

(9) 河川の確率水質推定に関する2, 3の検討

SEVERAL CONSIDERATIONS ON ESTIMATION OF
PROBABILISTIC WATER QUALITY OF A STREAM

河原 長美*

Osami KAWARA*

ABSTRACT: Compliance with water quality standards for a stream is evaluated by comparing the probabilistic water quality with the given stream standards. In Japan, the probabilistic water quality corresponding to cumulative probability of 75% is used for the above evaluation. At first, factors giving effect to the accuracy of probabilistic water quality estimated from limited data are discussed theoretically. As the above factors, the population parameters of a water quality variable, cumulative probability and a sample number are picked up. For the case of a water quality variable distributed log normal the confidence limits are nearly determined by σ^2/n and the selected cumulative probability, where σ^2 and n are the variance of logarithm of a water quality variable and a sample number respectively. Secondly, the estimation method used by a water quality management agency is discussed based on observations. The confidence limits of the estimated values are shown comparing with the observations.

KEYWORDS; water quality , stream , monitoring , standard compliance , statistics

1. はじめに

公共水域の水質環境基準には、人の健康の保護に関する環境基準と生活環境の保全に関する環境基準とがある。後者の環境基準においては、通常の状態において基準値を満足することとされており、河川については低水量以上の流量がある状態が対象とされている。

環境基準の適合性は上述の通常の状態下において、基準値を満たしているデータが75%以上あるかどうかで判定されている。具体的には、次に述べる75%水質値と環境基準値との大小比較により判定されている。75%水質値とは、年間の日間平均値の全データをその値の小さいものから順に並べ、 $0.75 \times n$ 番目 (n : データ数) の値、もしくは、 $0.75 \times n$ が整数でない場合は、端数を切り上げた整数番目の水質値を意味している。

本研究では、限られたデータから推定される非超過確率水質値の確率分布関数を、順序統計の理論に基づいて誘導し、理論式により推定水質の推定精度について検討するとともに、旭川における観測値に基づいて、75%水質値の意味や現行判定法の精度を明らかにすることを試みる。

2. 流域の概要と解析に用いたデータ

Fig.1 に旭川流域と採水地点とを示す。採水地点は、Fig.1 において白丸印で示される旭川非感潮部下流端に位置する兵団地点であった。

旭川は、流域面積 1800km²、平水量約40m³/s の一級河川であり、流域内人口密度が低く、人為的汚染の

* 岡山大学 Okayama University

比較的少ない河川であり、水質は比較的清浄である。
旭川には2基のダムがあり、流量が比較的安定している。

解析に用いたデータは、昭和55年11月より昭和56年10月までの12ヵ月分の定時観測データである。水質項目は、SS（ガラスファイバーろ紙法）、濁度（吸光度法）、COD(Mn)（酸性法）、TN（還元法）、および、TP（混合試薬法）であった。定時観測データにおける欠測については、別途求めておいた流送能力・供給関数型モデルを用いて推定した。

3. 解析方法¹⁾

確率密度関数 $f(x)$ に従う母集団から得られた n 個の標本を考える。今、 n 個の標本のうち、小さい側から k 番目の値を u 、 $k+1$ 番目の値を v とすると ($u \leq v$)、トーマスプロットによれば、 u 、 v それぞれに対応する非超過確率は、 $k/(n+1)$ および $(k+1)/(n+1)$ となる。

非超過確率 $P\%$ は、 $k/(n+1) \leq P/100 \leq (k+1)/(n+1)$ を満足するものとし、 $P\%$ に

相当する水質値 w を、 u 、 v の値およびそれぞれに対応する非超過確率値から求めるものとする。

上記の不等式において、等号が成立するか否かで式の形が異なってくるので、2つの場合に分けて検討する。

3.1 小さい側から k 番目と $k+1$ 番目の間に非超過確率 $P\%$ に相当する値 w が存在する場合

非超過確率 $P\%$ に相当する値 w を内挿により求めるものとし、 $w = h(u, v, n, P)$ とする。

非超過確率 $P\%$ に相当する値が w になる確率 $g(w)dw$ は、内挿関数 h を定めると、小さい側から k 番目および $k+1$ 番目の値が、それぞれ、 u 、 v になる同時確率に等しい。ところで、互いに素な領域 $I_1 \sim I_5$ に分割し、それぞれの領域に属する確率を $P_1 \sim P_5$ で表わすと次のようになる。

