Indonesian Journal of Chemical Analysis
Homepage: https://journal.uii.ac.id/IJCA

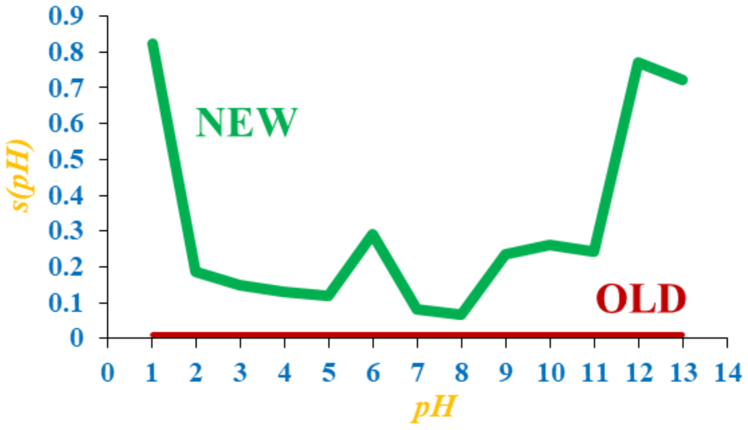
Uncertainty of Routine *pH* Measurements Evaluated by the Principle of Pooled Calibrations

Jens E.T. Andersen*, Mercy Menong

Botswana International University of Science and Technology, School of Pure and Applied Sciences, Department of Chemical and Forensic Sciences, Plot 10071, Boseja Ward, Private Bag 016, Palapye, Botswana
* corresponding author: andersenj@biust.ac.bw

DOI: 10.20885/ijca.vol8.iss1.art2

GRAPHICAL ABSTRACT



ARTICLE INFO

Received : 01 October 2024
Revised : 29 December 2024
Published : 30 March 2025
Keywords : Buffer, pH calculations, uncertainty of measurement, preparation by mass, principle of pooled calibrations.

ABSTRACT

In the digital age, taking logarithms of data is no longer necessary when assessing pH measurement uncertainties. The focus is now on using raw, unaltered data. Routine pH measurements often differ from those in reference labs, so quantifying these differences is necessary. The uncertainties of pH values at 25°C were evaluated using mass-by-mass calculations, along with comparisons between two pH meters. Four buffer systems and several beverage samples were tested. Expanded uncertainties (U) of up to $U = 0.36$ were observed for pH values between 2 and 11, significantly higher than those reported by manufacturers. While measured pH values were slightly higher than calculated ones, overlapping confidence intervals allowed the data to be combined. Due to significant uncertainties, the reliable pH range was limited to $1 < pH < 11$, potentially narrower ($1 < pH < 6.5$). Routine pH measurement uncertainties did not match those based on the PoPC, with notably high CV values for proton activities at pH values below the buffers' pKa.

1. INTRODUCTION

Advanced pH meters and analytical chemistry methods can detect measurements at very low concentrations, such as ng levels g.kg^{-1} or pg.kg^{-1}. Therefore, it may be a little surprising to some users that pH measurements of electrodes that are considered relatively simple can detect proton concentrations of 10^{-14} g.kg^{-1}, or 0.01 pg.kg^{-1}, with excellent levels of precision and trueness [1]. However, the determination of pH values is not straightforward, as was illustrated in the recommendations by the International Union of Pure and Applied Chemistry (IUPAC) [2–4], where interferences such as Na^+, K^+, and Cl^- are expected to influence both the *pH* values and the corresponding uncertainties [5–9]. The concept of proton activity was introduced to explain that the measured *pH* value should be adjusted by the activity coefficient, which is typically close to unity at low concentrations, to calculate the proton concentration used to prepare the solutions [10]. A good uncertainty to the Harned cell was reported as 0.002 *pH* units [11], but an overall IUPAC traceable uncertainty of 0.012 was assigned to the *pH* value determined by the Harned cell [12].

The measurements of IUPAC were performed using the Harned cell [11–13] under very controlled laboratory conditions where IUPAC recommends an uncertainty of 0.01 in diluted solutions (< 0.1 mol.kg^{-1}) within the temperature range of 0-50 °C [2]. The accuracy of the *pH* values determined by the Harned cell depends on many parameters, such as the structural design of the reference electrodes, where cylindrical structures performed better with response time than planar or spherical designs [14].

In a key comparison of a carbonate buffer with a nominal *pH* value of 10.0 at 25°C, the participating laboratories reported that the measured average value was significantly higher than the nominal value [4]. Interestingly, this result was reported after excluding outliers and calculating the uncertainty using a weighted method. It is unclear how the authors can expect others to reproduce the result to three decimal points without excluding outliers and potentially using different or no weighting schemes in their analyses [4]. It appears there is a significant misunderstanding of the concepts of trueness and reproducibility [1]. In the vocabulary of metrology 3rd edition (VIM3) [1] are these concepts explained, and they refer to the compatibility of repeated measurements of identical samples in different locations by different operators, which often shows to be challenging to achieve without rejection of outliers [4].

The uncertainty of the ionic strength is negligible [13, 15]. By considering the influence of ionic strength less than 0.1 on the *pH* measurements, Bates and Guggenheim [16] suggested a standard deviation of 0.02 that would correspond to an expanded uncertainty of 0.04 (k = 2) [17] within the interval $2 < pH < 12$, thus also reducing the number of decimal points to only two. Leito et al. [18] also suggested two decimal points, but that was a number related to routine pH measurements. A total of 67 different sources of uncertainty were identified, and the benzoic acid expanded uncertainty (U) of the negative logarithm to the acid-dissociation constant, pKa = 4.219 value was determined as 0.034 when the pH meter was calibrated using low-pH buffer (pH = 1.679) [19]. A similar level of uncertainty was established by Villasana et al. [20] for measurements of pH values for each concentration 0.05 M, 0.1 M, and 0.7 M of $MgSO_4$ where the *SD*s up to 0.02 were tentatively assigned to properties of the liquid junction potential. However, the significant random fluctuations in *pH* values as a function of [$MgSO_4$], without considering temperature, were not explained. The overall standard deviation across all results was estimated at 0.14, with an expanded uncertainty of U = 0.28 (k = 2) [20]. If this latter value could be attributed to uncertainties, the suggested number of decimal places for *pH* measurements, based on the abovementioned examples, could range from one to three.

For routine laboratory conditions of pH measurements [18] with the application of secondary standards to the measurements [21], it was realized that single calibrations followed by application to real samples could lead to large day-to-day variations, thus compromising the level of trueness of the method. Consequently, a more thorough investigation was initiated to determine the appropriate measurement uncertainty for *pH* values in calibration and sample solutions. These solutions were prepared by mass to minimize or eliminate concentration uncertainties, thereby allowing for the identification of other sources of uncertainty. Also, the influence of temperature on the measurements of pH values was minimized by equilibrating all solutions at 25°C. The principle of pooled calibrations (PoPC) [22] was applied to estimate the expected minimum level of uncertainty that could be assigned to routine pH measurements. As opposed to conventional approaches [10], the PoPC allows for assessing an overall level of universal uncertainty of all types of routine pH measurements. The PoPC [22] uses many independent series of operational calibrations to obtain uncertainty that addresses trueness or accuracy rather than focusing on precision.

