DE GRUYTER
OPEN

DOI: 10.1515/ssa-2017-0022

SOIL SCIENCE ANNUAL
Vol. 68 No. 4/2017:174–181

IZABELA KUNA-BRONIOWSKA[1], HALINA SMAL[2*]

[1]*University of Life Sciences, Department of Applied Mathematics and Computer Science*
ul. Głęboka 28, 20-612 Lublin, Poland
[2]*University of Life Sciences, Institute of Soil Science, Engineering and Environment Management*
ul. Leszczyńskiego 7, 20-069 Lublin, Poland

Statistical measures of the central tendency for H^+ activity and pH

Abstract: Despite the numerous papers on the statistical analyses of pH, there is no explicit opinion on the use of arithmetic mean as a measure of the central tendency for pH and H^+ activity. The problem arises because the transformation of the arithmetic mean for one does not give the arithmetic mean for the other. The paper presents 1) the theoretical considerations on the distribution of pH and H^+ activity and relation between them, properties of these distributions, the choice of distributions which should be consistent with the distribution of pH and the distribution of H^+ activity and measures of central tendency for features of such distributions and 2) examples of calculations of measures of central tendency for pH and H^+ activity based on the literature data on soil and lake water pH. These data analyses included distributions of pH and H^+ activities, properties of distribution, descriptive statistics for pH and for the H^+ activity and comparison of arithmetic mean with the geometric mean. From the results, it could be concluded that a uniform approach to the choice of measure for the central tendency of pH and H^+ activity requires the determination of the type of measure (mean) for one of them and then consistent transformation of this measure. The choice of measure of the central tendency for the variable should be preceded by determination of its distribution. Normal probability distribution of pH and thus lognormal distribution of H^+ activity indicate that the arithmetic mean, and its corresponding geometric mean should be used as proper measures of the central tendency for pH and for H^+ activity. Besides, the position statistic that is a median can be used for each of those variables, irrespective of their probability distributions.

Keywords: H^+ activity, pH, arithmetic mean, geometric mean, median, normal distribution, lognormal distribution

INTRODUCTION

In the natural environment, the reaction measured as H^+ activity or its log function (pH) is a very important factor that controls many chemical, physiochemical and biological processes. pH is one of the basic properties commonly determined in soil, water and other environmental studies. Very often, analyses of the results and conclusion require statistical calculations including the presentation of a single mean pH or H^+ activity value for the set of measurements characterizing a given environment. As an example, for monitoring of wet deposition of H^+ ions introduced to the surface (soils, waters) with atmospheric precipitation, the calculation of mean hydrogen ions H^+ activity is required.

As a result of the use of two measurement scales, i.e. H^+ activity and its logarithm, the question arises as to how to perform statistical analysis of the measured results of this feature. Especially, a problem arises when one would like to use the arithmetic mean, which needs a decision for which featured pH or H^+ activity this descriptive statistic should be used. As the subject literature shows, according to some scientists (Pace et al. 1979, Boyle 1991), the arithmetic mean should be calculated directly from the pH values according to the equation:

$$\overline{pH} = \frac{pH_1 + pH_2 + \cdots + pH_n}{n} \tag{1}$$

This is how many authors calculate the mean value of pH (Dangles and Guérold 2000, Simon et al. 2006, Larsen et al. 2007, Forsberg et al. 2008).

Another opinion (Barth 1975, Murphy 1981) is that before computation of mean pH one should transform the original results of pH into H^+ activities, calculate the arithmetic mean of H^+ activity $(\overline{H^+})$ and then reconvert it to pH, according to the equation:

$$-\log \overline{H^+} = -\log\left(\frac{10^{-pH_1} + 10^{-pH_2} + \cdots 10^{-pH_n}}{n}\right) \tag{2}$$

but $\overline{pH} \neq -\log \overline{H^+}$

This method is also used by researchers (Wesselink et al. 1996, Nőges et al. 2003, Smal and Olszewska 2008). The US Geological Survey recommends it in reporting water pH of over time or in space (USGS 2008). Also, in Poland it is officially recommended for averaging (as weighted mean) of pH values in the studies of environmental monitoring (Kostrzewski et al. 2006, Bochenek 2014).