$I_1 = (-\infty, u)$	$P_1 = F(u)$	$k-1$ 個が属する
$I_2 = [u, u+du)$	$P_2 = F(u+du) - F(u) \doteq f(u) du$	1 個が属する
$I_3 = [u+du, v)$	$P_3 = F(v) - F(u+du) \doteq F(v) - F(u)$	0 個が属する
$I_4 = [v, v+dv)$	$P_4 = F(v+dv) - F(v) \doteq f(v) dv$	1 個が属する
$I_5 = [v+dv, \infty)$	$P_5 = 1 - F(v+dv) \doteq 1 - F(v)$	$n-k-1$ 個が属する

ここに、 $F(x) = \int_{-\infty}^x f(x) dx$

それ故、求める確率 $g(w)dw$ は次のように表わされる。

$$g(w)dw = \frac{n!}{(k-1)!(n-k-1)!} F(u)^{k-1} (1-F(v))^{n-k-1} f(u)f(v)dudv. \quad \text{---(1)}$$

ところで、ある一組の u 、 v の値、 u_1 、 v_1 の値を用いて内挿関数 $h(u_1, v_1, n, P)$ より求める w の値を w_1 とおき、着目している w の値を w_0 とすると、 $w_1 \leq w_0$ となる確率は、 w の分布関数を用いると、 $G(w_0) (= \int_{-\infty}^{w_0} g(w)dw)$ に等しい。今、 $-\infty < u \leq w_0$ なる任意の u_1 が与えられた時に、 $w_1 \leq w_0$ を満足するためには v_1 は、 $u_1 \leq v_1 \leq v_*$ (u_1, w_0, n, P) を満足する必要がある。ここに、 v_* (u_1, w_0, n, P) は、 $w_0 = h(u_1, v_1, n, P)$ を v_1 に関して解いた値である。それ故、 $G(w_0)$ は、次のように表わされる。なお、 w_0 については添字を省略した。

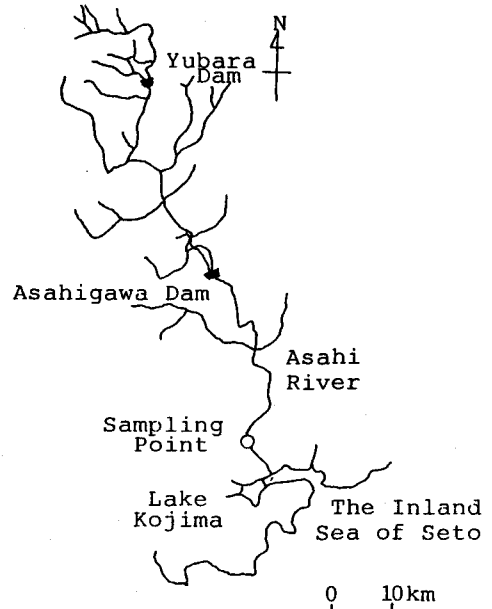


Fig.1 Sampling point

$$G(w) = \frac{n!}{(k-1)!(n-k-1)!} \int_{-\infty}^w F(u)^{k-1} f(u) du \int_{u_1}^{v_*} (1-F(v))^{n-k-1} f(v) dv$$

$$= \frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^w F(u)^{k-1} f(u) \{ (1-F(u_1))^{n-k} - (1-F(v_*))^{n-k} \} du \quad \text{---(2)}$$

3.2 小さい側からk番目が非超過確率P%に相当する場合

n個のデータが存在するとき、トーマスプロットに従うと、小さい側からk番目の値は、 $P = 100k/(n+1)$ なる非超過確率を有する。k番目の値の非超過確率がP%であるので、小さい側からk番目の値がwになる確率が、とりもなおさず 非超過確率P%に相当する値がwになる確率である。

(1)の場合と同様に、 $I_1 \sim I_3$ の互いに素な領域に分割し、それぞれの領域に属する確率を $P_1 \sim P_3$ とすると次のようになる。

$$\begin{aligned} I_1 &= (-\infty, w) & P_1 &= F(w) & k-1 \text{ 個が属する} \\ I_2 &= [w, w+dw) & P_2 &= F(w+dw) - F(w) \doteq f(w) dw & 1 \text{ 個が属する} \\ I_3 &= [w+dw, \infty) & P_3 &= 1 - F(w+dw) \doteq 1 - F(w) & n-k \text{ 個が属する} \end{aligned}$$