The results indicated that uncertainties much larger than those obtained under very controlled laboratory conditions [12], including those of the manufacturers' specifications [23, 24], should be assigned to the *pH* values that were obtained at room temperature (25 ºC) under routine laboratory conditions. It was also indicated that corrections of concentrations by activity coefficients might be unnecessary as the corresponding adjustments would fall short of the overall level of measurement uncertainty [22].

2. EXPERIMENTAL METHODS

2.1. Chemicals and samples

The solutions were prepared as buffer solutions to minimize the influence of interferences on the pH measurements. All solutions were prepared in deionized water. Buffers with pH = 0.92 and 0.93 were prepared with sodium hydrogen sulfate ($NaHSO_4$) assay 98%, batch 2575 + disodium sulfate (Na_2SO_4). Buffers with $2.8 < pH < 4.7$ were prepared with acetic acid (CH_3COOH), assay 99.5%, 090215AA + sodium acetate trihydrate ($CH_3COONa.3H_2O$), assay 99- 101.0 %, buffers with $6.5 < pH < 11$ were prepared with ammonium chloride (NH_4Cl), assay 99%, Uni lab SAAR1122700EM + ammonia solution ($NH_3(aq)$), assay min. 25%, 070515AM, and buffers with $pH = 13.5$ and $pH = 13.6$ were prepared with sodium hydrogen carbonate ($NaHCO_3$), assay 99%, supplied by Rochelle chemicals + sodium hydroxide (NaOH), assay 97%, batch 05121450 that were purchased from Merck (Pty) Ltd. Standard buffer solutions (Thermo Scientific $pH = 4.01$, 7.00 and 9.00) were used for the three-point calibrations. A solution of 3 M KCl was used to store the electrode between each series of measurements. The samples of UHT milk, fruit juice, Red Bull^{TM}, and diet Coca-Cola^{TM} were sourced from local supermarkets.

2.2. Instrumentation

The pH measurements were performed with two different apparatuses: (1) the Thermo Scientific Orion star A111 *pH* meter (Thermo) equipped with glass electrode Thermo Scientific Orion N9107BN Low Maintenance Triode [23] and (2) the Basic 20 *pH* meter (Crison) equipped with a 5010T electrode [25]. The Thermo *pH* meter was capable of measurements in the interval $0 < pH < 14$ at a readability of 0.01 and an accuracy of 0.002. The Basic 20 *pH* meter allowed for a broader interval of measurements $-2 < pH < 14$ with readability and accuracy equal to 0.01. Both electrodes were able to measure temperatures within the interval $-20°C < t < 150°C$ at an uncertainty of 0.2°C for the Basic 20 *pH* meter, whereas the specified temperature interval was $0°C < pH < 90°C$ with uncertainty 0.1°C for the Thermo *pH* meter. Hence, the chosen *pH*-interval $1 < pH < 14$ and 25°C for the present investigations were well covered by both electrodes.

2.3. Analytical methodology

Four sets of acid-base pairs of buffer solutions were prepared as follows: acetic acid (CH_3CO_2H, $Ka(CH_3CO_2H) = 1.8x10^{-5}$) and sodium acetate ($CH_3CO_2Na$); ammonium chloride ($NH_4Cl$, $Ka(NH_4^+) = 5.6x10^{-10}$) and 25% ammonia solution (NH_3); sodium hydrogen sulfate ($NaHSO_4$, $Ka(HSO_4^-) = 1.0x10^{-2}$) and disodium sulfate ($Na_2SO_4$); Sodium hydrogen carbonate ($NaHCO_3$, $Ka(HCO_3^-) = 4.7x^{10-11}$) and sodium hydroxide (NaOH) [26]. The solutions were prepared using the Henderson-Hasselbalch equation modified for mass-by-mass calculations [27]. 250 ml of each buffer was prepared by dissolving the weighed compounds in volumetric flasks and filling them to the mark with deionized water. The total mass of each solution was then measured using a top-pan balance (Mettler) with a readability of 0.001 g. The solutions were placed in water baths at 25°C for one day before taking measurements. Electrodes were inserted into the solutions, and measurements were recorded once the temperature stabilized at 25°C and the *pH* value reached a constant level. Calibration measurements and measurements of the buffers and samples were performed according to the manufacturer's instructions.

3. RESULTS AND DISCUSSIONS

In typical laboratory *pH* measurements, the impact of temperature is often overlooked. However, in this study, all solutions were equilibrated at 25°C to minimize the influence of temperature on the uncertainty. Acid and base concentrations were measured by mass, reducing preparation uncertainty. The *pH* values were then calculated using the following specific mass-based formulas [27]:

$$pH = -log\left(\frac{\sqrt{u^2+4\cdot K_a\cdot n(bH)\cdot m(tot)}-u}{2\cdot m(tot)}\right) = pK_a + log\left(\frac{n(wb)+z(bH)}{n(bH)-z(bH)}\right) \tag{1a}$$

$$z(bH) = \frac{1}{2}\cdot\left(\sqrt{u^2 + 4\cdot K_a\cdot\frac{m(bH)}{M(bH)}\cdot m(tot)}\; - u\right) \tag{1b}$$

$$u = \frac{m(wb)}{M(wb)} + K_a\cdot m(tot) \tag{1c}$$

$$m(tot) = m(s) + m(bH) + m(wb) \tag{1d}$$

Where *n(bH)* and *n(bH)* are the total moles of acid and base, respectively, *m(bH)* and *m(wb)* are the corresponding masses, and *m(s)* is the mass of water that was added to prepare the total mass *m(tot)* of the solutions. The parameters (Eqs, 1a-1c), *M(bH)*, and *M(wb)* represent the molar mass of weak acid and the molar mass of weak base, respectively. As pointed out previously, the number of moles of acid displaced (*z(bH)*) (Eqs. 1a and 1b), according to the equilibrium, cannot be zero, which renders the Henderson-Hasselbalch equation unsuitable for many combinations of the parameters in Eqs. 1a-1d [27].

Although the considerations around the uncertainty of *pH* values are relevant for routine *pH* measurements [18], it should be noted that the *pH* value itself is not the measurement. The *pH* value is not measured, but it is calculated as a voltage (*V*) that originates from the potential difference (*DE*) that is created between the inner and outer part of a glass electrode by a difference in proton activity across the glass membrane, according to the Nernst equation [28], as follows:

$$V = \Delta E = \ln(10)\cdot\frac{R\cdot T}{F}\cdot pH = -\ln(10)\cdot\frac{R\cdot T}{F}\cdot log\big(a(H^+)\big) \tag{2}$$

R is the gas constant, *T* is temperature, and *F* is Faraday's constant. Regarding the uncertainty budget (Equation 3) relative to Equation 2, temperature and *pH* value are the primary factors, while the contributions from *R* and *F* are negligible. The voltage (*V*) is also not the measurement, as it is produced by the *pH* meter after the transducer amplifies the signal. Therefore, the proton activity, which creates the potential difference (Equation 2), is the actual measurement to consider when calculating and evaluating uncertainties. However, it may still be useful to estimate the combined relative uncertainty (*uc(V)*) corresponding to Equation 2 using the law of propagation of uncertainties [17]:

$$\left(\frac{u_c(V)}{V}\right)^2 = \left(\frac{s_R}{R}\right)^2 + \left(\frac{s_T}{T}\right)^2 + \left(\frac{s_F}{F}\right)^2 + \left(\frac{s\left(a(H^+)\right)}{a(H^+)\cdot ln\left(a(H^+)\right)}\right)^2 \tag{3}$$