* *Prof. dr hab. H. Smal, halina.smal@up.lublin.pl*

Besides the mathematical considerations of applying one of the above methods, researchers attempted experimentally to answer the question “which procedure is better in expressing a central tendency of acidity in the series of solutions (samples) that differed in *pH* values”. The results were not explicit. For example, Giesecke (1979) mixed five solutions of HCl of known *pH* and measured the *pH* of the resulting solution. From that experiment the author concluded that the arithmetic mean of hydrogen ion concentration was a better measure of the central tendency of acidity in a series of solutions than the arithmetic mean of their *pH* values. Contrarily, Baker et al. (1981) revealed the close agreement between arithmetic mean *pH* values of individual soil samples and the *pH* values of the mixed samples. Moreover, he found that the agreement was consistently and considerably poorer when *pH* values were converted to H^+ activities before averaging. In another study, Boutilier and Shelton (1980) stated that both the mean of *pH* and of hydrogen ion concentrations seemed to be equally good and acceptable in statistical analysis on the basis of *pH* values of blood samples. This is true when the distributions are symmetrical. However, the central tendency measure of a set of data is not the same as the value of *pH* or mixed sample hydrogen ion concentration value. An arithmetic or geometric mean is the only information about mean level of *pH* or H^+ activities for a set of separate samples. Therefore, one should be very careful while using these means in calculations of *pH* values for mixtures. In mixed samples, the reactions influencing the ions activity and so called “border solutions“ may occur. For the reasons mentioned above, using the arithmetic mean as a measure for *pH* or as a measure for the value of H^+ activity of mixtures can lead to opposite conclusions, as can be seen from the work of Giesecke (1979) and Baker et al. (1981).

Yang et al. (2004) studied the consistency of *pH* and H^+ activity distributions with normal distribution using the data of three soil systems. The authors found a better consistency with normal distribution for soil *pH* than for H^+ activity. In spite of that, the authors suggested that calculating arithmetic mean for *pH* values may be inappropriate because of its lack of consistency with the arithmetic mean calculated for H^+ activity.

In the monthly Journal Water Pollution Control Federation 1974–1975, 27 authors suggested the use of the arithmetic mean as a measure of the central value for *pH*. Finally, following a series of articles in this journal, a summary discussion was posted in which the disputants supported the predecessors in favour of the arithmetic mean for *pH*, which is equivalent to the geometric mean for H^+ activity (Middleton and Rovers 1976). However, it should be pointed out that the discussion carried out by the authors did not fully justify such a view. Neither did they analyse the *pH* nor H^+ activity distributions nor the relationships between them.

Recently, Gruba et al. (2010) carried out statistical analyses on two sets of soil *pH* dat: one with the distribution consistent with normal distribution and the other with distribution not consistent with normal distribution. In their conclusions, the authors suggested the median as a measure of a central tendency for *pH*. However, the authors did not attempt to identify the type of distribution for H^+ activity, which has a significant effect on the choice of an appropriate measure of the central tendency for H^+ activity, and limited their choice to the positional statistics such as that which is the median.

In the light of the above, one can say that the choice a measure of central tendency for *pH* and H^+ activity as well as the question which of the presented methods is better are still valid. The research conducted so far and presented in the literature merely concentrated on checking the consistency of H^+ activity distribution with the normal distribution (Yang et al. 2004). Normality of distribution is the prerequisite for the applicability of most statistical procedures such as ANOVA or geospatial analysis distribution. These studies did not include an analysis of the nature of the H^+ activity distribution and the impact of logarithmic transformation on this distribution.

This paper aims to discuss and organize the relationship between distributions of pH and the distribution of H^+ activity and the relationship between measures of their central tendencies. It presents the analyses of the distribution of *pH* values and H^+ activity carried out in parallel. The properties of these distributions such as symmetry, left- or right skewness and kurtosis were determined. Based on the analysis, the distributions were selected together with their consistent *pH* and H^+ distributions. Also, the appropriate measures of central tendency for features of such distributions were chosen. Then, it was checked whether the logarithmic transformation leads to a mutual correspondence between the selected distributions, as well as between the measures of central tendency related to these distributions. Also, the properties of arithmetic and geometric means and their relationship with normal and lognormal distributions were widely described. The distributions of *pH* values and H^+ activities were analysed for the soil and lake water *pH* data. Moreover, arithmetic and geometric means were calculated and compared in terms of their median position.

THEORETICAL CONSIDERATIONS

Properties of logarithmic scale of pH

Hydrogen ions concentration is distributed over a wide range, i.e. from 10^{-14} in 1M NaOH to 1.0 in 1M HCl. In 1909, in order to simplify the recording and "flattening" of the the molar concentration scale, Sørensen defined a logarithmic scale (*pH*) as the negative logarithm of the hydrogen (hydronium) ions activity: $pH = -\log10(H_3O^+)$.