求める確率 $g(w)dw$ は

$$g(w)dw = \frac{n!}{(k-1)!(n-k)!} F(w)^{k-1} (1-F(w))^{n-k} f(w)dw$$

$$\therefore g(w) = \frac{n!}{(k-1)!(n-k)!} F(w)^{k-1} (1-F(w))^{n-k} f(w) \quad \text{---(3)}$$

4. 非超過確率水質値の推定精度

4.1 水質値の確率分布

式(2)もしくは式(3)を用いると、非超過確率水質値の推定値の分布を理論的に検討することができる。ただし、式(2)および式(3)では、対象としている水質項目の確率分布が既知であることを前提としている。そこで、最初に水質項目が従う確率分布について検討を加える。

旭川における定時観測値を基に、対数正規分布、ピアソンⅢ型分布(1母数、2母数、3母数)、対数ピアソンⅢ型分布の各分布を取り上げ検討した結果、Kolmogorov-Smirnov検定により5%水準で帰無仮説が否定されるのは、濁度、COD、TNの場合である。帰無仮説が否定された分布だけでなく、帰無仮説が否定されない場合については、適合性が最も良好な分布を取り上げるならば、Table 1に示すような結果が得られた。なお、TNについては、対数正規分布も帰無仮説が否定されたが、ピアソンⅢ型2母数分布のほうが、適合性が良好であった。

以上の結果より、旭川の定時観測値の場合、対数正規分布は多くの水質項目について適合する分布と言えよう。また、対数正規分布に従う水質項目が多く存在することは従来から認められているところであるので²⁾、以下の検討においては、対数正規分布を対象として検討を加える。

4.2 非超過確率水質の推定値の確率分布

Table 1 Probability density function

Q	$f(x) = \frac{1}{0.8126x(2\pi)^{1/2}} \exp\left\{-\frac{(\ln x - 3.711)^2}{2 \times 0.6603}\right\}$
SS	$f(x) = \frac{1}{1.039x(2\pi)^{1/2}} \exp\left\{-\frac{(\ln x - 1.570)^2}{2 \times 1.078}\right\}$
TUR	$f(x) = \frac{1}{0.9807x(2\pi)^{1/2}} \exp\left\{-\frac{(\ln x - 2.307)^2}{2 \times 0.9617}\right\}$
COD	$f(x) = \frac{1}{0.6261x(2\pi)^{1/2}} \exp\left\{-\frac{(\ln x - 0.5749)^2}{2 \times 0.3920}\right\}$
TP	$f(x) = \frac{1}{0.7460x(2\pi)^{1/2}} \exp\left\{-\frac{(\ln x + 3.586)^2}{2 \times 0.5565}\right\}$
TN	$f(x) = \frac{1}{0.2270x\Gamma(2.936)} \left(\frac{x}{0.2272}\right)^{1.936} \exp\left\{-\left(\frac{x}{0.2272}\right)^{2.936}\right\}$

非超過確率水質の推定値に関して、詳細な検討を加える前に、式(2)および式(3)に基づき推定水質値の確率分布の特徴を検討しておく。Table 1 に示したSSの場合について、非超過確率水質の推定値の分布関数ならびに確率密度関数をFig. 2 およびFig. 3 に示す。なお、この場合、線形な内挿関数 $w = au + bv$ ($a = 1 - P(n+1)/100 + [P(n+1)/100]$, $b = 1 - a$,ここに、 $[]$ はガウスの記号)を用いている。

なお、本研究では、水質の自己相関を無視した解析を行なっているが、旭川における従来の結果によれば³⁾、多くの場合、3日程度で相関性がなくなる。

Fig. 2 およびFig. 3 より、データ数が多くなると推定値の分布が集中し、正規分布に近くなることや、データ数の異なる各分布関数が必ずしも同じ累積確率の点で交わらないことがうかがえる。