The *SD*s of gas and Faraday's constant are given by $sR = 0.0000048$ J/K/mol and $sF = 0.00059$ C/mol [29], respectively, and the temperature *SD* is estimated at $sT = 0.5$ K. The temperature term and the term with proton activities dominate the overview of relative contributions to the uncertainty budget at all *pH* values of the different buffer systems (Eq. 3, Fig. 1). At $pH < 4$ and $6 < pH < 8$, the proton activity constitutes the significant contribution to the uncertainty of the voltage whereas the contribution from temperature becomes more critical when pH > 10. At $pH = 4.7$, which corresponds to the *Ka*-value of acetic acid, and at $pH = 9.0 \sim pKa$ of the ammonium ion, the temperature and proton activity contributed almost equally to the uncertainty (Fig. 1).

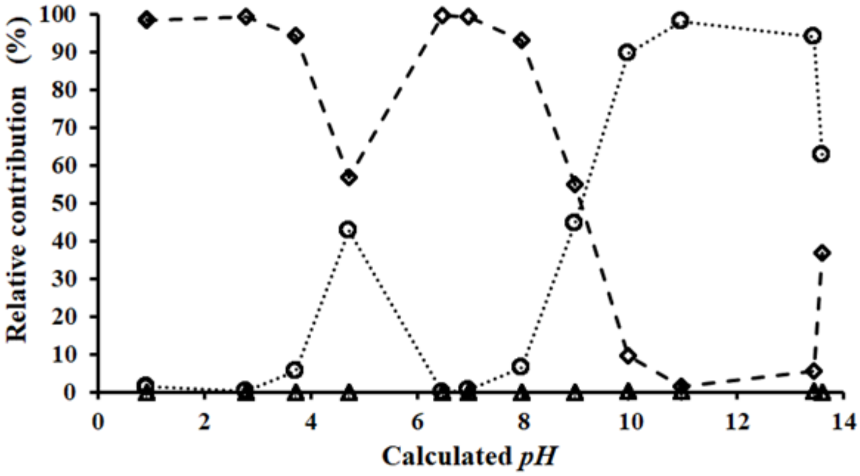


Figure 1. Relative contributions (%) to the uncertainty budget for all the studied *pH*-values, according to the parameters comprising (+) Ideal gas constant (*R*), (○) Temperature (*T*), (△) Faraday's constant (*F*), and the (◇) the proton activity ($a(H^+)$) of Eq.3. The broken and dotted lines are guides to the eye.

The relative uncertainties (*RSD*s) of the proton activity may be obtained by using the K_a-value where each parameter represents contribution to the law-of-propagation of uncertainties [17], as follows:

$$\left(\frac{s\left(\mathrm{a}(H^+)\right)}{\mathrm{a}(H^+)}\right)^2 = \left(\frac{s\left(m(bH)\right)}{\mathrm{m}(bH)}\right)^2 + \left(\frac{s\left(m(wb)\right)}{m(wb)}\right)^2 + \left(\frac{s\left(m(tot)\right)}{m(tot)}\right)^2 + \left(\frac{s(K_a)}{K_a}\right)^2 + + \left(\frac{s(MbH)}{M(bH)}\right)^2 + \left(\frac{s\left(M(wb)\right)}{M(wb)}\right)^2 \tag{4}$$

When all the solutions were prepared with concentrations by mass, the terms with relative masses were very small compared to the *Ka* term. It can be shown that the relative standard deviation (*RSD*) of the *Ka*-value dominates at pH-values lower than 11, above which the *RSD* of *m(bH)* starts to become more important (Fig. 1, Eq. 4). Also, in comparison with the *RSD* of *Ka* [19], the *RSD*'s of molar masses are considered negligible [30]. Hence, the *RSD* of the *Ka*-value [19] is the only important term of Eq. 4, which renders the corresponding uncertainty budget uncomplicated. Many methods are available to determine the *pKa* values [31] experimentally, and the values that are listed with sizeable uncertainties [32] are mostly determined using potentiometry [2, 19, 20]. The relations between $s(a(H^+))$ and $s(pH)$ may be expressed by the following equations:

$$s(pH) = \frac{1}{ln(10)} \cdot \frac{s\left(\mathrm{a}(H^+)\right)}{\mathrm{a}(H^+)} = \frac{10^{pH}}{ln(10)} \cdot s\left(\mathrm{a}(H^+)\right) \tag{5a}$$

$$\Rightarrow \quad s\left(\mathrm{a}(H^+)\right) = ln(10) \cdot s(pH) \cdot 10^{-pH} \tag{5b}$$

$$\Rightarrow \quad CV\left(\mathrm{a}(H^+)\right)\,\% = \ln(10) \cdot s(pH) \cdot 100\,\% \tag{5c}$$

The maximum absolute value of the uncertainty was calculated as $2 \cdot s(a(H^+)) = 0.068$ [17] for $pH = 1$ whereas the corresponding *CV* value of Eq. 4 depended on the *pH* value as portrayed in Fig. 2. With the result of Koort et al. [19] $s(a(H^+))$ could be estimated at 0.078 (Eq. 5b) the $s(K_a)/K_a$ (Eq. 4) and this corresponds well to *RSD* values of samples [32].

The proton activity was calculated as $a(H^+) = 10^{-pH}$ with 20 repetitions to calculate the corresponding *CV*-value of each *pH*-value of both electrodes (Fig. 2).

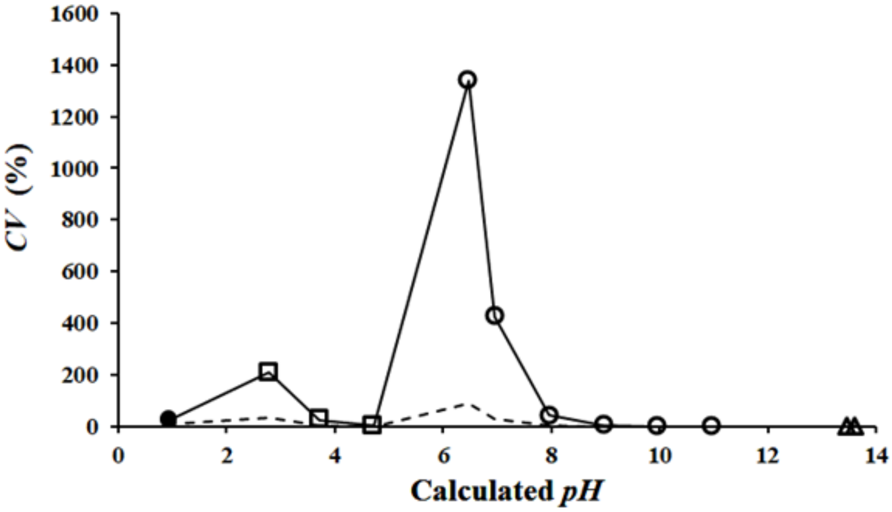


Figure 2. The *CV*-values of voltages (broken line, Eq. 3) and average proton activities (solid line, Eq. 4) depicted as a function of calculated *pH*-values for the four buffers systems ((●) HSO_4^-/SO_4^{2-}, (☐) CH_3COOH, CH_3COO^-, (○) NH_4^+/NH_3, (△) HCO_3^-/CO_3^{2-} that were applied to the analysis. Very high *CV* values were observed when the calculated *pH* values were significantly lower than the pK_a values. This pattern was not apparent when examining the average *CV* values of *pH* versus the calculated *pH* values (Fig. 4b, below). The broken and solid lines in the figure serve as visual aids.