A logarithmic scale is a nonlinear measurement scale. This scale does not maintain the distances in absolute terms, whereas it maintains the distances in relative (percentage) terms. A logarithmic transformation "flattens" the scale because it reduces longer distances to a greater extent than shorter distances. This transformation is monotonic, i.e. it preserves data ordering. The ordering of values remains the same regardless of whether they are used in the original values or their logarithms (*if* $a < b$ *then* $\log a < \log b$ *and* $-\log a > -\log b$).

The important property results from the monotonicity of the logarithmic transformation; namely the observation located in the middle of the original scale will also be in the middle of the logarithmic scale. The logarithmic scale can only be used to map the positive values. It is, in addition to a linear scale, the scale most often used in graphs. It should be noted that in some cases, a logarithmic scale is a natural due to the fact that the senses (sight, hearing or sense of temperature) react to stimuli in a logarithmic not linear manner (Weber's-Fechner's law).

Statistical measures of central value

Several measures of the central tendency of the examined feature can be found in the statistical literature. The most important is the arithmetic mean, but the geometric and harmonic means are also used. Positional measures like the median for example can also be found. The median may be a better indicator of the most typical value if a set of scores has an outlier. However, when the sample size is large and does not include outliers, the mean score usually provides a better central tendency measure.

Analysis of the feature's distribution should precede feature selection. If the feature distribution is consistent with the normal distribution, or is at least symmetrical, the arithmetic mean of such a feature is an appropriate measure of its central tendency. However, if the test feature distribution shows right skewness, the appropriate measure of central tendency of this feature is the geometric mean.

The problem of measuring of central tendency for *pH* and H^+ activity is more complicated because *pH* is the negative logarithm of H^+ activity. Regardless of which measure we take for *pH* and H^+ activity, the question will always arise as to whether the transformation (log or antilog) of the selected measure of central tendency for one of these features leads to a proper measure for the other.

The properties and conditions for use of the arithmetic and geometric means

Arithmetic mean

The arithmetic mean is the most popular measure of location. For the variable that takes values the arithmetic mean is calculated according the formula:

$$\bar{x}_a = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3}$$

The arithmetic mean from the data set is a good measure of the central tendency but it is greatly influenced by outliers (Svincov and Cambell 2002, Krysicki et al 2007). For skewed distributions, particularly the right skewed and lognormal ones, the arithmetic mean may not accord with one's notion of „middle", and the geometric mean or robust statistics, such as the median, may be a better description of central tendency.

Geometric mean

It is defined as the root of the product *n* from *n* values of the variable:

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots\ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} \tag{4}$$

According to the above definition, the geometric mean can be determined only when the observations are positive numbers and different from zero. Since the extraction of high degree root is difficult and the number *n* in a statistical series can be large, the logarithmic form of the geometric mean is often used; i.e. the arithmetic mean of the logarithms of values of the variable:

$$\log \bar{x}_g = \log \sqrt[n]{x_1 \cdot x_2 \cdot \ldots\ldots \cdot x_n} = \frac{1}{n}\sum_{i=1}^{n} \log x_i = \overline{\log x}$$

or

$$\log \bar{x}_g = \frac{1}{n}\sum_{i=1}^{n} f_i \cdot \log x_i = \overline{\log x}$$

The geometric mean reflects the effect of extreme values to a lesser extent than the arithmetic mean. The geometric mean is always smaller or equal to the arithmetic mean: $\bar{x}_g \le \bar{x}_a$

The geometric mean is related to the lognormal probability distribution. This distribution is a right-skewed distribution, where the measure of the skewness depends on the expected value and variance of the variable logarithm (Gaddum 1945).

The geometric mean is a better measure of central tendency than the arithmetic mean when the values of a statistical feature are relative measures (indexes, percentages), the collected results are several orders of magnitude and show a clear right skewness, several observations have a value far greater than the others and there are no essential arguments for the omission of extreme values. An indication can be found in the literature that the highest value to be omitted must be at least three times higher than the lowest.

When the data cover a narrow range of scale, or if they show left skewness distribution, the geometric mean and logarithmic transformation may be inappropriate. One cannot use the geometric mean for data that are logarithmically transformed, such as *pH* or decibels (dB).

Characteristics of *pH* and H^+ activity distribution

By introducing a *pH* as the negative logarithm of the H^+ activity, Sorensen's transformation reduces H^+ variability and normalizes its distribution. The logarithmic transformation reduces the right skewness by a greater compression of results from the right end of the original set and stretches the results from the left end of the set. For this reason in general, it is more likely that the distribution of *pH* will tend to have a distribution consistent with the normal distribution than H^+. Given the logarithmic transformation of H^+ activity, pH should characterized by symmetric distribution or distribution of a small left or right skewness (Young et al. 2004, Gruba et al. 2010), which in turn indicates that one should expect right skewness of the H^+ activity distribution. The lognormal distribution is right skewed, for this reason one should look for the consistency of distribution of H^+ activity with log-normal distribution but not with normal distribution.