4.3 非超過確率水質の推定値の信頼区間

(A) 実測値から直接求めた信頼区間と理論的に求めた信頼区間の比較

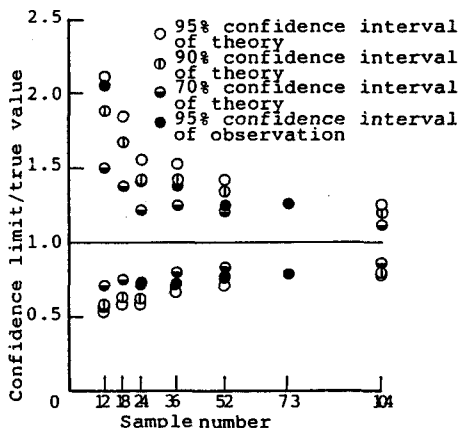


Fig. 4 Comparison of confidence interval for SS between theory and observation

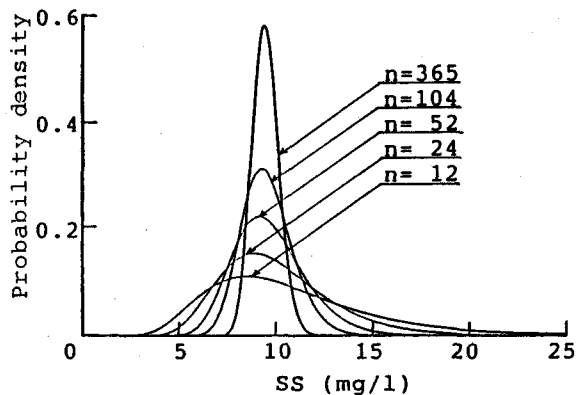


Fig. 2 Probability density functions of a sample estimate for SS corresponding to 75% cumulative probability

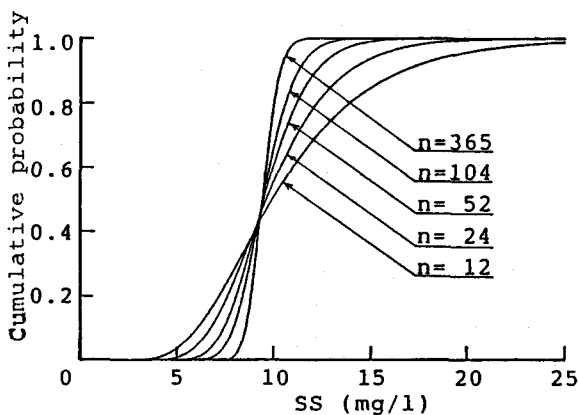


Fig. 3 Cumulative distribution functions of a sample estimate for SS corresponding to 75% cumulative probability

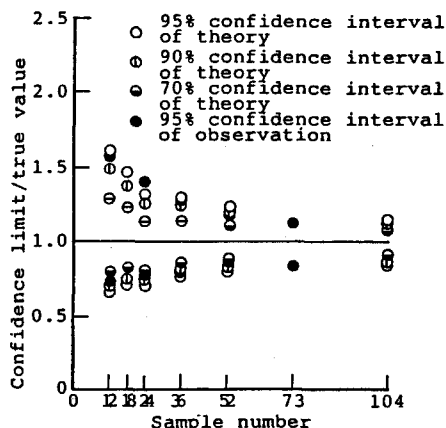


Fig. 5 Comparison of confidence interval for COD between theory and observation

対数正規分布を仮定して、(2)式から算出される信頼区間を、観測値から直接求めた信頼区間とを比較したのが、Fig. 4およびFig. 5である。観測値から直接求める際には、一年間の観測値から、等間隔サンプリングを仮定して得られる各観測値のグループ毎に、1個の非超過確率水質を求め、得られた非超過確率水質が、対数正規分布に従うことを仮定して、信頼区間を算出した。

Fig. 4は、分布の検定において帰無仮説が否定されなかったSSの場合であり、Fig. 5は帰無仮説が否定されたCODの場合である。どちらの図においても、実測値から直接推定された信頼区間は、(2)式に基づいて算出された信頼区間よりも幾分小さいようであるが、ほぼ一致しており、対数正規分布を仮定して

(2)式を用いることにより、非超過確率水質の推定値の分布を推定できるものと考えられる。

(B) 信頼区間に影響を及ぼす因子

対数正規分布で近似できる場合、水質値を x として、 $y = \ln x$ なる y に対する平均および分散を μ_y 、 σ_y^2 とすると、信頼区間の真値に対する比に影響を及ぼすのは、 σ_y^2 のみであり、 μ_y はほとんど影響を及ぼさない。そこで、データの個数を5種類に変化させて、 σ_y^2 の、信頼区間に対する影響を検討したのがFig. 6である。Fig. 6には、非超過確率75%に対する推定水質の95%信頼区間が示されている。なお、旭川での観測水質の場合、現行判定法にいうところの通常の状態において、 σ_y^2 の値は0.2~0.8程度であった。

Fig. 6よりうかがえるように、データ個数 n と σ_y^2 は、信頼区間区間に大きな影響を及ぼす。 σ_y^2 が大きくなると信頼区間が広くなり、 n が大きくなると信頼区間が狭くなるが、 σ_y^2/n が同じであれば、信頼区間はほぼ同じである。この様子をFig. 7に示す。なお、 n の値が増加するに従い、幾分信頼区間は下方に変化する。

