

〔資料〕

pHの統計解析方法の検討

石川県保健環境センター 環境科学部 金 曾 将 弘

〔和文要旨〕

水素イオン濃度指数 (pH) は水の性状の指標として用いられており、その定義は水素イオン活量 (a_{H}) を用いるものとガラス電極計によるものがある。ガラス電極計のpHから a_{H} を計算する場合を考えると、pHの平均値はガラス電極pH計に発生する起電力の平均値に対応する値であり、 a_{H} の平均値は活量係数が1とみなせる緩衝作用のない希薄溶液において溶液を等量混合した時の水素イオン活量の値である。pHと a_{H} の変換にあたっては、pHの平均をとってから a_{H} に変換した値と a_{H} に変換してから平均をとった値は異なることや、端数処理後の値から計算される統計値がサンプル母集団の統計値には必ずしも一致しないことに留意する必要がある。

キーワード：pH, 水素イオン活量, 統計解析, 分布関数, 平均

1 はじめに

水素イオン濃度指数 (pH) は水の基本的な性状を示すものとして河川の水質調査等に広く利用されている。しかしながら、年平均値の取り方などその統計解析の方法については統一的な方法が必ずしも見当たらないのが現状である。国の酸性雨調査の解析¹⁾では「pHは H^+ 濃度の降水量重みつき平均値のpH換算値を(中略)月間及び年間についてそれぞれ算出した」としている。一方、水質汚濁防止法に基づく公共用水域の水質常時監視では、pHの平均値について重みづけ等の特別の規定はなく、他の項目と同様「平均値の計算に当たっては、有効数字を2桁までとし、その下の桁を四捨五入する。」²⁾が適用されているものと思われる。

通常、pHの測定方法は標準液による校正を行いガラス電極法により行われている。この方法により観測されたpHに対する数値解析の方法については、公共用水域の水質常時監視のようにpHのまま解析を行う方法と、酸性雨調査のように常用対数をとって符号を逆転した a_{H} に変換して行う方法の2つに大別される。この両者の方法については、例えば玉置ら³⁾が雨水における実データをもとに比較検討しているが、両者の統計値の意味等

について詳細に検討した資料はあまり確認できない。

今回、このpH及び a_{H} に対する統計解析について簡単ではあるものの考察を行ったので、報告する。

2 考 察

(1) 測定方法・定義

pHの測定については様々な方法が用いられており、IUPACではpHを $\text{pH} = -\log_{10} a_{\text{H}}$ (a_{H} は水素イオンの活量)と定義している。測定方法としてはHarned cellと呼ばれる電池にBates-Guggenheimの慣用式を適用して、一次標準液を不確かさのある値として決定する⁴⁾。この標準液を利用して、すべてのpHの観測値を不確かさを持った値として順次評価していく。

一方JIS K0211(分析化学用語(基礎部門))(2013)では、pHは「水素イオン活量の逆数の常用対数」とされているものの、注釈として「概念上の定義で実測できない値である」としており、実際の測定ではJIS Z8802を参照するものとしている。JIS Z8802:2011(pH測定方法)では、pHは「…pH標準液のpH値を基準とし、ガラス電極pH計によって測定される起電力から求められる値」とされている。これは測定の結果によって得られる値をpHと名付けており、その実態が水素イオン活

A Theoretical Discussion on the Statistical Analysis of pH. by KANESO Masahiro (Environmental Science Department, Ishikawa Prefectural Institute of Public Health and Environmental Science)

Key words : pH, Activity of Hydrogen Ion, Statistical Analysis, Distribution Function, Average

量と対応しているかを問題としていない。

JIS Z8802の定義による pH 値は濃度 0.1mol/L 以下, pH3 ~ 10の緩衝液については水素イオン活量によって定義される pH 値と ± 0.02の範囲で一致すると考えられている⁵⁾。しかしながら, 海水や工場排水などこの前提を満たさない液は多く, 実際にガラス電極法によって計測された pH が水素イオン活量とどう結びついているかは不明な点も多いと考えられる。