At the normality of the pH distribution and right skewed H^+ activity distribution, one should use the geometric mean as a measure of central tendency for H^+ activity and the arithmetic mean for *pH* values.

Due to the logarithmic relationship between *pH* and H^+ activity, the geometric mean for H^+ activity, through logarithmic transformation, is equal to the arithmetic mean for *pH*.

The logarithmic transformation of the geometric mean

$$\overline{H}_g^+ = \sqrt[n]{H_1^+ \cdot H_2^+ \cdot \ldots\ldots \cdot H_n^+}$$

of H^+ activity leads to the arithmetic mean of $\overline{pH}_a$.

In addition, the expected value m_L and variance σ^2_L of random variable X with lognormal distribution depend on the expected value m and variance σ^2 of ln$X$ (logarithmically transformed variable X) according to the formula (Krishnamoorthy and Mathew 2003):

$$\mu_L = e^{\mu + \frac{\sigma^2}{2}}, \sigma_L^2 = \left(e^{\sigma^2} - 1\right) \cdot e^{2\mu + \sigma^2} \qquad (5)$$

Based on the dependencies (5) one can see that a measure of the expected value of the variable X (with the lognormal distribution) must take into account both the arithmetic mean and the variance (calculated based on logarithmically transformed values of the random variable X). Logarithmic transformation of H^+ activity with use of normal logarithm of X can be written as follows:

$$\log X = \frac{\ln X}{\ln 10}$$

Furthermore, log X is normally distributed with the expected value

$\frac{\mu}{\ln 10}$ and variance $\frac{\sigma^2}{\ln^2 10}$ if ln X is normally distributed. Therefore, the arithmetic mean calculated based on the value of the random sample for the H^+ activity, in the case when the probability distribution of H^+ is lognormal distribution, is not a good measure of the central tendency of such a variable.

RESULTS AND DISCUSSION OF SELECTED LITERATURE DATA ANALYSES

Theoretical considerations are illustrated with examples of central tendency measure calculations for H^+ activity and *pH*. The first example concerns the *pH* of lake waters determined in sixty-six European shallow lakes (Moss et al. 2003), hereinafter referred to as example lake waters. The other concerns the pH_{KCl} of humus horizon in sixty-nine post-agricultural soils afforested with Scots pine, further referred to as an example of the soil *pH*.

The source of soil pH data is a doctor's thesis by Olszewska (2006) and Pietruczyk (2010), partly published (Smal and Olszewska 2008). Lake waters are characterized by a wider range of *pH* values than soils hence the presentation of such examples allowing

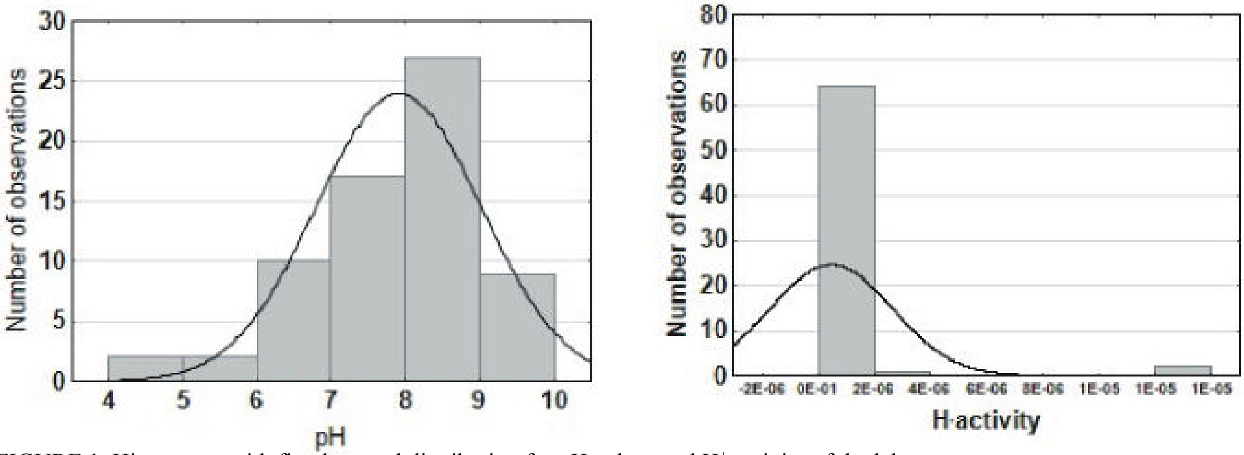

FIGURE 1. Histograms with fitted normal distribution for pH values and H^+ activity of the lake waters

