

The handling of numerical ages and their random uncertainties

MEBUS A. GEYH^{*}

Abstract: The correct handling of numerical ages and their standard deviations and a proper introduction to error propagation or propagation of uncertainty and statistical evaluation are important to avoid misleading chronological conclusions and statements even though based on properly determined and reliable numerical dates. The conclusions may also be erroneous if dates were taken from databases without sufficient background information on the origin of the dated material and the applied analytical techniques. This paper is an introduction into the field of mathematical handling and testing of numerical ages. The most common and simple calculations and statistical tests that are needed are described and the steps involved are demonstrated on examples. The problems involved in the visualization of numerical dates in the form of normal histograms and dispersion histograms are discussed.

[Der Gebrauch numerischer Alter und ihre Standardabweichungen]

Kurzfassung: Die korrekte Auswertung numerischer Alter mit ihren Plus-Minus-„Fehlerangaben“ oder besser Standardabweichungen und die Anwendung der damit verbundenen statistischen Auswerteverfahren sind für Geowissenschaftler und Geographen oft eine schwierige Aufgabe, weil ihnen dafür die Ausbildung fehlt. Entsprechend finden sich in der Literatur viele nur schwer nachvollziehbare geochronologische Folgerungen. Sie können allerdings auch dann falsch sein, wenn sie sich auf Ergebnisse aus Datenbanken ohne ausreichende Angaben über den Ursprung des datierten Materials und die angewandten analytischen Techniken beziehen. Diese Arbeit führt in die Beurteilung und Handhabung numerischer Altersangaben ein. Mit ihr können elementare Rechnungen und einfache statistische Tests korrekt ausgeführt werden. Entsprechende Beispiele veranschaulichen den Weg dazu. Die mit der Visualisierung von numerischen Altern in Form von einfachen und Dispersions-Histogrammen verbundenen Probleme werden diskutiert.

Keywords: numerical ages, propagation of uncertainty, histogram

1 Introduction

Owing to the growing number of databases open to the public and the increasing production of numerical ages using different dating methods and semi-automated equipment. Quaternary scientists are thus faced with the task of handling large sets of numerical ages with their random uncertainties rather than dealing with single numerical ages. Reliability checks based only on the values themselves are indis-

pensable as in many cases information on the dated material and the analytical techniques is insufficient or not accessible.

Unqualified handling of numerical ages may result in conspicuously erroneous scientific conclusions. An example is a statistically evaluated cycle in the formation of speleothem and calcareous tufa in Europe which appeared to correlate with the 25-ka cycle of the precession of the Earth's axis (BRUNNACKER & HAUSMANN 1987; GEYH 1991). Reliable calculation of error

^{*}Address of author: M. A. Geyh, Rübeland 12, 29308 Winsen/Aller, Germany,
E-Mail: Mebus.geyh@t-online.de

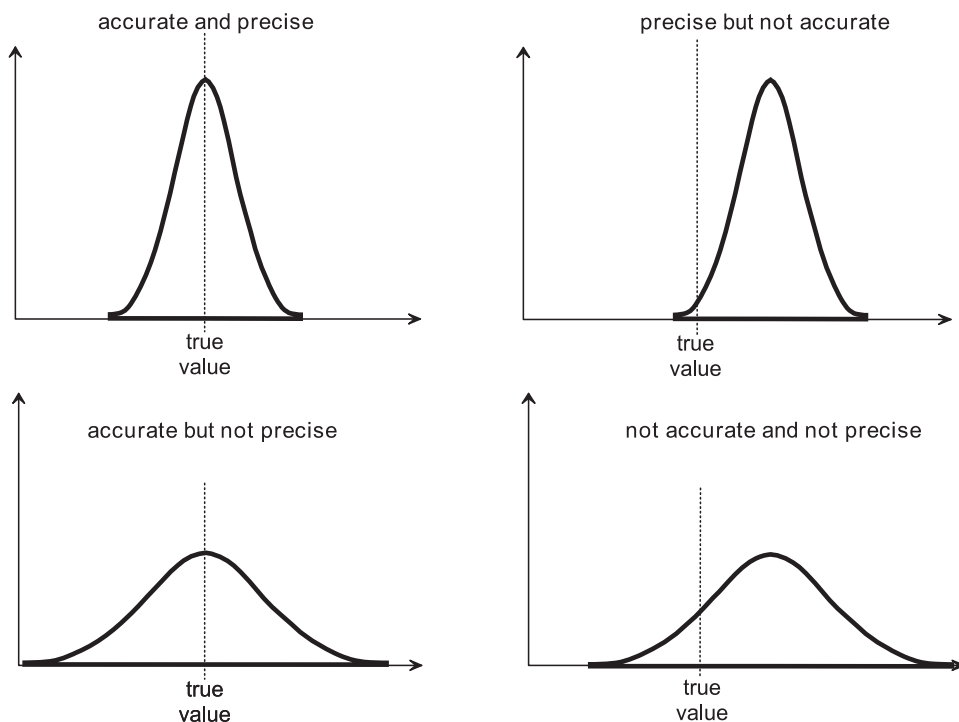


Fig. 1. The meanings of accuracy and precision: Accuracy describes the approximation of a mean to the „true“ value; precision is a measure of the reproducibility of a result.

Abb. 1. Bedeutung von Genauigkeit und Präzision: Genauigkeit beschreibt die Annäherung an den „wahren“ Wert. Die Präzision ist ein Maß für die Reproduzierbarkeit eines Ergebnisses.

propagation avoids such conclusions and also allows manipulated numerical ages to be identified (GEYH 1991). This paper provides the basics necessary for proper mathematical treatment of geochronological and analytical data of any origin with their random uncertainties. Complicated mathematical functions are not discussed. This paper uses the terminology of HUNTLEY (2001) in which „age“ refers to a range of time something has existed, while a „date“ is a particular time on a time scale. The functions given below are valid for any data, not only numerical ages.

2 Precision, accuracy and reporting of data

Precision and accuracy have distinct meanings as illustrated in Figure 1.

1) The *precision* of a result is a measure of the reproducibility of an observation and

corresponds to the *uncertainty* of a result. It is expressed as random error or *random uncertainty* and is mainly determined by the analytical technique and equipment used.

2) The *accuracy* is a measure of the correctness of an observed value and its approximation to the „true“ value. It depends on a large number of often unknown systematic errors, which may be eliminated by carrying out routine measurements on standard materials. Aliquots of the same set of samples of different standard materials are analyzed by laboratories throughout the world and the results compared. Reducing systematic errors increases the accuracy rather than the precision.