以下ではガラス電極法により観測された pH を, 水素イオン活量と理論的に結び付くものではない観測値であることを明示的にして「pH^{obs}」と記載し, 水素イオン活量「a_H^{obs}」を, pH^{obs}を pH = -log₁₀ a_Hの関係式に代入して得られる pH^{obs} = -log₁₀(a_H^{obs})の関係式により定義する。

(2) 平均値

まず, pH^{obs}についてみる。簡単のため, 以下では温度及び内部液の濃度などの変動は無視できるものとする。pH^{obs}はガラス電極 pH 計によって測定される起電力から計算される値であり, 物理的には対象溶液, 参照電極, 内部液, ガラス電極など特定の構成の機器により発生する起電力に対応する値である。

これをもとに考えると, pH^{obs}の平均値はこの起電力が平均的にどの程度の値であるか, ということを意味している。注意すべき点として, pH は溶液の混合と直接対応しないことが挙げられる。すなわち, pH8.00の液と pH7.00の液を等量混合しても, pH7.50にはならない。pH^{obs}の平均値は溶液の混合とは対応しないため, 最頻値・中央値等によってサンプルの母集団を評価する方が良い場合もあると思われる。

次に, a_H^{obs}についてみると, 定義は pH^{obs} = -log₁₀(a_H^{obs})であり, その物理的意味については上述のとおり一定の条件を満たす場合に概ね理論的な水素イオン活量と一致する量である。

このことから a_H^{obs}の平均値について考えると, 活量係数が 1 とみなせるような希薄溶液であって緩衝作用のない場合には, 各溶液を等量混合した溶液の a_Hに対応する。

(3) 観測値の分布

観測誤差について考えると, pH^{obs}は直接的な観測値である起電力 V に対して, 温度等で決まる定数 A, B を用いて pH^{obs} = A × V + B と線形の関係にある。仮に観測値 V の偶然誤差が正規分布に従うとき, この線形性から pH^{obs}についても誤差は正規分布に従う。pH^{obs}の分布が

$$f(pH^{obs}) = N(\mu_{pH}, \sigma_{pH}^2) = \frac{1}{\sqrt{2\pi\sigma_{pH}^2}} \exp\left[-\frac{(pH^{obs} - \mu_{pH})^2}{2\sigma_{pH}^2}\right] \dots (1)$$

であらわされる平均 μ_{pH}, 標準偏差 σ_{pH} を持つ正規分布 f(pH^{obs}) となると, a_H^{obs} の分布関数 g(a_H^{obs}) は

$$g(a_H^{obs}) = \frac{1}{\sqrt{2\pi\sigma_{pH}^2 \times \ln 10 \times a_H^{obs}}} \exp\left[-\frac{(\log_{10} a_H^{obs} + \mu_{pH})^2}{2\sigma_{pH}^2}\right] \dots (2)$$

となり, 正規分布しない。つまり, a_H^{obs} に対して正規分布を仮定する誤差解析を行うことは控えるべきである。

(4) 観測値の取り扱い

次に, 観測値の取り扱いについて考える。JIS K0102 (工場排水試験方法) (2019) では pH は 3 回測定 of 平均値として算出するとしており, 自明ではあるが, 平均値をとってから変換した値と, a_H^{obs} に変換してから平均値をとった値は異なる。

同一サンプルに対する複数回測定のうち, 個々の値の記録がない場合の適切な変換を考える。ある同一の検体に対して複数回 pH^{obs} の測定を行って得られる分布関数 f が, 上と同様平均 μ_{pH}, 標準偏差 σ_{pH} の正規分布である式(1)に従うと仮定する。この時, a_H^{obs} = 10^{-pH^{obs}} の期待値 E[a_H^{obs}] について考えると, 定義から

$$E[a_H^{obs}] = \int_{-\infty}^{\infty} 10^{-pH^{obs}} f(pH^{obs}) dpH^{obs} \\ = \int_{-\infty}^{\infty} \frac{10^{-x}}{\sqrt{2\pi\sigma_{pH}^2}} \exp\left[-\frac{(x - \mu_{pH})^2}{2\sigma_{pH}^2}\right] dx \dots (3)$$