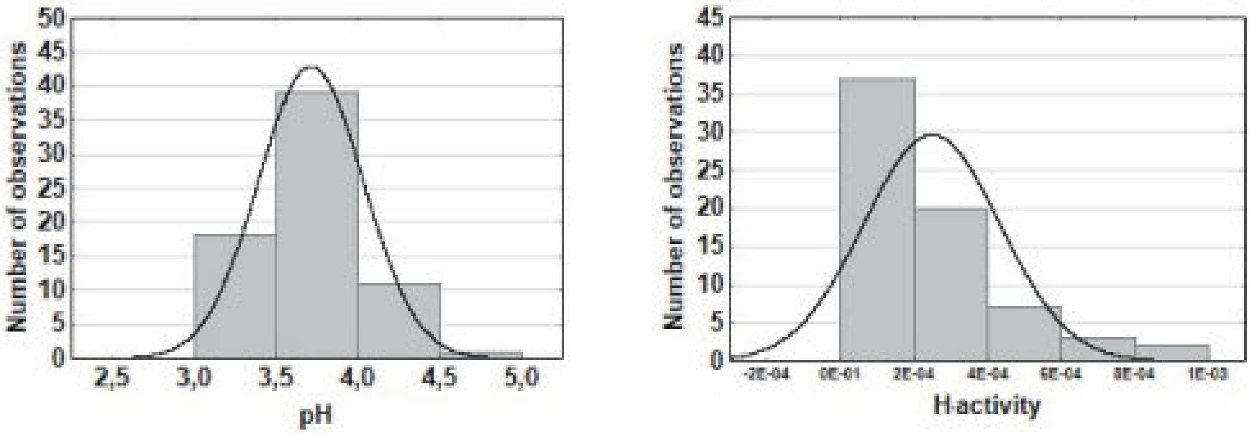

FIGURE 2. Histograms with fitted normal distribution for *pH* and H^+ activity of the afforested soils

fuller analysis of *pH* and H^+ activity distribution. In both cases, H^+ activity was calculated according to the formula: $H^+ (mol \cdot dm^{-3}) = 10^{-pH}$.

Using histograms with fitted normal distribution, the distribution properties' characteristics were presented for the studied features. Moreover, the values of basic descriptive statistics for *pH* and for the H^+ activity were calculated and discussed. Later, the arithmetic mean was compared with the geometric mean in the context of central tendency measures for the set of results.

The histogram with fitted normal distribution (Fig.1) for *pH* of lake waters and for the afforested soils (Fig. 2) shows a good consistency with the normal distribution. The histogram with fitted normal distribution for the H^+ activity of lake waters (Fig.1) and for the afforested soils (Fig. 2) indicates right skewed distribution. The distribution of *pH* values of lake waters is characterized by a small left skewness (-0.5424) and a small peaked κ =0.0733, while the distribution of the H^+ activity is a right skewed distribution of high skewness (7.555) and a very large peaked κ =59.21 (Fig. 1, Table 1). H^+ activity is rather inconsistent with normal distribution, but its logarithm *pH* is consistent with this distribution.

In probability theory, a lognormal distribution is a probability distribution of a random variable whose logarithm is normally distributed. If *pH* is a random variable with a normal distribution, then H^+ activity has a lognormal distribution; likewise, if H^+ activity

TABLE 1. Tests for normality distribution of *pH* and H^+ activity

Test	Afforested soils				Lake waters			
	pH		H^+ activity		*pH*		H^+ activity	
	test value	p-value	test value	p-value	test value	p-value	test value	p-value
Kolmogorov-Smirnov	0.0687	> 0.20	0.1996	< 0.01	0.1258	> 0,20	0.4421	< 0.01
Shapiro-Wilk	0.9548	0.1394	0.1996	<0.01	0.9608	0.0334	0.2254	0.0000

is lognormally distributed, then $pH = -\log(H^+)$ is normally distributed.

Multiplying or dividing lognormal random variables will result in lognormal distributions. According to these definitions and these tests results (Table 1) we can assume that H^+ activity of the soils and H^+ of the analysed lake waters are consistent with lognormal distribution. Kolmogorov-Smirnov and Shapiro-Wilk tests do not reject the hypotheses that the features of *pH* for the soil and *pH* for the lake waters are normally distributed (Table 1) (Svincov and Cambell 2002).