The expression $\pm \sigma A$ is termed *random* or *statistical uncertainty* and has to be distinguished from *systematic errors*. A *systematic error* arises from a specific physical or chemical

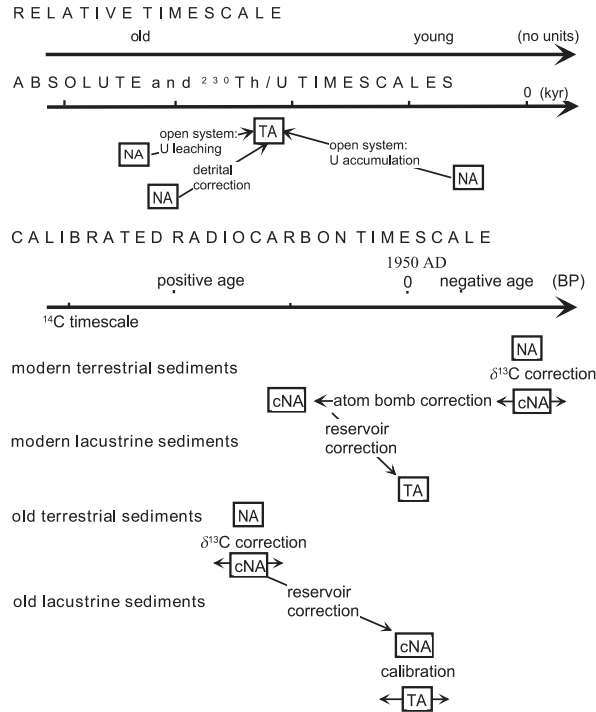


Fig. 2. Scheme for the correction and calibration of radiometric and mass spectrometric ^{14}C and $^{230}\text{Th}/\text{U}$ ages, in order to correlate them with the absolute time scale (GEYH & GROSJEAN 2000). TA – true age; NA – numerical age, cNA – corrected numerical age

Abb. 2. Schematische Darstellung der Korrektur und Kalibration radiometrischer und massenspektrometrischer ^{14}C - und $^{230}\text{Th}/\text{U}$ -Alter zur Korrelation mit der absoluten Zeitskala (GEYH & GROSJEAN 2000). TA – wahres Alter; NA – numerisches Alter, cNA – korrigiertes numerisches Alter

process, only changing the measured values in one direction or the other. If the process is understood and the relevant parameters are known, the result can be corrected. As an example, the relevant processes that affect ^{14}C and $^{230}\text{Th}/\text{U}$ ages are shown in Figure 2. Contamination with material older than the sample to be dated (e.g. with coal, fossil lime or detrital thorium) increases ^{14}C and $^{230}\text{Th}/\text{U}$ ages, respectively. Five percent matter containing no radiocarbon increases the ^{14}C ages by around 400 years. The same amount of uranium of a material in which the ^{238}U series is in radioactive equilibrium will increase the $^{230}\text{Th}/\text{U}$ ages by around 5000 years. A correction of these ages is possible if the proportion of the contaminating material is known. The conventional rules for reporting of quantitative data are often neglected. One rule is that

any quantitative result should be accompanied by a statement of the intrinsic uncertainty. The *precision of the random uncertainty* (the measure of its uncertainty) is usually not better than $\pm 10\%$. An example illustrates this rule. A random uncertainty of ± 213 has an uncertainty of about ± 21 a. Hence, it is reasonable to report the uncertainty as ± 210 a rather than ± 213 a. This means that the uncertainty should be rounded to two digits only (TAYLOR & KUYATT 1994). Another rule is that a set of values should always have the same number of places behind the decimal point. Hence, an age of 2.87 ± 0.20 ka should not be given as 2.87 ± 0.2 ka. This rule is especially important for the compilation of data in a table. Another rule is that only the final result should be rounded off rather than the values used in a calculation.

3 Numerical age, random uncertainty and frequency distribution

Any geochronological framework has to be based on numerical ages determined with suitable physical dating methods (GEYH & SCHLEICHER 1990; WAGNER 1998; GEYH 2005). The resulting ages are often inappropriately termed „absolute“ ages. If an age date is obtained, for example, from stratigraphic or palynological studies it is a relative age. Physical age dating also does not yield absolute ages; the results are better called „numerical ages“. The term absolute age should be reserved only for age dates related to the solar time scale.

The collected samples have to fulfill numerous criteria to be suitable for reliable dating. Different kinds of numerical ages have to be corrected for different sources of systematic error (e.g., ^{14}C ages for contamination, reservoir effect, and ^{13}C isotope fractionation; and $^{230}\text{Th}/\text{U}$ ages for detrital ^{230}Th ; e.g., GEYH & GROSJEAN 2000; Fig. 2)

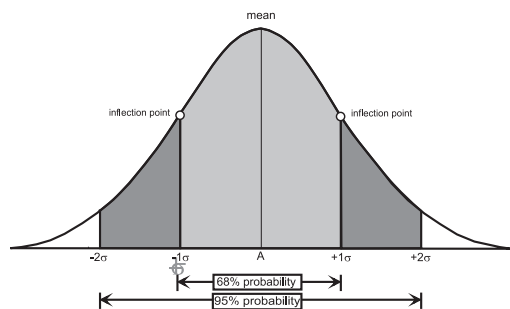


Fig. 3. Gaussian frequency distribution (also bell-shaped curve, normal distribution): The 1-sigma confidence interval lies between the two inflection points of the frequency distribution. The light gray area corresponds to nearly 68 % of the total area, the 2-sigma confidence interval represents a probability of about 95 %.

Abb. 3. Gauß'sche Häufigkeitsverteilung (auch Glockenkurve, Normalverteilung): Das 1-Sigma Vertrauensintervall liegt zwischen den beiden Wendepunkten der Glockenkurve. Die graue Fläche entspricht nahezu 68 % der Gesamtfläche, das 2-Sigma Vertrauensintervall repräsentiert eine Wahrscheinlichkeit von 95 %.

and in the case of dating with cosmogenic isotopes calibration is needed (e.g., ^{14}C dates). This is particularly important when numerical ages obtained from different materials and methods are compared with each other and/or with the absolute time scale (GEYH & SCHLEICHER 1990; WAGNER 1998; GEYH 2005).