It should be noted that the application of different buffers to the series of measurements influenced the level of *CV*-values of the proton activities, as the *CV*-values attained tremendous values at *pH*-values where the solutions contained little or no corresponding base (Fig. 2). This can be observed for both the acetic acid/acetate buffer starting at *pH* = 2.8 and ending at *pH* = 4.7 (Fig. 2, ☐). The ammonium/ammonia buffer starts at *pH* = 6.5 and ends at *pH* = 11 (Fig. 2, ○). The peaks in *CV*-values were also found in the data of the voltage (Fig. 2, broken line), according to Eq. 3. The large *CV*-values of buffers that have *pH* values distant from the *pKa*-values of the parent acid originate from the measurements only, which means that the buffers are perfectly stable and can be prepared with confidence, although perhaps with lower buffer capacity, for various purposes [27]. The only caveat is that the *pH* values cannot be appropriately measured with the *pH* meter when the *pH* value is much lower than the corresponding *pKa* value (Fig. 2, *pH* << *pKa*). Although the buffer capacity is at maximum when *pH* = *pKa*, it can also be increased by increasing the total concentration of the buffer

[28], which allows for the preparation of the buffer with other *pH* values. The present results add to this argumentation by including the increased uncertainties found using *pH* meters to measure *pH* values in buffers, especially with $pH < pKa$. Hence, buffers may be prepared within a wider interval around the *pKa*-value. Still, the target *pH*-value of the buffer should be measured by another method, such as ultra-violet visual spectrophotometry or fluorescence spectrophotometry. Notably, these results (Fig. 2) could only be identified after calculating the $s(a(H^+))$ from the *pH* values; the *pH* data did not reveal this trend (Figs. 3a-3c).

In-house experience with routine *pH* experiments showed occasional deviations from expected *pH* values or *pH* values that differed significantly from a series of repetitions in an inexplicable fashion. These relatively frequent observations prompted a further investigation into the performance of the *pH* meters where *pH* values were determined within a wide range of *pH* values using the four different types of buffers HSO_4^-/SO_4^{2-}, CH_3COOH/CH_3COO^-, NH_4^+/NH_3, and HCO_3^-/CO_3^{2-} (Figs. 3a-3c). The results of the *pH* measurements were depicted as a function of calculated *pH* values (Eqs. 1a-1d), where the results of many repetitions ($N = 260$) with the Thermo *pH* meter are shown in Fig. 3a. The results with an equal number of repetitions ($N = 260$) with the Basic 20 *pH* meter are shown in Fig.3b. While the data spread is similar across the entire *pH* range in both Figures 3a and 3b, there are some notable differences, especially at low and high *pH* values. The Thermo *pH* meter (Fig. 3a) showed larger standard deviations than the Basic 20 *pH* meter (Fig. 3b). At very low and very high *pH* levels, the Thermo *pH* meter occasionally produced extreme values. In contrast, the Basic 20 *pH* meter had slightly elevated standard deviations at *pH* 1 and 6.5, indicating better overall performance. The large deviations in the Thermo *pH* meter might be due to aging electrodes or experimental errors. However, with 40 repetitions at both $pH = 1$ and $pH = 14$, it is convincingly shown that the Thermo *pH* meter did not consistently perform well at extreme *pH* values in this series of experiments (Fig. 3a).

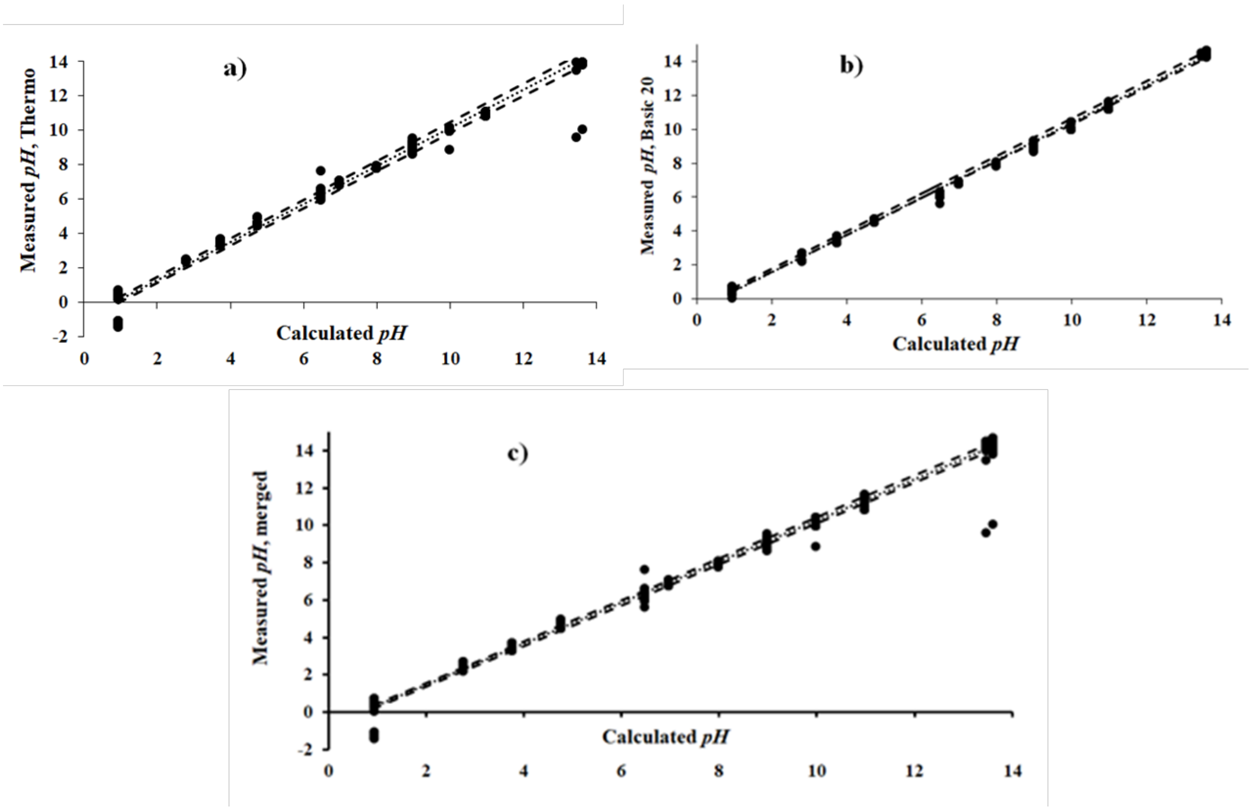


Figure 3. Measured *pH* values (*pH(meas)*) depicted as a function of calculated *pH* values (*pH(calc)*) using two different apparatuses (a) Thermo with the regression line, *pH(meas)* = (1.103 ± 0.017)·*pH(calc)* – (0.89 ± 0.14) and (b) Basic 20 with the regression line, *pH(meas)* = (1.0997 ± 0.0067)·*pH(calc)* – (0.604 ± 0.055). All solutions were equilibrated at 25 °C for the measurements. Dotted lines represent the least-squares regression lines, while broken lines indicate the corresponding confidence intervals. Due to 20 replicates per *pH* value, resulting in a total of 260 data points for

graphs (a) and (b), the confidence intervals are closely aligned with the regression lines. Significant overlaps between the confidence intervals led to the merging of all data, resulting in 520 data points in total for graph (c) with its regression line, $pH(meas) = (1.1015 \pm 0.0096) \cdot pH(calc) - (0.750 \pm 0.078)$.