The distributions thus obtained lead to the use of the arithmetic mean for *pH* values, while the geometric mean is used for H^+ activity. The *pH* probability distribution of the soils is a leptokurtic (κ =2.608) and skewed to the left (0.620), while the distribution of H^+ activity is also leptokurtic (κ =2.86) but skewed to the right (1.705) (Table 2).

The arithmetic and the geometric mean of *pH* for the lake waters and for the soils, are at the same side of the median, where there is inequality:

$$\overline{pH}_g < \overline{pH}_a < Me$$

Thus, the arithmetic mean of the *pH* is closer to 'middle" than the geometric mean (Table 2). The relationship between these three measures for the H^+ activity is as follows:

$$Me < H_g^+ < H_a^+$$

so the geometric mean for the H^+ activity is closer to 'middle" than the arithmetic mean.

Figure 3 shows the location of the arithmetic mean and geometric mean for the relevant (via the logarithmic transformation) of lake water H^+ activity in the background data set. The geometric mean is closer to the median of H^+ activity, whereas the arithmetic mean lies farther from the median and is shifted to the right in the direction of the greatest values (sensitivity to extreme values) (Table 2). Out of the 66 measurements, only a few show values greater than the arithmetic mean, while the remaining ones show lower values. It is easy to see from the data set of H^+ activity that the geometric mean is a better measure of its central tendency than the arithmetic mean.

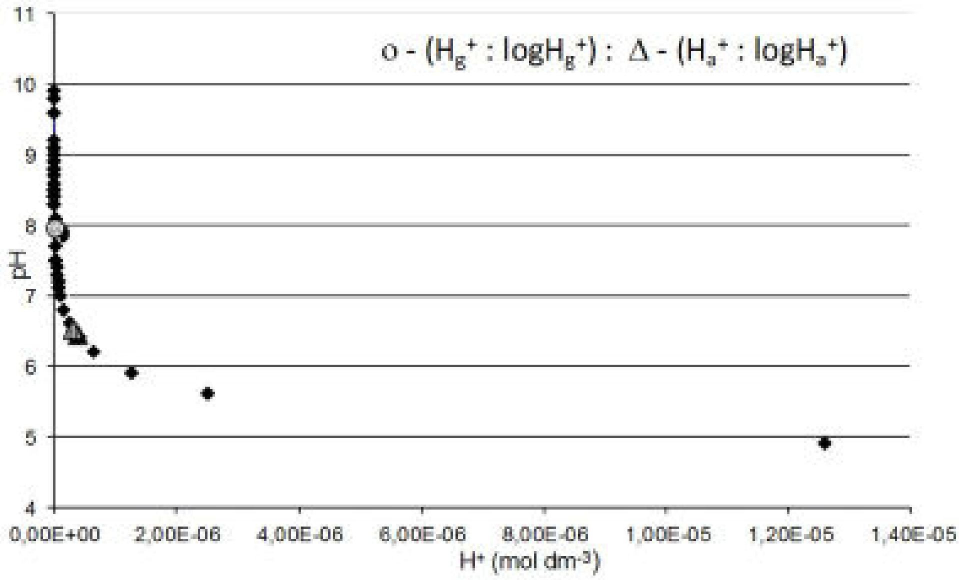

FIGURE 3. Scatter plot for *pH* and H^+ activity of the soils with marked geometric mean (H_g^+) and arithmetic mean (H_a^+) for H^+ activity

TABLE 2. Descriptive statistics for *pH* and H^+ activity of the lake waters and soils

Descriptive measure	Lake waters		Soils	
	pH	H^+ (mol ·dm^{-3})	*pH*	H^+ (mol ·dm^{-3})
Arithmetic mean pH_a, H^+_a	7.948	3.056E-07	3.712	2.48E-04
Geometric mean pH_g, H^+_a	7.874	1.126E-08	3.698	1.94E-04
Median *Me*	8.100	7.943E-09	3.720	1.91E-04
Mode *Mo*	8.400	0E+00	3.740	1.82E-04
Size of mode	7	7	4	4
Minimum	4.900	1.259E-10	3.050	1.02E-05
Maximum	9.900	1.259E-05	4.990	8.91E-04
Standard deviation *SD*	1.061	1.576E-06	0.321	1.86E-04
Skewness A	-0.543	7.555	0.620	1.705
Kurtosis κ	0.073	5.921E+01	2.608	2.86E+00