ところで、流量 Q と水質 C とが、ともに対数正規

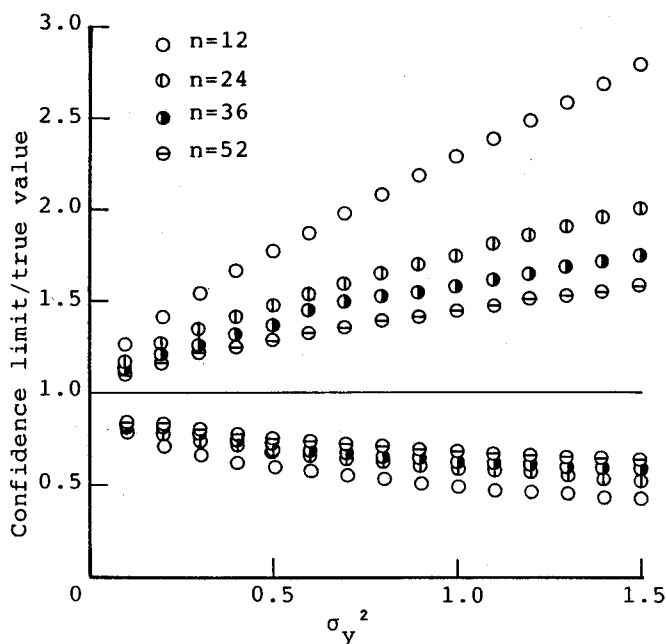


Fig. 6 Effect of variance on 95% confidence interval

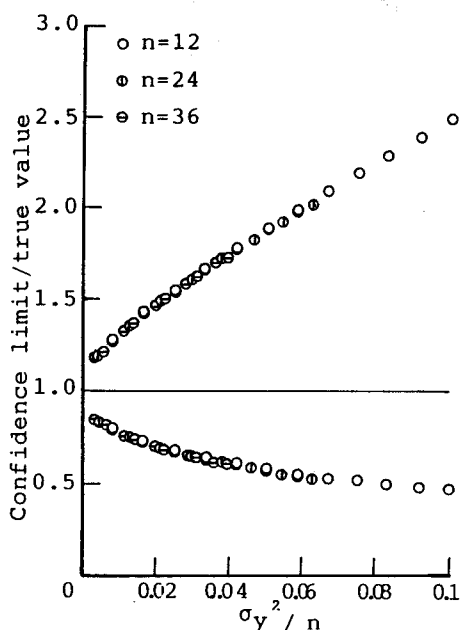


Fig. 7 Effect of σ_y^2/n on 95% confidence interval

分布で近似でき、かつ、 $C = aQ^b$ なる関係があるとする、 $\ln C = \ln a + b \ln Q$ が成立するから、 $\ln Q$ が $N(\alpha, \beta^2)$ に従うとすると、 $\ln C$ は $N(\alpha + \ln a, b^2 \beta^2)$ に従うことになり、流量に対する依存性を表わす b の大小により、水質間の相対的な σ_y^2 の大小関係を判定できる。

次に、信頼区間に対する非超過確率値とデータ数との影響を検討する。Fig. 8 に 1 例を示す。Fig. 8 は、95% 信頼区間が、非超過確率とデータ数とによって、どのように変化するかを示している。最初に確率値の影響を検討する。非超過確率 50% の場合の信頼区間を基準にすると、75% の場合は上限値が高くなり、25% の場合には下限値が低くなる。また、この傾向は、データ数が少ないほど顕著である。以上の結果は、データより直接算出した従来の結果⁴⁾ と同じ傾向を示すものであり、非超過確率が高くなると、信頼区間が高い側に移行することが理論的にも確認された。

データ数の増加による信頼区間の減少は 30~50 程度までが顕著であり、それ以後はゆるやかにしか減衰しない。

5. 水質環境基準の適合判定法の検討

5.1 現行判定法における 75% 水質値の信頼区間

水質環境基準の適合判定法においては、既述のように、全データを小さい側から順に並べ、 $n \times 0.75$ 番目もしくは、これが整数でない場合は、端数を切り上げた整数番目の水質値を 75% 水質値としている。このようにして推定された値の信頼区間を、トーマスプロットを用いた (2) 式もしくは (3) 式による理論値と対比して示したのが Fig. 9 である。なお、現行の適合判定法による 75% 水質値の推定値の分布は、(3) 式における k の値に上述の方法で算出される順位を代入することにより得られる。