When dealing with numerical ages, it is very important to understand the meaning of the common expression $A \pm \sigma A$, e.g., 2000 \pm 50 a before present (BP). *Standard deviation*, 1-sigma interval, random uncertainty, and statistical uncertainty are used as synonyms for $\pm \sigma A$. It defines the precision of the calculated value, and represents a frequency distribution. Theoretically it is assumed that the age of the sample was determined an *infinite* number of times. In practice, however, the standard deviation is precisely determined on the basis of a *finite* number of measurements. The ages determined on the same sample deviate from each other, even under most exacting measurement conditions, and scatter around a mean value \hat{A} , which is assumed to approximate the „true“ age of the sample. The frequency distribution of values is obtained by plotting all these values with the x-axis representing „age“ and the y-axis representing the frequency of occurrence of each value. The resulting curve looks like a vertical cut through a bell and is known as a *Gaussian distribution*, *normal distribution* or *bell-shaped curve* (Fig. 3). It is symmetrical around the *mean* value of all values.

In reality only a finite number of results are available and the frequency distribution is not perfectly symmetrical. In this case the curve peak is given by the median rather than the mean. It divides a frequency distribution into two equal areas. This case will, however, not be discussed here as it is of minor importance in geochronological studies due to the relatively small number of ages obtained from coeval samples.

The area enclosed by the bell-shaped curve, extended to infinity in both directions, represents the total number of values or 100 % probability. The area between the two inflection points, shown in light grey, contains nearly 68 % of all values (68 % probability) (*1-sigma interval* or

1σ confidence interval). The 2-sigma confidence interval contains approximately 95 % of all values (95 % probability), corresponding to a probability of error of 5 %. Hence, one and two sigma means that 2 out of 3 or 19 out of 20 results fit the 1σ and 2σ confidence intervals, respectively, provided a sufficiently large number of ages from coeval samples are available. The 1σ confidence interval for the age $A = 2000 \pm 50$ BP is 1950 BP to 2050 BP; one third of the values should be greater or smaller than these two values ($2050 > A > 1950$ BP).

The convention in science and engineering is to give the 1-sigma confidence interval for the uncertainty of a value. An exception is mass spectrometric ages in geochronology, for which the uncertainty is often given as the 2-sigma confidence interval.

The decision whether the 1-sigma or 2-sigma confidence interval is used depends on the question to be answered. An increase of the width of the confidence interval enlarges the probability that the true age is included, but it decreases the ability to distinguish two ages or to decide whether they belong to the same or different frequency distributions. Thus, it is indispensable to state in each publication whether the given uncertainty is 1σ or 2σ .

An example demonstrates this case. Two ages of 1850 ± 50 a and 2000 ± 50 a do not appear to belong to coeval samples due to the gap of 150 a between the 1σ confidence limits. In contrast, samples with the same ages but twice as large standard deviations – 1850 ± 100 a and 2000 ± 100 a – might have been considered as coeval as their confidence intervals overlap. The next chapter describes how such cases are correctly handled.

4 Arithmetic and weighted means

The *arithmetic mean* age \hat{A} (normally simply called the mean) of a set of n numerical ages (A_1 to A_n) with the same precision is given by the following equation:

$$(1) \quad \bar{A} = \frac{\sum_{i=1}^n A_i}{n}.$$

The *standard deviation* σA of the age values A_i in a set of ages A is calculated as follows:

$$(2) \quad \sigma A = \pm \sqrt{\frac{\sum_{i=1}^n (\bar{A} - A_i)^2}{(n-1)}}.$$

Using the three ages 2000 ± 50 a, 2050 ± 50 a and 2200 ± 100 a as an example, Equations 1 and 2 yield (for which the standard deviations of the individual ages are not used) a mean and standard deviation of 2083 ± 104 a. As explained above, this has to be reported as 2080 ± 100 a. (Contrary to the conventional rule for reporting values given in Chapter 2, the values and their uncertainties in this paper are given as calculated and not as they should be reported. This enables the reader to follow the calculations with the corresponding equations).

The larger the number of ages determined for the same sample, the higher the probability that the calculated mean approximates the „true“ age. This is reflected in the standard deviation of the mean ($\sigma \hat{A}$), which decreases with \sqrt{n} as follows:

$$(3) \quad \sigma \bar{A} = \frac{\sigma A}{\sqrt{n}} = \pm \sqrt{\frac{\sum_{i=1}^n (\bar{A} - A_i)^2}{n \cdot (n-1)}}.$$

Our example yields ± 60 a, while that of a single measurement is ± 104 (Eq. 2).

Numerical age values usually have differing precision, requiring the introduction of a *weighting factor* w . It is calculated from the standard deviation as follows:

$$(4) \quad w_i = \frac{1}{\sigma A_i^2}.$$

The weighting factors of our example given above are 0.0004, 0.0004 and 0.0001, respectively. This means the weighting factor for the two ^{14}C ages with a precision of ± 50 years is four times larger than the weighting factor of the ^{14}C age with a precision of ± 100 years. This is not only a mathematical artifact. The third radiometric ^{14}C measurement required four times the time needed for the first two.

x	y	sy	y-sy	y+sy
0,00	3,0	0,50	2,50	3,50
0,05	3,1	0,48	2,62	3,58
0,10	3,2	0,45	2,75	3,65
0,15	3,3	0,43	2,87	3,73
0,20	3,4	0,40	3,00	3,80
0,25	3,5	0,38	3,12	3,88
0,30	3,6	0,36	3,24	3,96
0,35	3,7	0,33	3,37	4,03
0,40	3,8	0,31	3,49	4,11
0,45	3,9	0,29	3,61	4,19
0,50	4,0	0,27	3,73	4,27
0,55	4,1	0,25	3,85	4,35
0,60	4,2	0,23	3,97	4,43
0,65	4,3	0,22	4,08	4,52
0,70	4,4	0,21	4,19	4,61
0,75	4,5	0,20	4,30	4,70
0,80	4,6	0,19	4,41	4,79
0,85	4,7	0,19	4,51	4,89
0,90	4,8	0,19	4,61	4,99
0,95	4,9	0,19	4,71	5,09
1,00	5,0	0,20	4,80	5,20

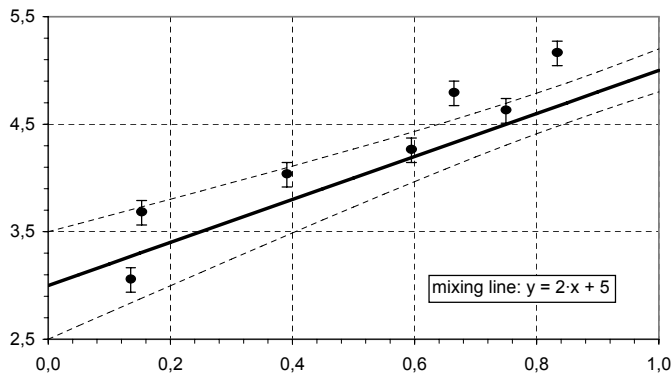


Fig. 4: Linear regression of values to obtain a binary mixing line.

