Geochronology accuracy score

Excerpted from: Sundqvist, Kaufman et al., 2014, Arctic Holocene proxy climate database – New approaches to assessing geochronological accuracy and encoding climate variables. Climate of the Past (doi:10.5194/cpd-10-1-2014)

To rate the accuracy of age models from ^{14}C-dated sediment sequences, we developed a procedure that is systematic and reproducible and that focuses on the most important factors that determine the overall accuracy (Appendix A). To facilitate its widespread utility, the simple algorithm is written in the open-source statistical package, R, and is based on the most basic and frequently published information about the materials used for ^{14}C analyses and the resulting age-depth trends. The input variables are: (1) the original ^{14}C ages, including their analytical uncertainties; (2) the age of the core surface (sediment-water interface) if known; (3) sample depths in the core; (4) number of ages rejected by the original author; and (5) material type (one category for the suite of ages; Appendix A).

The accuracy of sedimentary age models depends on the extent to which the constraining ages reliably represent the true timing of sedimentation. The precision of the analyses (the laboratory-reported counting error) might account for only a small part of the overall uncertainty. More important is the extent to which the material dated actually represents the age of the downcore property of interest. Dissolved carbon derived from old sources and incorporated into organisms and minerals that grow in the water (the so-called 'hardwater effect'), or the lag between when an organism grows versus the time it is incorporated into the sedimentary sequence (built-in age) can result in ages that are older than the true age, whereas post-depositional contamination by younger carbon can result in ages that are too young. Many studies have demonstrated the systematic offset between ^{14}C ages of bulk sediment versus the ages of plant macrofossils and tephra layers that they contain. The carbon within bulk sediment may be derived from multiple sources, some of which can be much older than the time of final deposition at the lake or ocean floor. For dating of marine sediments with ^{14}C, determining the marine reservoir effect at a given location through time is an additional challenge. This uncertainty, which is caused by the mixing of old deep water with younger shallow water, may be on the order of several hundreds of years. Identification of well-dated tephras has overcome this problem in some cases.

The accuracy of the geochronology is also determined by the number of ages used to delineate the trend in sedimentation rate relative to the extent to which the sedimentation rate varies at a core site. Where sediment accumulates uniformly through time, fewer ages are needed to determine the trend than for basins that experience variable sedimentation rates. Previous studies that have included an assessment of geochronological accuracy have relied on the difference between the age of the nearest dated sample and the event of interest. This strategy is difficult to implement for studies that do not focus on a particular time slice. In addition, the approach assumes that each ^{14}C age is equally accurate, rather than assuming that different ages might have different accuracies and that the overall trend defined by several ages might average out the random errors or shift a biased age toward the more accurate ages in a series.

We recognize that judging the quality of sample material and weighting the various factors that influence accuracy is subjective. Nonetheless, the rating scheme explicitly recognizes the key factors that influence the geochronological accuracy of sedimentary sequences that do not lend themselves to conventional statistical approaches, and assigns reasonable numerical ratings based on a simple, reproducible, and customizable procedure.

Appendix A

Geochronology accuracy score

The geochronology accuracy score (chron score) combines three indicators of the reliability of sediment-based age models, namely: the delineation (D) of the downcore trend, the quality (Q) of the dated samples, and the (P) precision of the ^{14}C ages.

Delineation of downcore trend. The accuracy of an age model depends on how well the analyzed samples delineate changes in sedimentation rate downcore. If the sedimentation rate is linear, then only two ages are needed to define it. With increasing variability of sedimentation rates, more ages are required to delineate accurately the downcore trend. In absence of stratigraphic information that attests to where within a sequence the sedimentation rate is most likely to have changed, evenly spaced samples increase the chances of capturing changes in sedimentation rate compared with the same number of ages clustered in small intervals. We therefore assess the extent to which an age model is accurately delineated by combining three attributes: (1) the frequency of ages, (2) the regularity of their spacing, and (3) the uniformity of the downcore trend.

The frequency of ages (f) is quantified as the number of ages relative to the length of time represented by the sedimentary sequence, or:

$f = (t_{max} – t_{min}) / n_{tot}$

Where t_{max} and t_{min} are the oldest and youngest ages, respectively, and n_{tot} is the total number of ages that were accepted by the author of the age model. The age of the core surface is included if the sediment-water interface was captured at the time of coring.

The regularity (r) of a series of ages is quantified by the standard deviation of the length of time that separates consecutive ages, or:

$r = s_{[t_n – t_{n+1}]}$

Where s is the standard deviation, and $t_n – t_{n+1}$ is the difference in time between the n^{th} age and the next older age, as assessed for each age in a series. The absolute value is used for downcore age reversals.