CONCLUSIONS

1. Before decision on the using the arithmetic mean as a measure of the central tendency for *pH* or for H^+ activity, the probability distribution of these variables must be determined. In addition, also skewness and kurtosis of distributions should be analysed.
2. At normality of the *pH* distribution and right-skewed of H^+ activity distribution, one should use the arithmetic mean for *pH* and the geometric mean for H^+ activity as a measure of the central tendencies of *pH* and H^+ activity.
3. For the skewed or the wide-scale range probability distributions of the *pH*, the median should be used as a measure of the central tendency for the *pH* and for the H^+ activity.
4. The median can be used as a measure of the central tendency both, for *pH* and for H^+ activity, irrespective of their probability distributions. It should be mentioned that the median does not use full information from the sample and cannot be the parameter verified in statistical parametric tests.
5. The measure of central tendency does not allow for complete *pH* or H^+ activity analysis therefore more descriptive statistics should be included in an analysis of this kind.

REFERENCES

Baker A.S., Kuo S., Chae Y.M., 1981. Comparisons of arithmetic mean soil pH values with the pH values of composite samples. Soil Science Society of America Journal 45: 828–830.

Barth E.F., 1975. Average pH. Journal-Water Pollution Control Federation 47: 2191–2192.

Bochenek W., 2014. Sezonowa zmienność i wieloletnie tendencje pH w opadzie atmosferycznym na Stacji Bazowej ZMŚP w Szymbarku w latach 1999–2013 (Seasonal variability and longterm trends of pH in precipitation on the Base Station of IMNE in Szymbark in the years 1999–2013). Monitoring Środowiska Przyrodniczego 16: 41–47.

Boutilier R.G., Shelton G., 1980. The statistical treatment of hydrogen ion concentration and pH. Journal of Experimental Biology 84: 335–339.

Boyle J.F., 1991. Acidity and organic carbon in lake water: variability and estimation of means. Journal of Paleolimnology 6: 95–101.

Dangles O.J., Guérold F.A., 2000. Structural and functional responses of benthic macroinvertebrates to acid precipitation in two forested headwater streams (Vosges Mauntains, northeastern France). Hydrobiologia 418: 25–31.

Forsberg L.S., Gustafsson J-P., Kleja D.B., Ledin S., 2008. Leaching of metals from oxidizing sulphide mine tailings with and without sewage sludge application, Water Air and Soil Pollution 194: 331–341.

Gaddum J.H., 1945. Lognormal distributions. Nature 156: 463–466.

Giesecke A.H., 1979. *In reply*. Anesthesiology 51: 482–483.

Gruba P., Błońska E., Socha J., 2010. Metodyczne aspekty pomiaru i statystycznej analizy wartości *pH* gleb (Methodological aspects of the measurement and statistical analysis of the soil *pH* values. Roczniki Gleboznawcze – Soil Science Annual 1: 29–37.

Kostrzewski A., Kruszyk R., Kolender R., 2006. Zintegrowany Monitoring Środowiska Przyrodniczego. Zasady organizacji, system pomiarowy, wybrane metody badań (The Integrated Monitoring of the Natural Environment. Principles of organization, measurement system, chosen methods), http://www.gios.gov.pl/zmsp/wyt2006/22_aneks_1.pdf

Krishnamoorthy K., Mathew T., 2003. Inferences on the means of lognormal distributions using generalized p-values and generalized intervals. Journal of Statistical Planning and Inference 115: 103–121.

Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M., 2007. Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, cz. 2. (Probability calculus and mathematical statistics in exercises, part 2). Wydawnictwo Naukowe PWN-SA, Warszawa: 330 pp.

Larsen B.M., Sanlund O.T., Berger H.M., Hesthagen T., 2007. Invasives, introductions and acidification: The dynamics of a stressed river fish community, Water Air and Soil Pollution Focus 7: 285–291.

Middleton A.C., Rovers F.A., 1976. Average pH (discussion). Journal-Water Pollution Control Federation 48: 395–396.

Moss B., Stephen D., Alvarez C. et al., 2003. The determination of ecological status in shallow lakes – a tested system (ECOFRAME) for implementation of the European Water Framework Directive. Aquatic Conservation Aquatic Conservation: Marine and Freshwater Ecosystems 13: 507–549.

Murphy M.R., 1981. Analyzing and presenting pH data. Journal of Dairy Science 65: 161–163.

Nőges P., Nőges T., Tuvikene L. et al., 2003. Factors controlling hydrochemical and trophic state variables in 86 shallow lakes in Europe. Hydrobiologia 506–509: 51–58.