Abb. 4: Lineare Regression von Werten, die einer Zweikomponenten-Mischgerade entsprechen.

The *weighted mean* \hat{A}_w is obtained using the following equation:

$$(5) \quad \bar{A}_w = \frac{\sum_{i=1}^n A_i \cdot w_i}{\sum_{i=1}^n w_i} = \frac{\sum_{i=1}^n \frac{A_i}{\sigma A_i^2}}{\sum_{i=1}^n \frac{1}{\sigma A_i^2}}$$

Equation 5 reduces to Equation 1 if the measurements were done with the same precision: $w_i = \text{constant}$.

The *standard deviation of the weighted mean* $\sigma \hat{A}_w$ is obtained by

$$(6) \quad \sigma \bar{A}_w = \pm \sqrt{\frac{1}{\sum_{i=1}^n w_i}}$$

The weighted mean of our example yields 2044 ± 33 a (to be reported as 2040 ± 30 a), deviating by 35 years from the arithmetic mean of 2083 ± 104 a (to be reported as 2080 ± 100 a) given above. The weighted mean and its standard deviation are shifted towards the more precise results.

Analogous to Equation 3 the *weighted mean standard deviation of a single value* is given by

$$(7) \quad \sigma A_w = \sqrt{n} \cdot \sigma \bar{A}_w.$$

The example yields ± 57 years.

According to Equations 3 and 6 the precision of the mean values $\sigma \hat{A}$ and $\sigma \hat{A}_w$ theoretically approaches zero for an infinite number of ages. In reality the precision has an (often unknown) upper limit resulting from the technical limitations of the method: for example, varying contamination of coeval samples that cannot be removed and slightly deviating measurement conditions (temperature, differing calibration of the equipment in different laboratories, impure chemical reagents, etc.). Ages above the upper limit of the method are given, for example, as $>40,000$ a. They express the minimum age of a sample or the maximum age which can be determined by the method.

5 Chi square test

Calculation of the arithmetic or weighted means \hat{A} or \hat{A}_w of a set of n ages $A_i \pm \sigma A_i$ makes sense only if the σA_i values were determined for coeval samples and therefore belong to the same frequency distribution. To test for this, there are two alternative but equivalent mathematical approaches:

a) The uncertainty $\pm \sigma A$ of the ages is calculated using Equation 2 and compared with the standard deviation σA_w of a single measure-

ment (Eq. 7). The *uncertainty of ages of different precision* is obtained (in analogy to Eq. 2) as follows:

$$(8) \quad \sigma A_g = \pm \sqrt{\frac{\sum_{i=1}^n (\bar{A}_w - A_i)^2 \cdot w_i}{(1 - 1/n) \cdot \sum_{i=1}^n w_i}}$$

It is permissible to calculate the mean \hat{A} only if σA (Eq. 2) or σA_g (Eq. 8) is more or less equal to σA_w (Eq. 7). Our example yields ± 57 a, ± 56 a and ± 104 a, respectively. Hence, the first two standard deviations differ so much from σA_w that the samples cannot be considered to be coeval and the mean should not be calculated.

b) χ^2 test

χ^2 is obtained as follows:

$$(9) \quad \chi^2 = \sum_{i=1}^n (\bar{A} - A_i)^2 \cdot w_i$$

$$\chi_w^2 = \sum_{i=1}^n w_i \cdot A_i^2 - \frac{(\sum_{i=1}^n w_i \cdot A_i)^2}{\sum_{i=1}^n w_i}$$

A mean can be calculated only if χ^2 and $\chi_w^2 \approx n-1$ for $n < 15$. For $n > 15$ the χ^2 value can be taken from any textbook on statistics (e.g., SACHS 1999). Our example yields $\chi^2 = 4.4$ and $\chi_w^2 = 3.2$, confirming the three ages do not belong to the same frequency distribution and thus a mean value would have no validity.

6 Error propagation: difference, rates, mathematical function

There are tasks for which more or less complicated mathematical functions are necessary, e.g., $y = f(x_1, \dots)$. Such functions have a dependent variable y and one or more independent variables $x_1, \dots, x_1, \dots, x_n$. The most simple case is calculation of the difference between two ages which equals the duration of a process. More complicated is, for example, calculation of the sedimentation rate for a certain time interval (Eq. 11). If the uncertainties for the independent variables σx_i are known, the standard deviation of the dependent variable σy can be calculated using Equation 10.

The standard deviation σf of the function $f(x_1,$

$x_2, \dots, x_n)$ is calculated by the law of propagation of uncertainty:

$$(10) \quad \sigma f(x_1, \dots, x_n)^2 = \left(\frac{\partial f}{\partial x_1} \right)^2 \cdot \sigma x_1^2 + \dots + \left(\frac{\partial f}{\partial x_n} \right)^2 \cdot \sigma x_n^2$$

When geological sections are studied, deposition periods and sedimentation rates r between two levels a and b (e.g., at a depth of 100 and 150 cm and ages of 2000 ± 50 and 2200 ± 100 years) are of interest. The function is given by:

$$(11) \quad r = \frac{b - a}{A_2 - A_1}$$

To solve this equation, the difference S between two ages ($A_2 - A_1 = 2200 \pm 100 - 2000 \pm 50$ years) has to be calculated. Neglecting the standard deviation, this example yields a difference of 200 years and a sedimentation rate of 0.25 cm/a. The standard deviation of a *sum* $S = x_1 + x_2$ or a *difference* $\Delta = x_2 - x_1$ is obtained from Equation 12:

$$(12) \quad \sigma S^2 = \sigma \Delta^2 = (\sigma x_1)^2 + (\sigma x_2)^2$$

Our example yields an age difference of 200 ± 112 years. The 1σ and 2σ confidence intervals are 88 to 312 and -24 to +424 years, respectively. The sample ages differ by at least 88 a with a 68 % probability (1σ), but they may be considered as similar with a probability of 95 % (2σ) because a negative age difference is meaningless.