Both electrodes captured elevated standard deviations close to $pH = 6.5$, which are related to the choice of buffers for the measurements (Fig. 2). However, the confidence lines (broken lines) were overlapping for most *pH* values, which indicates that the data of Figs. 3a and 3b belong to the same distribution of data. Hence, it is legitimate to merge the data according to the PoPC [33] (Fig. 3c) and perform the calculations of uncertainties according to all the data combined ($N = 520$). Notable deviations from the regression line were observed for data points at $pH = 1$, 6.5, 10, and 14, but in the present context, no outliers should be rejected [17]. Especially about *pH* measurements, the slight deviations between observed and expected values (Figs. 3-4) translate into large deviations when the proton concentrations and proton activities are considered (Eqs. 5a-c and Fig. 5). Thus, rejection of outliers may hide some of the characteristic features of the uncertainties that are key to understanding the method of measurement (Figs. 1-2). The rejection of outliers may be used to increase precision but only at the cost of loss of information regarding the uncertainties. Introducing the PoPC inevitably leads to increased uncertainties, which is also desirable, as it increases the chance of obtaining complete correspondence between predicted and observed *pH* values. Notably, it would not be right to reject outliers of Fig. 3 because that would make significant observations (Figs. 2 and 4) disappear. According to the PoPC, the figures of merits are the lower limit of analysis (*LLA*), the start of the best range (*SBR*), and the best relative uncertainty (*BRU*), which were determined as 0.051, 8.2, and 0.63 %, respectively. The LLA and SBR values suggest a wide interval between $0.051 < pH < 8.2$, where the *CV*-values decrease steadily from 50 % to the *BRU* of only 0.63 %. However, the *LLA* and the *LOD* would be expected to occur at high *pH values* where the proton activity is low and not at the low *pH values* suggested by the calculations. Thus, the meaning of the *LLA* in the present context refers to the limit of acceptable level of relative uncertainty rather than the lower limit of concentrations for the measurements. Therefore, other means of determining the *LLA* and *LOD* should be implemented. In clinical experiments, where large deviations in ionic strengths may be expected, the residuals around $pH = 7$ in a mixture of molecules were larger [34] than those observed in Figs. 3a-3b.

The *SD*s of the measured *pH* values of the merged data are shown in Fig. 4a (circles, each with $n = 40$), and the corresponding *CV* values according to the theories of IUPAC/QUAM and PoPC are indicated by the broken line and full line, respectively. Close to $pH = 1$, the experiments and theories are widely different, with a factor of two too large values as compared with those predicted by the IUPAC/QUAM formula and a factor of two predictions too small for the PoPC formula (Fig. 4a). Within the interval $2 < pH < 11$, the *SD*s were approximately constant with an average value of $\langle s(pH) \rangle \sim 0.18$ that translates into an expanded uncertainty of $U = 0.36$ ($k = 2$). Thus, this value of the expanded uncertainty was 36 times above the stipulated uncertainty of 0.01 indicated by the manufacturers of the *pH* meters and 30 times above the uncertainty of 0.012 obtained for the Harned cell under controlled conditions [5].

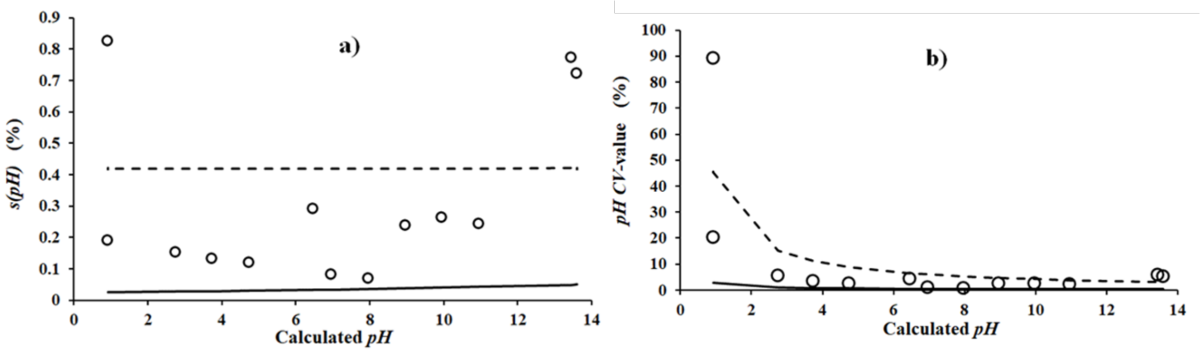


Figure 4. Evaluation of (a) *SD*s and (b) *CV* values as a function of calculated *pH* values (Eqs. 1a-1d), according to calculations that were performed by means of the IUPAC/Eurachem formula (broken lines) [17] and the PoPC (solid lines) [33] using the data of Fig. 3c. The experimental *SD*s displayed a

characteristic U-shape with large *SD*s at low and high *pH* values and almost constant *SD*s within the interval $2 < pH < 11$ in (a). In both (a) and (b), the experimental data fall amidst of the two theories, except at *pH* ~ 1 and 14.

Both the IUPAC/QUAM and PoPC models predicted *CV* values for the *pH* range of 2 to 11 (Fig. 4b), with the former generally predicting values above the experimental results and the latter below. The systematic variation in *CV* values of proton activities was only faintly reflected in the corresponding *CV* values of the *pH* results (Fig. 4b). At $pH = 6.5$, an increased deviation corresponding to observations in Figs. 3a-3c resulted in a *CV* of 4.5% in Fig. 4b. Although *CV* values ranged from 0.9% to 90%, most remained below 10% (Fig. 4b). The predominant *CV* values of 0.9% to 10% (Fig. 2b) represent nearly an order of magnitude increase when considering proton activity, which should be the guiding factor in determining measurement uncertainty. The results in Fig. 4a also show that the electrodes do not perform well at extreme *pH* values, although the *CV* value remained below 10%, even at *pH* ~14 (Fig. 2b).

To further illustrate the uncertainty issue, the *CV*-values according to the average value of proton activities (*CV(a(H*$^+$*))*) were calculated pairwise for the couples of proton activities of the 260 repetitions of each electrode (Figs. 5a-5b).