The uniformity (u) of the trend is quantified as the root mean standard error (RMSE) with respect to a cubic smoothing spline with degree of freedom (df) = 4. If the spline fit contains a reversal, then the df is lowered incrementally until there are no reversals in the spline fit.

The three attributes that make up the age-model delineation (D) can each be weighted to adjust their relative importance in the overall D score:

$$D = w_f f + w_r r + w_u u$$

where w_f , w_r, and w_u are weighting factors. Because the frequency of ages is fundamental to the accuracy of the age model, and because the RMSE is generally a low value, we chose to increase their weight in the overall score. Namely, we set the weighting factors = 2, 0.5, and 3, respectively. The D value increases with decreasing delineation.

Reliability of dated samples. The accuracy of age models generally depends on the type of material analyzed, with some material types typically yielding ages that more closely represent the timing of deposition than others. In addition, the influence of contamination by young carbon, or the reworking of older material into younger sediment is often indicated by ages that violate stratigraphic superposition. We therefore assess the "quality" of dated materials based on two criteria: (1) the proportion of outliers and stratigraphically reversed ages, and (2) a qualitative (categorical) score based on the type of material dated.

Standard practice is to report the results of all radiocarbon analyses from a core or series of cores, then to identify and exclude the outliers if they exist. These analyses are indicated as rejected by the authors of the original age model. Minor age reversals are often retained in the age model, and the sedimentation-rate smoothing function is used to average the differences. In our scoring scheme, the proportion (p) of outlier and stratigraphically reversed ages is the number of ages that were rejected by the original author, plus the number of stratigraphically reversed ages relative to the total number of dated samples, or:

$$p = 1 - (n_{rej} + n_{rev}) / n_{tot}$$

where n_{rej} is the number of ages rejected by the original author and therefore not included in the list of ages used to calculate the delineation (D) score, n_{rev} is the number of ages that are at least 100 year older than the next age downcore. The proportion is subtracted from 1 so that higher p values signify a higher proportion of accepted and monotonically arranged ages.

We developed a five-fold classification scheme for the types of material (m) used for the ^{14}C analyses. A value of 1 through 5 is assigned to the entire series of samples, depending largely on the extent to which they comprise reliable types of sample material based on a specific set of criteria (see below). A value of 5 is reserved for age models that have been checked by independently derived ages from correlated tephra layers or ^{14}C wiggle matching. Separate classification schemes were developed for lacustrine and marine materials. m values for lacustrine materials:

5 = at least one age can be confirmed by tephra or ^{14}C wiggle matches; no bulk-sediment

4 = mainly (> 90%) plant macrofossils

3 = 50-90% plant macrofossils; bulk-sediment ages can be reasonably adjusted

2 = < 50% plant macrofossils

1 = all bulk-sediment ages

m values for marine materials:

5 = >90% whole, monospecific forams with a constrained reservoir age (at least one well-dated tephra or wiggle match used to determine the reservoir correction)

4 = mainly (>90%) monospecific forams

3 = >50% monospecific forams and articulated mollusks

2 = mixture of sample types: fragmented and whole; monospecific and mixed species

1 = mainly (>90%) fragmented and unidentifiable tests and shells

To derive a single value for the quality (Q) of the dated samples, we take the product of the two attributes, the proportion (*p*) of accepted, monotonic ages, and the material (*m*) type category, or:

$Q = pm$

Q values increase with increasing sample quality.

Precision. All radiocarbon laboratories report the ± 1s analytical precision associated with the internal reproducibility of the counting statistics for ^{14}C ages. The analytical precision is controlled by the mass of carbon used for AMS analysis, or the activity of the sample used for decay counting methods, and the length of time that the sample is analyzed on the instrument. The extent to which the overall accuracy of an age model is influenced by the analytical precision is difficult to quantify. In general, analytical precision is on the order of decades, but the uncertainty is amplified when calibrated to calendar years. We developed a simple index for the precision (P), which is based on calibrated age ranges of the ^{14}C ages using the INTCAL04 calibration dataset.

$P = s^{-1}$

where *s* is the mean 2s range of all calibrated ^{14}C ages. The inverse function is used to stratify the precision scores over the most precise end of the range (decadal scale) while de-emphasizing the differences among the less-precise ages (centennial scale). P values increase with increasing precision.

Geochronology accuracy score. The overall score (G, chron score) is calculated by summing the weighted values of each of the three components:

$G = -w_D D + w_Q Q + w_P P$

Where w_D, w_Q, and w_P are the weighting factors, which we set = 0.001, 1 and 200, respectively, so that each component is of the same order of magnitude.