Calculation of the deposition rate r involves a division and two subtraction steps. The corresponding formula involves both multiplication and division:

$$(13) \quad y = \frac{x_1 \cdot x_2}{x_3}$$

The standard deviation of y is obtained as follows:

$$(14) \quad \sigma y = \pm y \cdot \sqrt{\left(\frac{\sigma x_1}{x_1} \right)^2 + \left(\frac{\sigma x_2}{x_2} \right)^2 + \left(\frac{\sigma x_3}{x_3} \right)^2}$$

This corresponding formula for σr is derived from Equations 10 and 12. Assuming depths a and b were determined precisely (i.e., $\sigma a = \sigma b = 0$), the standard deviation of the deposition rate r is obtained by combining Equations 12 and 14 as follows:

$$(15) \quad \sigma r = \pm r \cdot \sqrt{\frac{\sigma A_1^2 + \sigma A_2^2}{(A_2 - A_1)^2}}.$$

According to Equation 15, the standard deviation σr decreases with increasing difference between the ages. The example yields 0.25 ± 0.14 cm/a. The result has 1σ and 2σ confidence intervals (i.e., 68 % and 95 % probabilities) of 0.11 to 0.39 cm/a or 0 to 0.53 cm/a, respectively. Hence, the actual sedimentation rate can be smaller or greater than 0.25 cm/a by a factor of two.

It is worthwhile to note that the formulas for calculating numerical ages usually contain one or more physical parameters that are internationally agreed upon – half lives are an example. Although they were determined with a finite uncertainty, these uncertainties are not usually considered when the law of propagation of uncertainties is applied. Indeed they are not required as long as only numerical ages are compared with each other. But when the ages relate to the absolute time scale, the standard deviation has to include the uncertainties of the internationally agreed parameters and is usually larger than the standard deviation normally reported.

7 Regression analysis and least squares fit

Regression analyses are applied for prediction, testing of hypothesis, and modeling of causal relationships. Such analyses are carried out to determine correlations in numerical data consisting of values of the dependent variable y and the independent variables x_i of a function with one or more parameters, e.g., a and b in Equation (16). These parameters are adjusted to give a best fit of the function to the numerical data. This is usually done using the least squares method. Either the independent variable or the dependent variable is assumed to be a random variable with the uncertainty of the

measurement. The other variable(s) is assumed to be error-free.

The simplest case is a linear function:

$$(16) \quad f(x) = y = b \cdot x + a.$$

Quadratic, exponential and other functions may be transformed into a linear function with a corresponding recalculation of the measured data. Both coefficients a and b are obtained by minimizing the sum of the squares of the deviation of the values y_i from a straight line (the uncertainty in x is assumed to be negligible). The equation is obtained by a least-squares fitting of a straight line to the data:

$$\begin{aligned} \sum \Delta y_i^2 &= \sum [y_i - f(x_i)]^2 = \\ \sum [y_i - a - b \cdot x_i] &= \min. \end{aligned}$$

The values of the parameters a and b are calculated as follows:

$$\begin{aligned} \frac{\partial}{\partial a} \sum w_i \cdot (y_i - a - b \cdot x_i)^2 &= 0 \quad \text{and} \\ \frac{\partial}{\partial b} \sum w_i \cdot (y_i - a - b \cdot x_i)^2 &= 0. \end{aligned}$$

Thus,

$$(17) \quad \begin{aligned} a &= \left(\sum x_i^2 \cdot w_i \sum y_i \cdot w_i - \right. \\ &\quad \left. \sum x_i \cdot w_i \cdot \sum x_i \cdot y_i \cdot w_i \right) / \Delta \\ b &= \left(\sum w_i \cdot \sum x_i \cdot y_i \cdot w_i - \right. \\ &\quad \left. \sum x_i \cdot w_i \cdot \sum y_i \cdot w_i \right) / \Delta \quad \text{where} \\ \Delta &= \sum w_i \cdot \sum x_i^2 \cdot w_i - \left(\sum x_i \cdot w_i \right)^2. \end{aligned}$$

The standard deviations of a and b are given by Equation 18 and shown in Fig. 4:

$$(18) \quad \begin{aligned} \sigma a &= \pm \sqrt{\frac{1}{\Delta} \sum x_i^2 \cdot w_i} \\ \sigma b &= \pm \sqrt{\frac{1}{\Delta} \sum w_i}. \end{aligned}$$

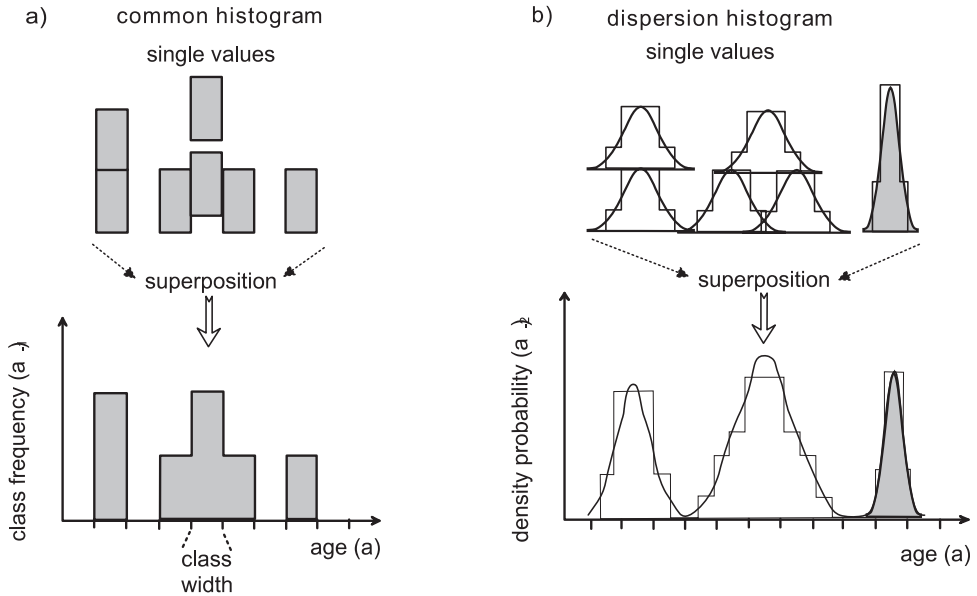


Fig. 5. a) Construction of a normal histogram of five ages based on rectangles, b) Construction of a dispersion histogram of five ages of low (white) and high (gray) precision based on polygons and bell curves, respectively.

Abb. 5. a) Konstruktion eines einfachen Histogramms mit Polygonen für fünf Alter, b) Konstruktion eines Dispersionshistogramms mit Polygonen und Glockenkurven für fünf ungenaue (weiß) und genaue (grau) Alter.

Many computer programs include a least-squares algorithm for fitting a straight line to the data. Often they can also handle composite curves consisting of the superposition of more than one function. A cubic spline can be used to fit a function to a data set. The parameter values of the function are individually chosen to obtain an optimum fit of a curve to all values or to smooth out irregularities in the curve.