となる。この積分結果から, E[a_H^{obs}] ≠ 10^{-E[pH^{obs}]}}, 書き換えると E[pH^{obs}] ≠ -log₁₀(E[a_H^{obs}]) である。このことから上述のとおり, 同一溶液で pH^{obs} を複数回観測した平均値を関係式 pH^{obs} = -log₁₀(a_H^{obs}) で変換した値は, a_H^{obs} の平均値と一致しない。つまり同一溶液で複数回観測した結果を平均した pH^{obs} をそのまま変換するのは誤りであり, 各観測値を先に a_H^{obs} に変換した後で平均した値か, 又は, 観測した pH^{obs} の平均と不偏分散から推定される μ_{pH} 及び σ_{pH} を用いて式(3)の積分を数値的に解いて得られる E[a_H^{obs}] を統計解析には用いるべきである。しかし, 通常の公表データでは標準偏差 σ_{pH} は記録されておらず, 実際にはこの操作は難しいと思われる。

ところで, σ_{pH} → 0 の極限では f(pH^{obs}) = N(μ_{pH}, σ_{pH}) → δ(pH^{obs} - μ_{pH}) が成立する (ここで, N(μ, σ) は平均 μ, 標準偏差 σ の正規分布, δ は Dirac のデルタ関数) ので, 式(3)は

$$E[a_H^{obs}] = \int_{-\infty}^{\infty} 10^{-pH^{obs}} f(pH^{obs}) dpH^{obs} \\ \rightarrow \int_{-\infty}^{\infty} 10^{-pH^{obs}} \delta(pH^{obs} - \mu_{pH}) dpH^{obs} \quad (\sigma_{pH} \rightarrow 0) \\ = 10^{-\mu_{pH}} \\ = 10^{-E[pH^{obs}]} \dots (4)$$

のように変形できて, E[a_H^{obs}] = 10^{-E[pH^{obs}]}} が成立する。すなわち, pH^{obs} の標準偏差 σ_{pH} が十分小さいとみなせる場合は, E[a_H^{obs}] = 10^{-E[pH^{obs}]}} による単純な変換が成立

する。実際の溶液について考えると、JIS K0102（工場排水試験方法）（2019）ではpHの測定値が±0.1（緩衝性の低い試料では±0.2）で一致した測定値を平均することとされており、溶液によってはこの程度の標準偏差 σ_{pH} を持つと考えられるため、この近似が本当に成立するかは、各測定対象に応じて検討する必要がある。

(5) 離散化処理（端数処理）

次に、測定結果の離散化処理に伴う問題について述べる。通常記録・管理される値は pH^{obs} として0.1又は0.01刻みの一定間隔で端数処理して離散化された形になっている。端数処理については、複数の考え方があるが、例えば水質汚濁防止法に基づく公共用水域の水質常時監視においては、小数第2位を四捨五入するものとされている⁴⁾。

連続な確率変数 X の平均値 $E(X)$ と、 X に対し（四捨五入や切り捨てなどの）離散化変換 R を作用させた $R(X)$ の平均値 $E(R(X))$ は一般的には一致しない。つまり、複数のサンプルからなる集団に対して、四捨五入や切り捨てなどの端数処理後の値から計算される平均値は、サンプル母集団の平均には必ずしも一致しない。例えば切り捨て処理を行った値の平均値はもとの値の平均値に比べて低くなることは明らかである。分散等についても同様であり、サンプルの数を増やしても離散化に伴う誤差の問題は解決しないと思われる。