Olszewska M., 2006. Wpływ zalesienia gleb lekkich porolnych sosną na zmiany ich właściwości. Praca doktorska (The effect of afforestation with Scots pine of light textured soils on changes in their properties, PhD Thesis), University of Life Sciences in Lublin: 200 pp.

Pace N.L, Ohmura A., Mashimo T., 1979. Averaging pH vs. H^+ values. Anesthesiology 51: 481–482.

Pietruczyk D., 2010. Właściwości chemiczne i fizykochemiczne gleb porolnych zalesionych sosną zwyczajną (*Pinus sylvestris* L.), Praca doktorska (Chemical and physicochemical properties of post-agricultural soils afforested with Scots pine (*Pinus sylvestris* L.), PhD Thesis), University of Life Sciences in Lublin: 260 pp.

Simon L.M., Cherry D.S., Curie R.J., Zipser C.E., 2006. The ecotoxicological recovery of fly creek and tributaries (Lee County, Va) after remediation of acid mine drainage. Environment Monitoring and Assessment 123: 109–124.

Smal H., Olszewska M., 2008. The effect of afforestation with Scots pine (*Pinus silvestris* L.) of sandy post-arable soils on their selected properties. II. Reaction, carbon, nitrogen and phosphorus, Plant and Soil 305: 171–187.

Svincov T., Cambell M., 2002. Statistics at square one. BMJ Publishing Group.

USGS-United States Geological Survey, 2008. National Field Manual for the Collection of Water-Quality Data. Techniques of water-resources Investigations, Book 9 Handbooks for Water-Resources Investigations, http://water.usgs.gov/owq/FieldManual/Chapter6/6.4_ver2.0.pdf

Wesselink L.G., Van Breemen N., Mulder J., Janssen P.H., 1996. A simple model of soil organic matter complexation to predict the solubility of aluminium in acid forest soils. European Journal of Soil Science 47: 373–384.

Yang J., Sun J., Hammer D., 2004. Distribution normality of pH and H^+ activity in soil. Environmental Chemistry Letters 2: 159–162.

Received: October 2, 2017
Accepted: November 30, 2017
Associated editor: B. Rutkowska

Miary statystyczne centralnej tendencji aktywności jonów H^+ i *pH*

Streszczenie: Mimo wielu prac dotyczących analizy statystycznej *pH* poglądy na użycie średniej arytmetycznej jako miary centralnej tendencji *pH* i aktywności H^+ są niejednolite. Problem powstaje, ponieważ przekształcenie średniej arytmetycznej dla jednej z tych cech nie daje średniej arytmetycznej dla drugiej. W pracy przedstawiono: 1) teoretyczne rozważania dotyczące rozkładów prawdopodobieństwa *pH* i aktywności H^+ oraz zależności między nimi, właściwości tych rozkładów, dopasowywania rozkładów empirycznych *pH* oraz aktywności H^+ do rozkładu teoretycznego, a także wyboru jednej z cech *pH* lub H^+, dla której będzie obliczana średnia arytmetyczna, 2) przykłady obliczeń miar centralnej tendencji *pH* i aktywności H^+ na podstawie danych literaturowych dla gleb i wód jeziornych. Analiza danych obejmowała sprawdzenie zgodności rozkładu empirycznego *pH* i aktywności H^+ z rozkładem normalnym, właściwości rozkładu, statystyki opisowe *pH* i aktywności H^+ oraz porównanie średniej arytmetycznej ze średnią geometryczną. Wyniki pozwalają stwierdzić, że jednolite podejście do wyboru miary centralnej tendencji dla *pH* i dla aktywności H^+ wymaga określenia rodzaju miary dla jednej z nich, a następnie konsekwentnego przekształcenia tej miary. Decyzja o zastosowaniu średniej arytmetycznej jako miary centralnej tendencji dla jednej z cech *pH* lub aktywności H^+ powinna być poprzedzona badaniem zgodności empirycznych rozkładów tych zmiennych z rozkładem normalnym. Normalny rozkład prawdopodobieństwa *pH*, a stąd logarytmiczno-normalny rozkład aktywności H^+ wskazuje, że średnia arytmetyczna i korespondująca z nią średnia geometryczna są właściwymi miarami klasycznymi tendencji centralnej *pH* i aktywności H^+ odpowiednio. Ponadto, mediana jako statystyka pozycyjna może być użyta dla każdej z tych zmiennych, niezależnie od ich rozkładu prawdopodobieństwa.

Słowa kluczowe: aktywność H^+, *pH*, średnia arytmetyczna, średnia geometryczna, mediana, rozkład normalny, rozkład logarytmiczno-normalny