8 Graphical presentation: normal and dispersion histograms

Visualization of the frequency distribution of data is often more instructive than a compilation of the data in tables. A *normal histogram* is the most simple way to present ages. The x-axis is assigned an age scale; the y-axis represents the frequency. The x-axis is divided into classes of an arbitrarily chosen width which depends on the age resolution or the average precision

of the available values. Each value is represented by a rectangle with a given area and a class width as base length. All rectangles are superimposed above the x-axis (Fig. 5a). The enveloping curve represents the frequency distribution.

Ages with a mean sigma of ± 50 years (Eq. 7) may be optimally resolved with a class width of 50 years as the base length of the rectangle. If a larger class width is chosen, e.g., 200 years, the optimum age resolution is lost. The advantage of a larger class width is that the number of values per class is increased, which reduces the random scatter of the frequency distribution per class. The uncertainty in a class population is approximated by the square root of the number of values. For instance, four values per class yield $\pm\sqrt{4} = \pm 2$ or $\pm 50\%$ scatter, for nine dates we obtain $\pm\sqrt{9} = \pm 3$ or $\pm 33\%$. As a consequence, the data density per

class governs the random fluctuation of the frequency distribution. If it is small, only minima rather than peaks provide reliable information. The random uncertainty of the histogram peaks may be larger than the actual fluctuations of the frequency of values per class. In contrast, lows always represent periods of missing values, which is why only they allow an unambiguous interpretation of a histogram. They may mirror, for example, arid periods without groundwater recharge, glacial periods without growth of trees, or periods of intensive erosion (see also Section 9.1).

A *dispersion histogram* (GEYH 1969; MICH-CZYNSKA & PAZDUR 2004) provides an alternative visualization of values. A polygon or, even better, a bell-shaped curve with an area representing one value replaces the rectangle of the normal histogram. The width of the area is related to the standard deviation. As the area is constant, its height is determined by the chosen width or standard deviation. Hence, a precise value yields a polygon or bell curve with a short baseline and large height, while a less precise value is represented by a wide baseline and low height (Fig. 5b top).

In a dispersion histogram, the precision of an age represented by its weighting factor w (Eq. 4) is taken into account. Hence, the y-axis represents a density probability function with the dimension of $1/(\text{time unit})^2$ rather than the class frequency with the dimension $1/(\text{time unit})$ (MICH-CZYNSKA & PAZDUR 2004).

There is a special problem with histograms of ^{14}C ages. The time scale is distorted with respect to the solar or absolute time scale. In certain time periods, ^{14}C years are systematically compressed with respect to the solar time scale, in others they are stretched out. This means in the case of a uniform frequency distribution of samples along the absolute time scale that the corresponding ^{14}C histogram may show apparent peaks and lows (GEYH 1980; Fig. 6). In contrast to what is expected, the distortion of the histogram does not disappear if calibrated age intervals are used instead of the ^{14}C ages. Modern calibration programs (e.g., Cal 125 Groningen Radiocarbon Calibration Program)

allow the automated construction of histograms with the calibrated ^{14}C time scale and overcomes the intrinsic problem of calibration of single values.

9 Two case studies

Two case studies are presented which may demonstrate the value of a correct statistical evaluation of ages. It is not the intention of this paper to become involved in the geological aspects of the cited papers or problems inherent in dating methods used. Here, it is mainly considered whether correct handling of the ages gives another numerical result than discussed in the papers.

In the first case study, evidence is provided that increasing the number of numerical age values may not increase the precision of the obtained values. In the second case study, it becomes ob-

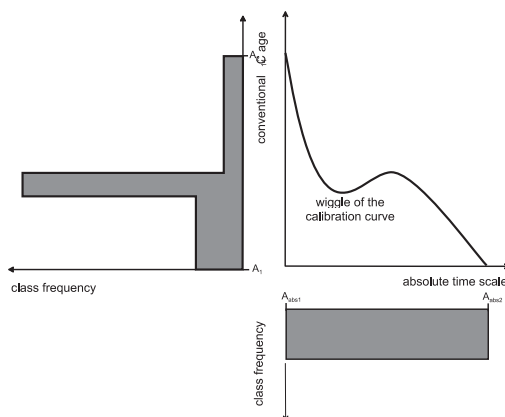


Fig. 6. The histogram between the absolute ages A_{abs1} and A_{abs2} is uniform whereas the histogram of the corresponding ^{14}C dates A_1 and A_2 is distorted. This is the consequence of the wiggles in the ^{14}C calibration curve. Peaks and lows in the ^{14}C histogram may be without actual relevance.

Abb. 6. Das Histogramm der gleich häufigen absoluten Alter A_{abs1} und A_{abs2} entspricht einem Rechteck und liefert für die entsprechenden ^{14}C -Alter ein Polygon. Die Ursache sind die Windungen der ^{14}C -Kalibrationskurve. Gipfel und Täler im ^{14}C -Histogramm zwischen den Altern A_1 und A_2 haben keine wirkliche Bedeutung.

Table 1: ^{14}C ages ordered in size from wood samples dating the beginning of the Younger Dryas in New Zealand (DENTON & HENDY 1994). The standard deviation of the mean (Eq. 7) and the calibrated ^{14}C age of the 2σ confidence interval are given in parentheses.

Tab. 1: Die nach ihrer Größe geordneten ^{14}C -Alter von Holzproben, die den Beginn der Jüngerer Dryas-Zeit in Neuseeland datieren sollen. Die Standardabweichung des mittleren Alters (Gleichung 7) und das 2-Sigma-Konfidenzintervall seines kalibrierten ^{14}C -Alters sind in Klammern angegeben.

^{14}C age (BP) Group I		^{14}C age (BP) Group II	
10,650 ± 100	11,040 ± 90	11,150 ± 160	11,255 ± 95
10,750 ± 100	11,045 ± 85	11,190 ± 60	11,290 ± 150
10,800 ± 90	11,080 ± 60	11,200 ± 50	11,340 ± 110
10,830 ± 110	11,090 ± 70	11,200 ± 120	11,350 ± 90
10,920 ± 90	11,110 ± 110	11,225 ± 60	11,350 ± 60
10,950 ± 100	11,110 ± 90	11,230 ± 130	11,365 ± 60
10,980 ± 90	11,110 ± 130	11,240 ± 100	11,370 ± 190
10,980 ± 100	11,115 ± 90	11,250 ± 70	11,520 ± 170
11,040 ± 70	11,150 ± 60	11,250 ± 50	11,520 ± 140
11,150 ± 14 BP (± 28 a) $\chi^2 = 148$			
13,020 – 13,103 BP (12,965 – 13,117 BP)			
11,011 ± 20 BP (± 32 a) $\chi^2 = 42$		11,265 ± 19 BP (± 13 a) $\chi^2 = 17$	
12,899 – 12,945 BP		13,125 – 13,193 BP	
(12,874 – 12,995 BP)		(13,101 – 13,223 BP)	
10,862 ± 34 BP (± 48 a)	11,097 ± 27 (± 9 a)		
$\chi^2 = 11$	$\chi^2 = 2$		
12,835 – 12,871 BP	12,960 – 13,052 BP		
(12,818 – 12,893 BP)	(12,930 – 13,086 BP)		

vious that the ages were not correctly evaluated in the laboratory and the published standard deviations are meaningless.