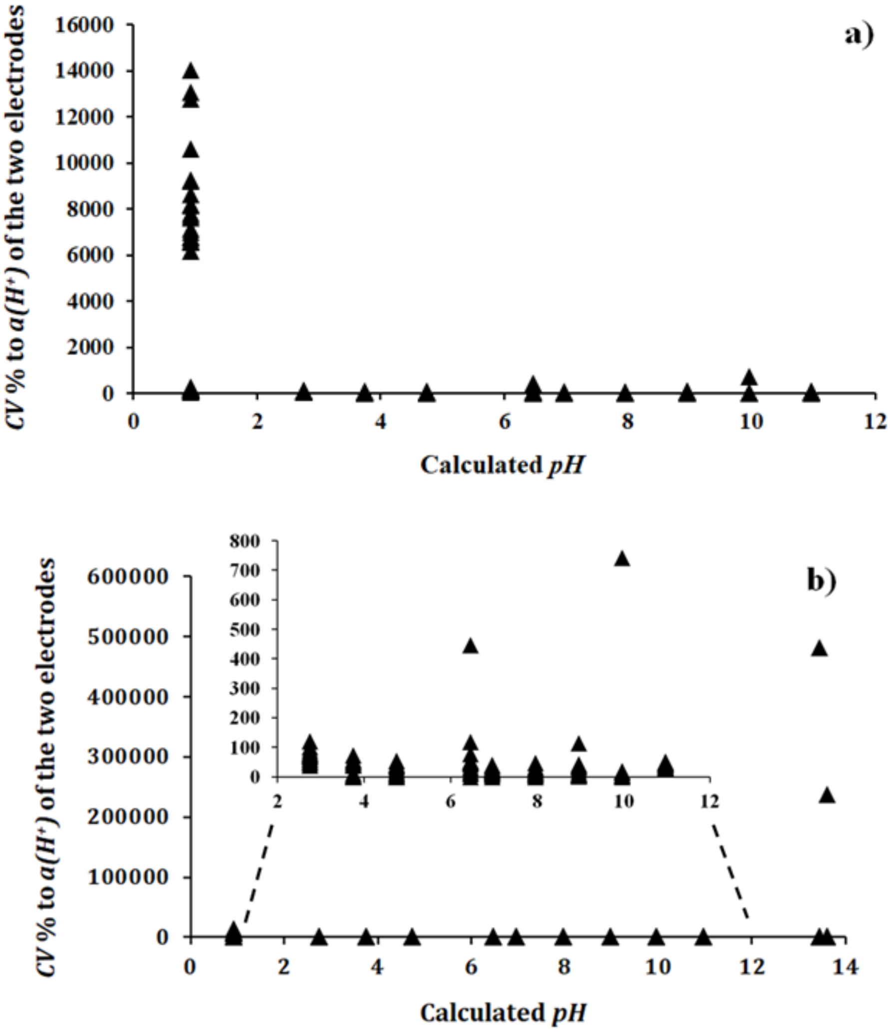


Figure 5. The *CV* values of average proton activities (*CV(a(H*$^+$*))*) are shown as a function of calculated *pH* values. Proton activities were estimated using Eqs. 1a-1d for each electrode, and the standard deviation between the two electrodes was determined for each of the 260 data pairs. To

enhance clarity, the diagram is divided into two sub-diagrams: (a) displaying *CV* values for $pH < 12$, and (b) encompassing all data, with an inset highlighting *CV*-values for $3 < pH < 12$. Notably large *CV* values were observed at *pH* values near 1 and 14.

At the extreme *pH* values, the results were very different for the two electrodes, as evidenced by the *CV* values reaching up to $1.4\text{x}10^4$% (Fig. 5a) and $5.0\text{x}10^5$% (Fig. 5b) for *pH values* close to 1 and 14, respectively, which amplifies the observations of Figs. 4a and 4b. Such *CV* values more strongly indicate that no valuable measurements can be obtained at the extreme *pH values* when the average values of the two electrodes were considered. The spikes at $pH = 10$ and 14 may be accidental, as they were primarily generated by the Thermo *pH meter*, whereas the smaller spike at *pH* around 6.5 (Fig. 5b, inset) was captured by both electrodes (Figs. 3a and 3b). While this Thermo electrode may have been faulty and struggled to perform well at extreme *pH* values, it occasionally produced the expected results at both ends of the *pH* interval (Fig. 3a). Within the *pH* range of 1 to 11, both electrodes demonstrated similar performance. The spikes observed at high *pH* values could indicate the method's tentative limit of detection (*LOD*) near *pH* 10. However, since these spikes are infrequent, more data than currently reported might be needed to determine the LOD accurately.

The *pH* values are measured to determine proton activities, as proton activity is the primary measurment. This makes Figs. 2, 5a, and 5b crucial tools for decision-making regarding proton activities. The definition of *pH* involves taking the negative logarithm of the proton activity to the base of 10, which obscures key information about measurement uncertainty (as shown in Figs. 4a and 4b) and eliminates spikes often seen at high *pH* values (Fig. 5b). Because the *pH* definition loses significant information about the uncertainty of proton activity ($a(H^+)$), it is proposed to avoid this mathematical transformation and instead focus on reducing noise levels to prevent potential spikes. Therefore, future *pH* meters should be optimized based on proton activities to provide linear responses, which will help identify key factors leading to the development of improved *pH* meters.

TABLE I. Overview of *pH*-values that were measured in the samples of milk, Red Bull™, fruit juice and diet Coca Cola™ consecutively each with 10 replicates ($n = 10$) at 25ºC according to the interval of *pH*-values with the lowest uncertainty (Fig. 4b). Abbreviations: Standard deviation (*SD*), 95 %-confidence interval (*CI*), expanded uncertainty (U, $k = 2$).

Sample	Milk		Red Bull™		Fruit juice		Diet Coca Cola™	
Apparatus	Thermo	Basic 20	Thermo	Basic 20	Thermo	Basic 20	Thermo	Basic 20
Average *pH*	6.495	6.371	3.197	3.264	3.106	3.189	2.564	2.696
SD(pH)	0.016	0.070	0.0067	0.0097	0.0070	0.014	0.018	0.033
CI(pH)	0.012	0.050	0.0048	0.0069	0.0050	0.010	0.013	0.024
CV(pH) %	0.25	1.1	0.21	0.30	0.23	0.45	0.69	1.2
U(pH) k =2*	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36
U(pH)/CI(pH)	31	7.2	75	52	72	35	28	15

*Average of the values in Fig. 4a excluding the spike values at $pH = 1$ and 14.

The *pH* values of samples (Table 1) were measured within the reliable range of measuring *pH* values $2 < pH < 6$, according to the *CV* values of Fig. 4b. The ten consecutive replicates yielded small standard deviations that indicated the *pH* value should be reported with 2 - 4 decimal points, following earlier suggestions [12, 19]. However, the results of the present study show that only one decimal point is more likely to be correct for routine *pH measurements*, as indicated by the expanded uncertainty of $U(pH) = 0.36$ (Table 1, $k = 2$). This expanded uncertainty did not exceed the difference of approx. 0.5 between *pH* values that were calculated with and without adjustments of concentrations with activities [9]. However, the expanded uncertainty exceeds the influence of temperature on *the pH* values of tartrate, phthalate, and phosphate buffers [10, 16]. Although numerically smaller, the results of CCQMK91 indicated that those *pH* values would also be determined to one decimal point if data of the measurements at 25°C were not rejected [35]. If unwanted data continues to be rejected, conducting statistical evaluations of interlaboratory comparisons would be pointless [36]. Table 1 highlights a common problem in contemporary analytical chemistry: the confusion between precision and accuracy.