Fig. 9 より、現行判定法の信頼区間に次のような特徴が読みとれる。第 1 に、現行判定法においては、データ数が $4\ell - 1$ (ℓ : 正整数) の場合には理論式と信頼区間が一致するが、それ以外の場合には、低い側の信頼区間を有する。これは、理論式においては、トーマスプロット ($i / (n + 1)$) により非超過確率を算定しているのに対し、現行の判定法では、整数番目の値を 75% 水質値とすることにより、結果的には、カリフォルニア法 (i / n) からトーマスプロットまで変化する方法で非超過確率を算出していることからである。すなわち、データ数が $4\ell - 1$ 個の場合のみ、トーマスプロットと一致し、

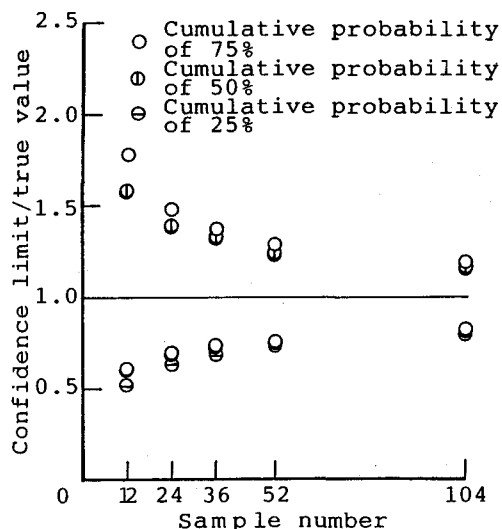


Fig. 8 Effect of cumulative probability on 95% confidence interval

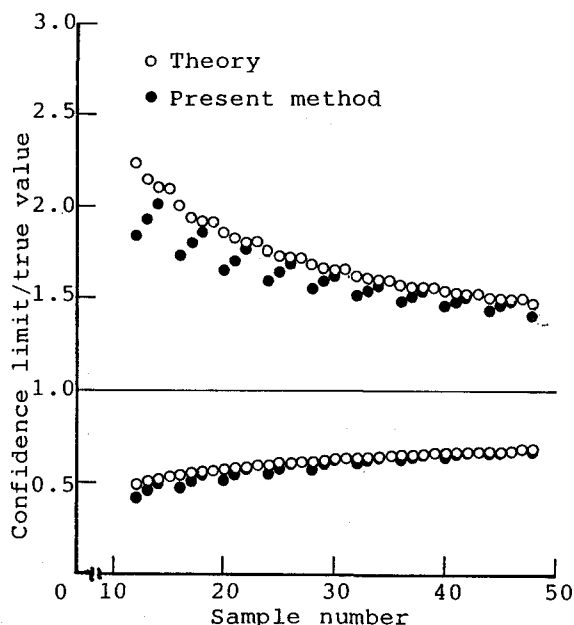


Fig. 9 Comparison of 95% confidence interval between theory and present method used by a water quality management agency

それ以外の場合には、トーマスプロットからすると、75%よりも低い非超過確率に相当する水質を算出していることになるからである。

第2に、現行判定法では、データ数の増加が単純に信頼区間の増加につながらず、データ数が4個の時に信頼区間が大きく減少し、そこからデータ数が増加すると、データ数とともに信頼区間が増加する。このことにも、上述のように非超過確率がデータ数とともに変化することが関係している。

第3に、現行の方法は、信頼区間が低い側に偏ってはいるが、上限値の差は大きく、下限値の差は小さいので、信頼区間としては、(2)式による値よりも狭くなっている。このことには、トーマスプロットで算出される非超過確率よりも、低い非超過確率に相当する値を算出していることの他に、50%を越える非超過確率の場合には、確率値が小さくなると、Fig. 8に示すように信頼区間の上限値のほうが大きく変化することが関係している。

5.2 75%水質値の検討

環境基準の適合判定に用いられている75%水質値は、河川の通常の状態を対象として算定するために、低水量以上の流量時における観測値に基づいて算定される。通常の状態という場合には、流量の上限値も問題になるが、必ずしも明確にはされていないので、ここでは、日平均流量が $400 \text{ m}^3/\text{s}$ 以上の場合（対象とする期間内に5日）も除いて、75%水質値について検討を加えた。