9.1 The beginning of the Younger Dryas in New Zealand

DENTON & HENDY (1994) tried to determine the beginning of the Younger Dryas in New Zealand, i.e., the age of a single event. They collected 25 pieces of wood from short-lived trees and covered by Younger Dryas moraines of the Franz-Josef Glacier in the Southern Alps of New Zealand. The holocellulose extract yielded 36 ^{14}C ages determined in three laboratories (Table 1). It was assumed that all samples were coeval and deposited at the beginning of the Younger Dryas.

The weighted mean and its standard deviation of the ^{14}C ages is $11,150 \pm 14$ BP (Eqs. 5 and 6) and the weighted mean standard deviation of a single ^{14}C age is ± 82 a (Eq. 7). DENTON & HENDY (1994) published a mean of $11,170 \pm 14$ BP. The difference between these age values results from the inclusion by the authors of a ^{14}C date which they considered to be older than the target event of the Younger Dryas. The calibrated ^{14}C age interval is $13,020$ – $13,103$ BP ($12,965$ – $13,117$ BP for the 2σ interval). The authors give the mean ^{14}C age as the date for the beginning of the Younger Dryas. The actual standard deviation of the weighted mean, however, is ± 28 a (Eq. 6) and the actual weighted mean standard deviation of the individual ages is ± 169 a (Eq. 8). This finding is not compatible with the assumption that the published ^{14}C

ages belong to a single frequency distribution and, therefore, to a single event. This is supported by the χ^2_w value of 148 for the 36 values.

In order to check whether the published ^{14}C ages represent more than one frequency distribution, the ages were ordered according to size and divided into two groups of 18 values each and the first group into two subgroups of nine values (Table 1). Only the weighted means of the second group and the first subgroup belong to a single frequency distribution that represents coeval samples. The other groups represent more than one frequency distribution. The too low χ^2 value for the second subgroup is an artifact of the ordering of the ^{14}C ages according to size.

The present consensus places the beginning of the Younger Dryas between 12,900 and 12,700 BP on the basis of calibrated ^{14}C ages (ALLEY 2000). The calibrated ^{14}C -age range of the weighted mean of the published ^{14}C ages of 12,965–13,117 BP (2σ confidence interval) considerably deviates from that. Only the first group and first subgroup with calibrated ^{14}C age ranges of 12,874–12,995 BP and 12,818–12,893 BP, respectively, agree with this consensus. In both cases several ^{14}C ages have to be discarded as the χ^2 values of 42 and 11 deviate from the target χ^2 values of 17 and 8, respectively.

For the purpose of this paper, it is not of interest to determine the events responsible for this. The wood samples may stem from trees in forests containing trees of different ages that were buried at different times by two or three separate glacier advances. Other causes are possible. Of importance is only that the assumption of the authors the dated samples were coeval is not valid. The precision for the beginning of the studied event was not narrowed to ± 14 years. The actual precision of the date for the beginning of this event is at least twice as large.

The reader may claim that the difference in precision between ± 14 and ± 26 years is small. However, one has to realize that as few as nine ^{14}C ages would have been sufficient to obtain a precision of ± 26 years. This means that up to 75 % of the dating costs could have been saved.

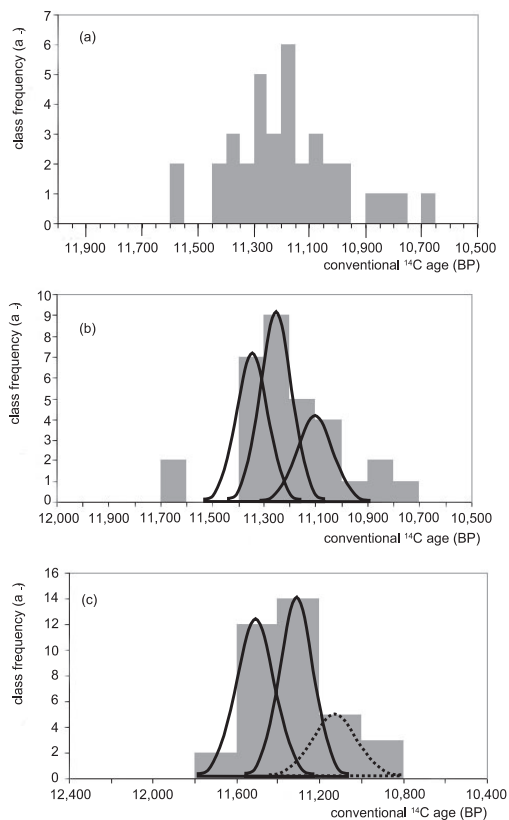


Fig. 7: Normal histogram of 36 ^{14}C ages from 25 wood samples which were assumed to represent the beginning of the Younger Dryas in southern New Zealand (DENTON & HENDY 1994). The arbitrarily chosen class widths of (a) 50, (b) 100 and (c) 200 years shows the influence on the temporal resolution and the scatter of the histogram. The optimum choice may be 100 years (b), which is close to the mean standard deviation of ± 88 years for single ages.

Abb. 7: Einfaches Histogramm von 36 ^{14}C -Altern, die von 25 Holzproben vom südlichen Neuseeland bestimmt worden sind. Sie sollen den Beginn der Jüngerer Dryas-Zeit festlegen (DENTON & HENDY 1994). Die willkürlich gewählten Klassenintervalle von (a) 50, (b) 100 und (c) 200 Jahren zeigen die Auswirkungen auf die altersmäßige Auflösung und auf die Unruhe des Verlaufs des Histogramms. Eine optimale Wahl scheint 100 Jahre zu sein (b), ein Wert, der nahe der mittleren Standardabweichung von Einzeldaten von ± 88 Jahren liegt.

Table 2: ESR ages (ka) of interglacial mollusks from Estonia (MOLODKOV & RAUKAS 1987).