Although repeated measurements within a short period yield small standard deviations and averages that match certified values by chance, significant discrepancies in *pH* values were found between the two *pH* meters for all the samples. These discrepancies became evident when comparisons were made using confidence intervals (*CI*s, $p = 0.05$). Conversely, the *pH* values appeared similar when the expanded uncertainty was applied for comparison. Therefore, one possible approach is to average the *pH* values obtained from the two electrodes when presenting results to potential customers. However, this method would likely result in customers obtaining significantly different results from another laboratory, which is unacceptable in analytical chemistry. Reproducibility is a critical requirement. Ratios as high as 75, but typically around an order of magnitude, were observed between the expanded uncertainty and the confidence interval (*CI*) (Table 1). This highlights the distinction between intra-day and inter-day precision of the methods. Intra-day precision, determined by the equipment manufacturer, is considered less relevant to analytical chemistry and does not need to be repeatedly reported in scientific publications.

4. CONCLUSIONS

Significant differences were observed between the uncertainties in routine *pH* measurements and those encountered under highly controlled laboratory conditions. This study highlights the significance of retaining outliers in the data set. Inter-laboratory comparisons often use weighting schemes to calculate uncertainties, eliminate outliers, and report low uncertainties. This practice is likely intended to create certified reference materials in response to customer demands for low uncertainty levels, which may be mistakenly perceived as an indicator of measurement quality. Quality of measurement should not be defined as merely producing results that satisfy customers' often unrealistic expectations or those reported with low uncertainties. Instead, it should be associated with providing evidence of the results' closeness to the valid values, considering the conditions of traceability and statistical control. Due to the extensive independent replicates conducted with *pH* meters from two different brands in this study, an expanded uncertainty of 0.36 is universally applicable for routine *pH* measurements. The magnitude of the expanded uncertainty of 0.36 indicates that adjustments of activities using activity coefficients and temperature to obtain the proton concentrations are redundant or unnecessary at low to medium proton concentrations. Moreover, the reliable range for *pH* measurements should be limited to $1 < pH < 11$, or possibly even more narrowly to $1 < pH < 6.5$ if the spikes were to be considered.

Within the *pH* range of 1 to 11, routine *pH* measurements were found to have an uncertainty more than 30 times greater than the manufacturers' specifications and those conducted under controlled conditions. Consequently, the pH meter's readability should only be reduced to one or two decimal points. Therefore, it is suggested that buffers be prepared with *pH* values close to the *pKa* value of the corresponding acid, not because it simplifies *pH* calculation using the Henderson-Hasselbalch equation but because the measurement uncertainty is lower around $pH = pKa$ and beyond. However, while buffers can be conveniently prepared at *pH* values much lower than the *pKa*, the reliability of measuring these *pH* values is significantly compromised.

Acknowledgments

We sincerely appreciate the financial support provided by the Botswana International University of Science and Technology. Additionally, we extend our heartfelt thanks to the Organising Committee of the North American Chemical Residue Workshop (NACRW) for accepting our oral presentation at the workshop in 2022.

References

[1] JCGM, Joint Committee for Guides in Metrology 200 : 2008 “International vocabulary of metrology — Basic and general concepts and associated terms (VIM) Vocabulaire international de métrologie — Concepts fondamentaux et généraux et termes associés (VIM)”, International Organization for Standardization Geneva ISBN 3 (2008) 104. https://doi.org/10.1016/0263-2241(85)90006-5.

[2] R.P. Buck, S. Rondinini, A.K. Covington, F.G.K. Baucke, C.M.A. Brett, M.F. Camões, M.J.T. Milton, T. Mussini, R. Naumann, K.W. Pratt, P. Spitzer, G.S. Wilson, “Measurement of pH. Definition, standards, and

procedures (IUPAC Recommendations 2002)", *Pure and Applied Chemistry* 74 (2003) 2169–2200. https://doi.org/10.1351/pac200274112169.

[3] F.G.K. Baucke, "New IUPAC recommendations on the measurement of pH - Background and essentials", *Anal. Bioanal. Chem.* 374 (2002) 772–777. https://doi.org/10.1007/s00216-002-1523-4.

[4] J.F. Waters, M. Máriássy, F. Bastkowski, T. Asakai, S. V. Prokunin, F.B. Gonzaga, L. Dimitrova, R.C. Amaya, H.T. Quezada, A. Vospelova, M. Rozikova, A. Snedden, D. Stoica, B. Sander, B. Jakusovszky, I. Maksimov, S. Onuma, T. Suzuki, J.M. Ruiz, J.L. Ortiz-Aparicio, G. Tincona, J. Vasquez, M. Pawlina, W. Kozlowski, V.I. Dobrovolskiy, A. V. Aprelev, Z. Hanková, P. Rodruangthum, E. Uysal, L. Liv, N. Nakiboglu, V. Gavrilkin, A. Petrenko, O. Manska, R.A. Easley, S. Fajardo, "Key comparison CCQM-K18.2016 pH of carbonate buffer final report (15.07.2020)", *Metrologia* 57 (2020). https://doi.org/10.1088/0026-1394/57/1A/08022.

[5] G.M. Marion, F.J. Millero, M.F. Camões, P. Spitzer, R. Feistel, C.T.A. Chen, "PH of seawater", *Mar. Chem.* 126 (2011) 89–96. https://doi.org/10.1016/j.marchem.2011.04.002.

[6] R.A. Easley, R.H. Byrne, "Spectrophotometric calibration of pH electrodes in seawater using purified m-cresol purple", *Environ. Sci. Technol.* 46 (2012) 5018–5024. https://doi.org/10.1021/es300491s.

[7] R. de Levie, "A pH centenary", *Electrochim. Acta* 135 (2014) 604–639. https://doi.org/10.1016/j.electacta.2014.04.006.

[8] S. Lainela, I. Leito, A. Heering, G. Capitaine, B. Anes, F. Camões, D. Stoica, "Toward unified ph of saline solutions", *Water* (Switzerland) 13 (2021) 1–12. https://doi.org/10.3390/w13182522.

[9] M. Hippler, G.D. Metcalfe, "Using Activities to Correct the Henderson-Hasselbalch Equation", *Bunsen-Magazin*, 22 (2020) 102–105. https://doi.org/10.26125/y7p7-an56.

[10] M.F. Camoes, M.J.G.H.M. Lito, M.I.A. Ferra, A.K. Covington, "Consistency of pH standard values with the corresponding thermodynamic acid dissociation constants (Technical Report)", *Pure and Applied Chemistry* 69 (1997) 1326–1333. https://doi.org/10.1351/pac199769061325.

[11] I. Maksimov, M. Ohata, S. Nakamura, A. Hioki, K. Chiba, P. Spitzer, "pH determination on a carbonate buffer by Harned cells of different designs", *Accreditation and Quality Assurance* 13 (2008) 381–387. https://doi.org/10.1007/s00769-008-0399-1.