実際の例として pH^{obs} の側で端数処理を行う場合について考える。 a_H^{obs} が $1.1 \times 10^{-8} \sim 1.0 \times 10^{-7}$ の間で、 0.1×10^{-8} 刻みで一様分布する90個のサンプル集団（以下、この値を真の値という。）を測定して、上述の公共用水域の水質常時監視の方法に基づき数値処理を行うことを考える。例えば $a_H^{obs} = 5.00 \times 10^{-8}$ のサンプルは、 $-\log_{10}(a_H^{obs}) = 7.301\dots$ であるから、小数第2位を四捨五入して pH^{obs} は7.3として記録される。ここで、 $10^{-7.3} = 5.01\dots \times 10^{-8}$ であるから、 pH^{obs} として端数処理を行って記録された値を a_H^{obs} に戻した値は、真の値から 0.01×10^{-8} だけずれが生じる。（以下、四捨五入後の pH^{obs} に対して $10^{-pH^{obs}}$ により求めた値を記録値という。）この変換と端数処理に伴う数値のずれの関係を図1に示す。図のとおり、記録値の間隔は pH^{obs} とは違って等間隔でなく、また a_H^{obs} の値の区間は点によ

て大きく異なるため、真の値から作る分布と記録値から作る分布は違いが生じる。図2にこの90サンプルについて a_H^{obs} で分けけた場合のヒストグラムを示す。刻み幅は 0.1×10^{-8} とした。真の値から作る分布は一様分布であるが、記録値から作る分布では形状が歪んでいる。

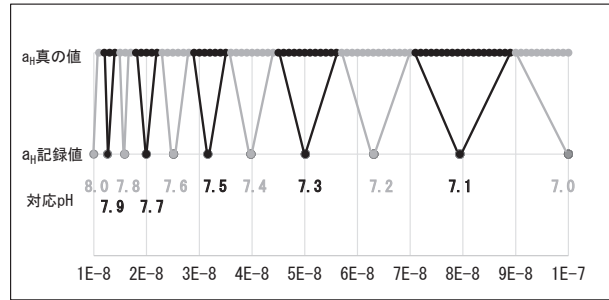


図1 真の値と記録値の対応関係

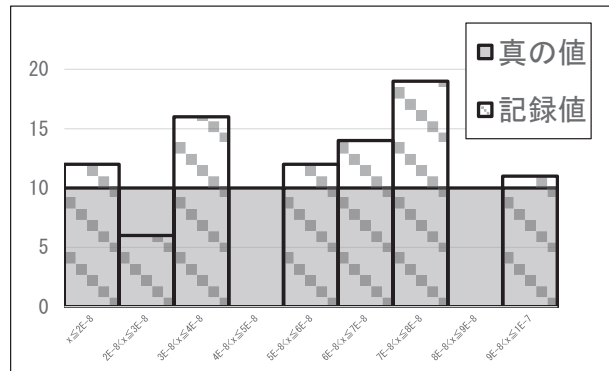


図2 真の値と記録値の分布

表1に真の値と記録値とで各種の統計値がどの程度変化するかを示す。特に分散では5.7%，中央値では9.7%の乖離が真の値と記録値との間で生じており、 pH^{obs} として記録されている結果をそのまま戻した場合には一定の影響がある。参考までに、このサンプル集団に対する pH^{obs} の統計値を四捨五入前後で比較したものを合わせて表1に示す。このサンプル集団の場合、端数処理に伴う誤差は a_H^{obs} に比べて pH^{obs} のほうが小さいが、この誤差は母集団の分布関数の形状に依存すると考えられ、 a_H^{obs} の誤差が pH^{obs} に比べて常に大きくなるかは判然としない。しかしながら、記録値の間隔が等間隔ではない a_H^{obs} に比べて、等間隔となる pH^{obs} の方が分布の歪みは

