利用单站单水质项目,分别计算每个水质项目的单独采样频率,并且简单地求它们的平均值,获得合适的年采样频率。

确定单站多水质项目采样频率的另一个方法是在采样频率范围内,计算几个项目的标准差和置信区间宽度的加权平均值,然后选择产生最理想结果的采样频率。

3.3 水质常规监测年采样频率的确定

3.3.1 算术平均法

安徽省淮河蚌埠闸以上干流和颍河干流现有各自独立的水质监测站 12 个,其中淮河干流 6 个,颍河干流 6 个,现利用式(6)对安徽省淮河蚌埠闸以上干流和颍河干流各水质监测站的年采样频率进行计算,然后将几个水质项目的年采样频率进行简单的算术平均,即得到每个水质监测站的年采样频率。由于所得到的水质数据样本较大,计算式中 $t_{\frac{\alpha}{2}}$ 值取 2,对于小样本的则根据 95% 的置信水平和自由度 $(N-1)$ 确定 $t_{\frac{\alpha}{2}}$ 值。 $\Delta L = 2t_{\frac{\alpha}{2}}s/\sqrt{N}$ 。计算各站各水质项目年采样频率所得结果见常规水质监测站年采样频率计算表。从表 4 中可以看出,氨氮与高锰酸盐指数的年采样频率基本一致。

表 4 常规水质监测站年采样频率计算

河流	断面	置信区间宽度 ($\text{mg}\cdot\text{L}^{-1}$)		单项采样 频率/ n		年采样 频率/ \bar{n}	
		$\text{NH}_3\text{-N}$	COD_{Mn}	$\text{NH}_3\text{-N}$	COD_{Mn}	算术 平均法	加权 平均法
淮 河	王家坝	1.253	2.540	18	10	14	13
	鲁台子	0.878	0.878	18	18	18	18
	凤台大桥	1.724	0.956	11	11	11	11
	田家庵	1.097	0.608	18	18	18	18
	大涧沟	1.067	0.523	14	14	14	14
	蚌埠闸上	0.792	0.481	18	18	18	18
颍 河	界首上	6.760	4.494	18	18	18	18
	界首下	6.789	4.339	18	18	18	18
	阜阳闸上	5.171	2.362	18	18	18	18
	阜阳闸下	4.723	8.562	18	18	18	18
	颍上闸上	3.636	2.644	18	18	18	18
	范台子	8.520	4.029	13	13	13	13

3.3.2 加权平均法

确定单站多水质项目年采样频率 \bar{n} 的加权平均的方法,即是计算各站几个项目的加权平均标准差和加权平均置信区间宽度 ΔL 的值,然后再计算各站的年采样频率。

从表 4 中可以看出,以上两种方法计算的各站年采样频率基本一致,只有两个站的年采样频率有很小的差别。

通过对安徽省清流河滁州市区监测断面和青弋江弋江监测断面采样频率的计算,说明本课题研究的确定常规水质监测采样频率的方法不仅适合于淮河,也适用于其他河流。

4 结 语

在水质监测站网设计中,水质采样频率的确定是一个比较复杂的过程,并不是计算某一两个指标就能完全定量地反映和确定水质监测的采样频率,如监测条件、经费和管理情况等,都可能直接影响水质监测的采样频率的确定。但是就本次课题研究的成果看,水质监测的采样频率一般可定为每月 1~2 次,并辅助以水污染紧急状态下进行水质动态变化过程的监测即可满足一般管理的需要,若采样频率过高,造成浪费,过低则达不到监测目的。但在一些特殊需要或特殊地区(如敏感的水源地、边远地区等),其采样频率可视情况增减。

参考文献:

- [1] Sanders T G. 水质监测站网设计[M]. 金立新译. 南京: 河海大学出版社, 1989.
- [2] 中山大学数学力学系. 概率论及数理统计[M]. 北京: 人民教育出版社, 1980.
- [3] Kottegoda N T. 随机水资源技术[M]. 金光炎译. 北京: 农业出版社, 1987.
- [4] 四川省环境保护科研监测所. ISO 数理统计方法标准译文集[M]. 成都: 四川科学技术出版社, 1984.
- [5] 邓英春, 金传良, 李怡庭, 等. 水质化学分析计算指南[M]. 合肥: 安徽科学技术出版社, 1993.
- [6] SL219—98 水环境监测规范[S].

(收稿日期 2004-12-02 编辑:傅伟群)