[12] R.J.C. Brown, P.J. Brewer, M.J.T. Milton, "On the validity of acidity function extrapolation to zero added ionic strength and its impact on the uncertainty of primary pH measurements using the Harned cell", *Measurement* (Lond) 46 (2013) 1726–1733. https://doi.org/10.1016/j.measurement.2013.01.007.

[13] M.J. Guiomar, H.M. Lito, M.F. Camões, C.M. Viçoso, "Improving the quality of potentiometric pH measurements", *Accreditation and Quality Assurance* 12 (2007) 447–453. https://doi.org/10.1007/s00769-007-0296-z.

[14] P.J. Brewer, R.J.C. Brown, "Effect of structural design of silver/silver chloride electrodes on stability and response time and the implications for improved accuracy in pH measurement", *Sensors* 9 (2009) 118–130. https://doi.org/10.3390/s90100118.

[15] B. Anes, R.J.N. Bettencourt da Silva, H.F.P. Martins, C.S. Oliveira, M.F. Camões, "Compatibility of activity coefficients estimated experimentally and by Pitzer equations for the assessment of seawater pH", *Accreditation and Quality Assurance* 21 (2016) 1–7. https://doi.org/10.1007/s00769-015-1163-y.

[16] R.G. Bates, E.A. Guggenheim, "Report on the Standardization of pH and Related Terminology", *Pure Appl. Chem.* 1 (1960) 163–168.

[17] CITAC, Eurachem, "Quantifying Uncertainty in Analytical Measurement", BIPM 3rd ed. (2012) 1–133. https://doi.org/0 948926 15 5.

[18] E.K.& V.P. I. Leito, Liisi Strauss, "Estimation of uncertainty in routine pH measurement", *Accreditation and Quality Assurance* 7 (2014) 242–247. https://doi.org/https://doi.org/10.1007/s00769-002-0470-2.

[19] E. Koort, K. Herodes, V. Pihl, I. Leito, "Estimation of uncertainty in pKa values determined by potentiometric titration", *Anal. Bioanal. Chem.* 379 (2004) 720–729. https://doi.org/10.1007/s00216-004-2586-1.

[20] Y. Villasana, N. Moradi, C. Navas-Cárdenas, G.S. Patience, "Experimental methods in chemical engineering: pH", *Canadian Journal of Chemical Engineering* 100 (2022) 1703–1717. https://doi.org/10.1002/cjce.24393.

[21] P. Spitzer, S. Seitz, "Metrological traceability chain for pH measurement results", *Mapan - Journal of Metrology Society of India* 25 (2010) 191–196. https://doi.org/10.1007/s12647-010-0019-5.

[22] J.E.T. Andersen, K. Moemedi, K. Katse, "Modeling uncertainty by the principle of pooled calibrations and issues with method validations may influence consensus in science", *Measurement* (Lond) 188 (2022) 110382. https://doi.org/10.1016/j.measurement.2021.110382.

[23] T. Scientiific, "pH measurement handbook", *PH Measurement Handbook* (2023) 1–28. https://assets.thermofisher.com/TFS-Assets/LSG/brochures/pH-Measurement-Handbook-S-PHREFBK-E.pdf (accessed May 14, 2024).

[24] C. Instruments, "pH-meter Basic 20", *Julho* (2008) 1–10.

[25] Crison, Technical data sheet, "Basic 20" (2008) 1–2. https://www.crisoninstruments.com/site/pdfs_en/basic_20.pdf (accessed May 14, 2024).

[26] W.P. Jencks, F.H. Westheimer, R. Williams, "pKa Data Compiled by R. Williams", *ACS, Organic Division* (2022) 1–33. https://organicchemistrydata.org/hansreich/resources/pka/pka_data/pka-compilation-williams.pdf (accessed May 14, 2024).

[27] J.E.T. Andersen, "Comments on the Concept of Buffers", *Chemistry Africa* (2024). https://doi.org/10.1007/s42250-023-00848-4.

[28] D.C. Harris, "Quantitative Chemical Analysis, 8th ed.", Clancy Marshall, W.H. Freeman, New York, 2010.

[29] CODATA Committee on data, "CODATA Recommended Values of the Fundamental Physical Constants", *CODATA Recommended Values of the Fundamental Physical Constants* (2014). https://codata.org/blog/2015/08/04/codata-recommended-values-of-the-fundamental-physical-constants-2014/ (accessed May 14, 2024).

[30] T. Prohaska, J. Irrgeher, J. Benefield, J.K. Böhlke, L.A. Chesson, T.B. Coplen, T. Ding, P.J.H. Dunn, M. Gröning, N.E. Holden, H.A.J. Meijer, H. Moossen, A. Possolo, Y. Takahashi, J. Vogl, T. Walczyk, J. Wang, M.E. Wieser, S. Yoneda, X.K. Zhu, J. Meija, "Standard atomic weights of the elements 2021 (IUPAC Technical Report)", *Pure and Applied Chemistry* 94 (2022) 573–600. https://doi.org/10.1515/pac-2019-0603.

[31] J. Reijenga, A. van Hoof, A. van Loon, B. Teunissen, "Development of methods for the determination of pKa values", *Anal. Chem. Insights* 8 (2013) 53–71. https://doi.org/10.4137/ACI.S12304.

[32] J. Wiora, "Measurement Uncertainty Calculations for pH Value Obtained by an Ion-Selective Electrode", *Sensors* (2018). https://doi.org/10.3390/s18061915.

[33] J.E.T. Andersen, K. Moemedi, K. Katse, "Modeling uncertainty by the principle of pooled calibrations and issues with method validations may influence consensus in science", *Measurement* (Lond) 188 (2022) 110382. https://doi.org/10.1016/j.measurement.2021.110382.

[34] T. Ring, J.A. Kellum, "Modeling Acid-Base by Minimizing Charge-Balance", *ACS Omega* 4 (2019) 6521–6529. https://doi.org/10.1021/acsomega.9b00270.

[35] P. Spitzer, F. Bastkowski, B. Adel, L. Dimitrova, F.B. Gonzaga, P.T. Jakobsen, P. Fisicaro, D. Stoica, T. Asakai, I. Maksimov, Z.N. Szilágyi, A. Reyes, M. Monroy, G.T. Canaza, W. Kozlowski, M. Pawlina, V. Kutovoy, L. Vyskocil, A. Mathiasova, T. Nongluck, F. Ficicioglu, V. Gavrilkin, K.W. Pratt, "Final report on CCQM-K91: Key comparison on pH of an unknown phthalate buffer", *Physikalisch-Technische Bundesanstalt Report* (2013) 1–24. https://doi.org/10.1088/0026-1394/50/1a/08016.

[36] D.L. Duewer, K.W. Pratt, C. Cherdchu, N. Tangpaisarnkul, A. Hioki, M. Ohata, P. Spitzer, M. Máriássy, L. Vyskočil, "Degrees of equivalence for chemical measurement capabilities: primary pH", *Accreditation and Quality Assurance* 19 (2014) 329–342. https://doi.org/10.1007/s00769-014-1076-1.