Tab. 2: ESR-Alter (ka) interglazialer Mollusken von Estland (MOLODKOV & RAUKAS 1987).

No	ESR age (ka)	standard deviation (ka)	χ^2 value
1	95±5		0.0
2	90±8	± 3.2	0.3
3	92±7	± 2.5	0.0
4	92±6	± 2.1	0.0
5	92±9	± 2.0	0.0
6	90±8	± 2.0	0.1
7	82±6	± 4.7	2.5
8	81±9	± 5.2	1.1
mean	89.9 ± 2.4 (Eqs. 5 and 6) ± 6.8 (Eq. 7)	88.6 ± 1.8 (Eqs. 3 and 8) ± 5.2 (Eq. 8)	$\chi^2 = 4.0$ target ~ 7

A normal histogram was used for visualization of the ^{14}C ages from New Zealand (Fig. 7). As there are a large number of determined ages, the shape of the normal histogram is very similar to that of a dispersion histogram. The effect of the choice of the class width is demonstrated using 50, 100 and 200 years. The histogram with the smallest class width (Fig. 7, top) shows a large random scatter of up to $\pm 50\%$, which may lead to incorrect statements on the temporal fluctuation of the sample population. The data density per class obviously is too small.

The histogram with a class width of 100 years seems to be the optimum solution. The class width is within the range of 82 to 169 a for the mean standard deviation of the individual ages (Eqs. 7 and 8). The asymmetry of the distribution in the middle histogram suggests the superposition of three frequency distributions with four outliers between 10,700 and 10,900 years and two outliers between 11,600–11,700 years. With the use of the largest class width of 200 years considerable temporal information is lost, but the histogram still clearly shows that the samples represented more than two events.

9.2 ESR ages from mollusks

In the majority of all case studies the scatter of the data corresponds to that expected from the size of the standard deviations or is wider. The-

re are, however, studies in which the opposite case occurs. In these cases, the defined criteria of the expression $A \pm \sigma A$ are not fulfilled and calculations with such dates should be done without the described error algorithms.

MOLODKOV and RAUKAS (1987) published ESR ages from mollusks collected from an interglacial layer on the coast of Estonia. The results are compiled in Table 2.

The actual uncertainty of the ESR ages (Eq. 3 derived from Eq. 8) is considerably smaller than the expected one calculated from the given standard deviations (Eq. 6). Consequently, $\chi^2_w = 4.0$ (Eq. 9), which is also considerably smaller than 7 (the number of ages minus 1). This result is evidence that the propagation of uncertainty of the published ESR ages was not done properly. This is a frequent observation in published ESR, TL and OSL ages. The standard deviations are too large or too small. A discussion of the reasons for this problem would be beyond the scope of this article. One of the main problems is that the statistical uncertainty of several of the parameters involved can only be estimated or only minimum and maximum values are known for the parameters. In addition, systematic errors are often underestimated. In any case, in such cases it is not justified to use weighting factors in the statistical handling of the numerical ages determined with these methods.

10 Conclusion

It is demonstrated that the handling of numerical ages with their random uncertainties requires at least a minimum knowledge of simple statistical methods, especially the algorithms used to determine the propagation of uncertainty. A careless or unqualified use of such dates bears the risk of misleading geoscientific conclusions.

References

- ALLEY, R.B. (2000): The Younger Dryas cold interval as viewed from central Greenland. – *Quaternary Science Reviews*, 19: 213-226.
- BRUNNACKER, K. & HAUSMANN, R. (1987): Absolutes Alter klimarelevanter Gesteine im jüngeren Quartär. – *Forschung in Köln. Berichte aus der Universität*, 1: 67-70.
- DENTON, G.H. & HENDY, C.H. (1994): Younger Dryas age advance of Franz Josef Glacier in the southern Alps of New Zealand. – *Science*, 264: 1434-1437.
- GEYH, M.A. (1969): Versuch einer chronologische Gliederung des marinen Holozäns an der Nordseeküste mit Hilfe der statistischen Auswertung von ^{14}C -Daten. – *Zeitschrift der deutschen geologischen Gesellschaft*, 118: 351-360.
- GEYH, M.A. (1980): Holocene sea-level history: Case study of the statistical evaluation of ^{14}C dates. – *Radiocarbon*, 22: 695-704.
- GEYH, M.A. (1991): Determination of absolute dates for terrestrial materials (Last Interglacial to Holocene). An appeal for careful interpretation. – In: Frenzel, B. (ed.): *Klimageschichtliche Probleme der letzten 130 000 Jahre*: 251-265; Stuttgart (Fischer).
- GEYH, M.A. (2005): *Handbuch der physikalischen und chemischen Altersbestimmung*. – 211 p.; Darmstadt (Wissenschaftliche Buchgesellschaft).
- GEYH, M.A. & GROSJEAN, M. (2000): Establishing a reliable chronology of lake level changes in the Chilean Altiplano: A result of close collaboration between geochronologists and geoscientists. – *Zentralblatt für Geologie und Paläontologie, Teil I (7/8)*: 985-995.
- GEYH, M.A. & SCHLEICHER, H. (1990): *Absolute Age Determination. Physical and Chemical Dating Methods and Their Application*. – 503 p.; Berlin (Springer).
- HUNTLEY, D.J. (2001): Some notes on language. – *Ancient TL*, 19 (1): 27-28.
- MICHCZYNSKA, D.J. & PAZDUR, A. (2004): Shape analysis of cumulative probability density function of radiocarbon dates set in the study of climate change in the late glacial and Holocene. – *Radiocarbon*, 46 (2): 733-744.
- MOLODKOV, A. & RAUKAS, A. (1987): The age of Upper Pleistocene marine deposits of the Boreal transgression on the basis of electron-spin resonance (ESR) dating of subfossil mollusk shells. – *Boreas*, 17: 267-272.
- SACHS, L. (1999): *Angewandte Statistik. Anwendung statistischer Methoden*. – 881 p.; Berlin (Springer).
- STUIVER, M., REIMER, P.J., BARD, E., BECK, J.W., BURR, G.S., HUGHEN, K.A., KROMER, B., MCCORMAC, G., VAN DER PLICHT, J. & SPURK, M. (1998): INTCAL98 radiocarbon age calibration, 24,000-0 cal BP. – *Radiocarbon*, 40: 1041-1983.
- TAYLOR, B.N. & KUYATT, C.E. (1994): *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. – NIST Technical Note 1297: 20 p.; Washington D.C (National Institute of Standards and Technology).
- WAGNER, G.A. (1998): *Age Determinations of Young Rocks and Artifacts*. – 466 p.; Stuttgart (Enke).