Fig. 10およびFig. 11に、SSとCODについて、流量との関係とともに、通常の状態における平均水質、270個のデータを用いて算出した75%水質値、および、現行判定法による推定値の信頼区間が示されている。75%水質値と平均水質値とは、ほぼ同様な値をとり、18個程度のデータによる信頼区間の中には、平均値も含まれている。

同じデータを用いて、水質および流量の非超過確率に対してプロットしたのが、Fig. 12およびFig. 13である。これらの図には、270個のデータを用いて算出した75%水質値、および、現行判定法による信頼区間が、365個のデータに対する非超過確率で示されている。

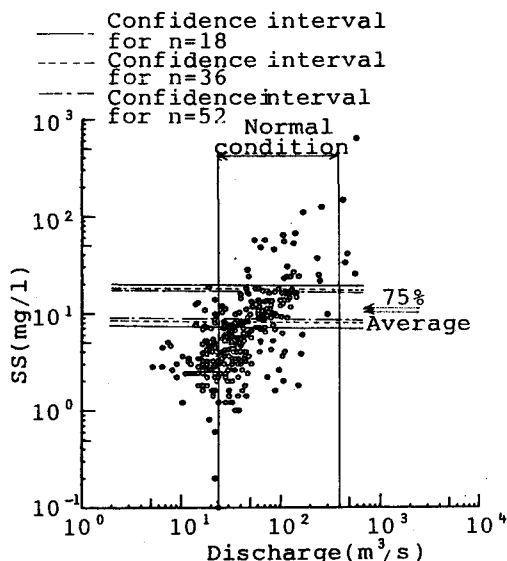


Fig. 10 Confidence interval of SS corresponding to 75% cumulative probability

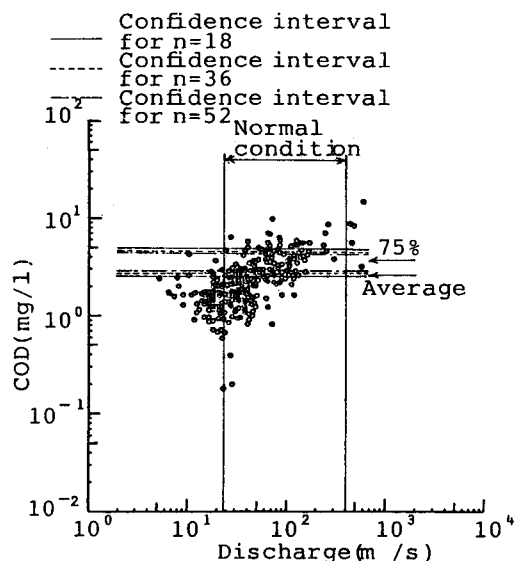


Fig. 11 Confidence interval of COD corresponding to 75% cumulative probability

非超過確率で表わすと、水質と流量との関係は不明瞭となる。75%水質値は、ここには示していない濁度、TPおよびTNの場合を含めて、全水質データの75~80%程度の非超過確率値に相当した。旭川の場合図からも明らかなように、低水量以下の場合、75%水質値を越えることはごくわずかであり、平水量ないし豊水量以上ある場合に、75%水質値を越えることが多い。

また、18個~52個のデータより75%水質値を算出する場合、ここに示されていない項目も含めて、下限値は、非超過確率60%弱~70%に、上限値は90%前後に相当した。なお、非超過確率で表わした信頼区間は、 σ_y^2 にはあまり影響を受けないが、 σ_y^2 の小さい場合のほうが幾分広くなる傾向も認められた。

6. まとめ

本研究では、水質環境基準の適合判定に用いられている確率水質をとりあげ、推定精度に影響を及ぼす因子、ならびに、現行判定法によって推定される75%水質値の特徴に関して検討を加えた。得られた主要な結論は次の通りである。

最初に、非超過確率水質の推定値が従う確率分布の理論式を誘導し、これによって、観測値より直接算出した推定値の信頼区間が、確率分布の検定に必ずしも合格しない場合も含めてほぼ説明できることを確認した上で、推定精度に影響を及ぼす因子について理論的に検討を加え、次の結果を得た。なお、本研究では、対数正規分布の場合のみを検討した。

1) 非超過確率水質の推定値に関する信頼区間は、データ数を n 、水質 C の対数値 y の分散を σ_y^2 とすると、 σ_y^2 / n によってほぼ決定される。また、流量を Q 、 a 、 b を定数として、 $C = a Q^b$ なる関数が認められる場合は、 σ_y^2 は b^2 に比例するので、 b の大小によって水質間の相対的な精度の目安が与えられる。