表1 真の値と記録値における統計値の比較及び誤差

	a_H			pH		
	真の値	記録値	記録値/真の値	四捨五入前	四捨五入後	四捨五入後/四捨五入前
相加平均	5.550E-8	5.577E-8	100.5%	7.318E+0	7.318E+0	100.0%
相乗平均	4.812E-8	4.811E-8	100.0%	7.313E+0	7.313E+0	100.0%
標本分散	6.749E-16	7.132E-16	105.7%	6.322E-2	6.546E-2	103.6%
中央値	5.550E-8	5.012E-8	90.3%	7.256E+0	7.300E+0	100.6%
最頻値	-	7.943E-8	-	-	7.100E+0	-

小さく、誤差は小さいことが多いのではないかと推察する。

この誤差に対するアプローチについてはいくつかの方法が考えられる。1つは pH^{obs} の端数処理に対応する範囲において $a_{\text{H}}^{\text{obs}}$ の分布を仮定し、その平均値を使用する方法である。 $a_{\text{H}}^{\text{obs}}$ の分布を一様分布と仮定し、 pH^{obs} に対して四捨五入の端数処理を行う場合における対応表を表 2 に示す。例えば $\text{pH}7.5$ と記録された結果があった場合、実際の詳細な観測値は $\text{pH}7.45 \sim 7.55$ の範囲にあったと考えられる。これに対応する $a_{\text{H}}^{\text{obs}}$ の範囲は $2.82 \times 10^{-8} \sim 3.55 \times 10^{-8}$ の範囲であるので、その平均値である 3.18×10^{-8} を使用して統計解析を行う。しかし、この方法では分布の形状は歪んだままであるため、大きな改善は期待しにくいと考えられる。別のアプローチとしては、 pH^{obs} の端数処理に対応する範囲において a_{H} の分布を仮定し、乱数を使用してその範囲の値を分布に従う形で生成する方法が考えられる。この方法については仮定した分布の形状を再現して種々の統計値を計算できるが、乱数の初期値依存性や計算結果の取束性について別途議論が必要であろう。

表 2 記録された pH から a_{H} への変換例

記録された pH	対応する pH 範囲	対応する a_{H} (mol/L) 範囲	区間平均値 (mol/L)
8.0	8.05 ~ 7.95	8.912509.E-9 ~ 1.122018.E-8	1.006635.E-8
7.9	7.95 ~ 7.85	1.122018.E-8 ~ 1.412538.E-8	1.267278.E-8
7.8	7.85 ~ 7.75	1.412538.E-8 ~ 1.778279.E-8	1.595408.E-8
7.7	7.75 ~ 7.65	1.778279.E-8 ~ 2.238721.E-8	2.008500.E-8
7.6	7.65 ~ 7.55	2.238721.E-8 ~ 2.818383.E-8	2.528552.E-8
7.5	7.55 ~ 7.45	2.818383.E-8 ~ 3.548134.E-8	3.183258.E-8
7.4	7.45 ~ 7.35	3.548134.E-8 ~ 4.466836.E-8	4.007485.E-8
7.3	7.35 ~ 7.25	4.466836.E-8 ~ 5.623413.E-8	5.045125.E-8
7.2	7.25 ~ 7.15	5.623413.E-8 ~ 7.079458.E-8	6.351436.E-8
7.1	7.15 ~ 7.05	7.079458.E-8 ~ 8.912509.E-8	7.995984.E-8
7.0	7.05 ~ 6.95	8.912509.E-8 ~ 1.122018.E-7	1.006635.E-7

(6) 実際の解析への適用

最後に、実際の解析にあたって pH^{obs} と $a_{\text{H}}^{\text{obs}}$ のどちらを使用すべきかを検討する。これまで述べたとおり、 $a_{\text{H}}^{\text{obs}}$ は希薄溶液では混合という観点を持つものの、 pH^{obs}

に比べて(3)式による変換が必要となる。 pH^{obs} のみが記録されている場合について考えると、溶液の混合が意味を持つような場合(例: 1回の降雨に対する雨水の解析)には $a_{\text{H}}^{\text{obs}}$ を使用することが良いと考える。液の性状の変化に重点が置かれる場合(例: ある地域における河川水の経年変化)には、標準偏差を要する(3)式の変換が不要な、 pH^{obs} で解析を行う方が良いと考える。

最後に、本文の概要を表 3 にまとめる。

3 まとめ

- (1) ガラス電極 pH 計によって観測された pH^{obs} を、 $\text{pH}^{\text{obs}} = -\log_{10}(a_{\text{H}}^{\text{obs}})$ によって変換した値 $a_{\text{H}}^{\text{obs}}$ は、必ずしも水素イオン活量とみなせるわけではない。
- (2) pH^{obs} の平均値はガラス電極 pH 計に発生する起電力の平均値に対応する値であり、 $a_{\text{H}}^{\text{obs}}$ の平均値は緩衝作用がなく、かつ活量係数が 1 とみなせる希薄溶液において溶液を等量混合した時の水素イオン濃度の値である。
- (3) 同一サンプルに対して複数回測定を行ったとき、 pH^{obs} の平均をとってから変換した値と、 $a_{\text{H}}^{\text{obs}}$ に変換してから平均をとった値は異なる。 pH^{obs} の平均と分散から $a_{\text{H}}^{\text{obs}}$ を求める場合は、式(3)を数値積分する。
- (4) 複数のサンプルからなる集団に対して、四捨五入や切り捨てなどの端数処理後の値から計算される統計値は、サンプル母集団の統計値には必ずしも一致しない。
- (5) 溶液の混合が意味を持つような場合には $a_{\text{H}}^{\text{obs}}$ を、液の性状の変化に重点を置いて解析を行う場合は pH^{obs} を解析に使用するのが良いと考える。

文 献

- 1) 環境省: 令和 3 年度国内モニタリングデータの取りまとめ, <https://www.env.go.jp/content/000109961.pdf>, 2023年10月17日閲覧
- 2) 環境省水・大気環境局長通知環水企第92号: 環境基本法に基づく環境基準の水域類型の指定及び水質汚濁防止法に基づく常時監視等の処理基準について、平

表 3 pH^{obs} と $a_{\text{H}}^{\text{obs}}$ の比較

	pH^{obs}	$a_{\text{H}}^{\text{obs}}$
定義	ガラス電極 pH 計によって測定される起電力から計算される値	$\text{pH}^{\text{obs}} = -\log_{10}(a_{\text{H}}^{\text{obs}})$
物理的意味	ガラス電極 pH 計によって発生する起電力	一定条件を満たすとき、水素イオン活量
相加平均値の意味	起電力の平均に対応する値	一定条件を満たすとき、溶液を等量混合した時の水素イオン濃度
観測値の分布 (同一サンプル)	起電力の偶然誤差が正規分布に従うとき、式(1)の形 (正規分布である)	起電力の偶然誤差が正規分布に従うとき、式(2)の形 (正規分布でない)
平均値の計算方法 (同一サンプル)	複数回測定を行ったとき、その pH の平均値	複数回測定を行ったとき、その pH を変換したものの平均値又は式(3)の数値積分により計算される値
記録値の間隔	通常一定間隔	pH^{obs} が一定間隔の時、一定間隔にならない
端数処理による統計値への影響	あり	あり
解析対象の適用範囲(例)	液の性状の変化に重点を置く場合	溶液の混合が意味を持つような場合

成13年5月31日（令和3年10月7日改正）

- 3) 玉置元則, 平木隆年: わが国各地における雨水pHの年平均値 (1), 環境技術, **15**, 188-192 (1986)
- 4) J.G.Frey et al.: (独) 産業技術総合研究所計量標準総合センター訳 IUPAC 物理化学で用いられる量・単位・記号 第3版, 90-91, 講談社 (2009) Web版

(https://unit.aist.go.jp/nmij/public/report/others/pdf/iupac_green_book_jp.pdf/) 2023年8月8日閲覧

- 5) 日本規格協会 編: JIS使い方シリーズ 詳解 工場排水試験方法 [JIS K 0102:2019] 改訂6版, 68, 日本規格協会 (2019)