2) 非超過確率50%の推定水質の場合の信頼区間と比較して、非超過確率が高い場合は、信頼区間の上限値が高くなり、逆に非超過確率が低い場合には、信頼区間の下限値が低くなる。

ついで、河川における水質環境基準の、現行判定法の特徴を、旭川での観測値も用いて検討し、次の結果が得られた。

3) 現行の判定法により算定される75%水質値は、 l を整数として、 $4l - 1$ 個のデータ数の場合を除いて、トーマスプロットからすると、75%値よりも低い確率に相当する水質を求めていることになり、

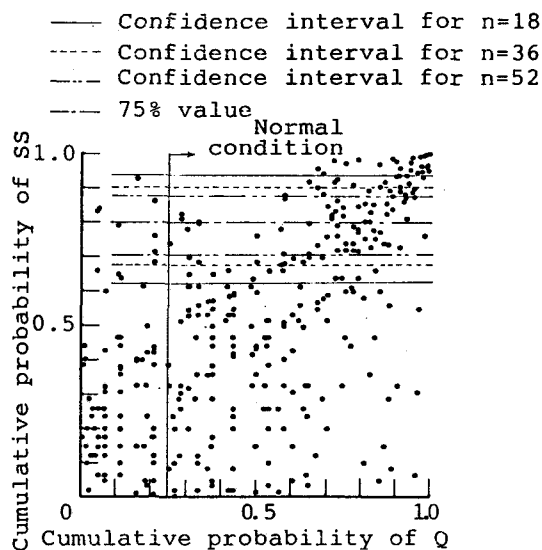


Fig. 12 Confidence interval of SS expressed by cumulative probability

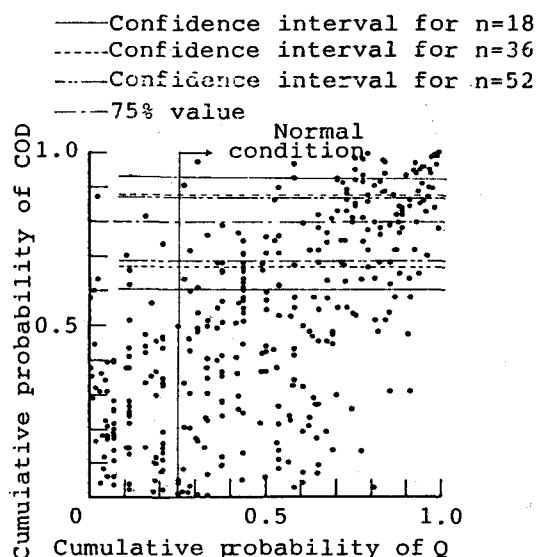


Fig. 13 Confidence interval of COD expressed by cumulative probability

そのため、信頼区間は、トーマスプロットと内挿とを用いて算定する場合よりも低い側に偏っているが、区間幅は狭い。なお、データ数が $4l-1$ 個の場合は、トーマスプロットを用いた場合と一致する。

- 4) 現行の方法では、データ数の増加が単純に信頼区間の減少につながらず、 $4l$ から $4l+3$ までは信頼区間が広くなり、 $4(l+1)$ のとき信頼区間が減少する。
- 5) 環境基準の適合判定にいうところの75%水質値は、旭川の場合、一年間 365個のデータにおける75~80%の水質に相当し、これを越える場合は、平水量~豊水量以上の場合であり、低水量以下の場合は、わずかであった。
- 6) 18~52個のデータから、75%水質値を推定すると、信頼区間の下限値は 365個のデータの80%弱から70%程度に、上限値は90%程度に相当することが判明した。

本研究では、非超過確率水質の推定値の特徴について、検討を加えた。今後、これらの特徴を踏まえて、非超過確率水質把握手法、ならびに、把握された水質に基づく判定手法の改善に関して検討を加える予定である。

〈参考文献〉

- 1) I.ガットマン; S.S.ウィルクス著, 石井恵一; 堀素夫 訳, 工学系のための統計概論, 培風館, 1984.
- 2) たとえば, 新田正, 河川の水質の変動に関する研究, 東京工業大学学報, No.20, 1961.
- 3) 河原長美, 河川の水質ならびに汚濁負荷量の代表値推定に関する2, 3の検討, 土木学会第12回環境問題シンポジウム講演論文集, 1984.
- 4) 同上