THE PRINCIPLE OF THE BAYESIAN METHOD

*By Franz Weninger, Peter Steier, Walter Kutschera, Eva Maria Wild**

Introduction

The Bayesian statistical method has now been used for about ten years in radiocarbon dating. It has become a widely used and commonly accepted method to consider additional information about the samples within the calibration process. (For details on the standard calibration procedure for radiocarbon age determination see e.g. Bowman 1990.) In most cases, the available additional information is the knowledge about time relations of the samples, derived from stratigraphic conditions. Depending on the shape of the calibration curve the common calibration procedure may produce calibrated ages with much larger uncertainty ranges than that of the uncalibrated radiocarbon ages. For instance large wiggles in the calibration curve enlarge the uncertainty of the calibrated age considerably, as shown in details below. Including additional information within the calibration process by means of Bayesian statistics can compensate for this increase of the uncertainty.

Although the Bayesian method is widely used, it is often difficult for the user to get a clear idea about the detailed procedure underlying the method. Therefore, it is our intent to give a simple description of the basic mechanism of the method in this article. Formulas, avoided in the text, are presented in a short appendix.

How the Bayesian Method Works

First Example

Before entering into Bayesian statistics, let us first have a look at the normal calibration process of a single radiocarbon sample. We start with a radiocarbon measurement with a known uncertainty. On the vertical axis in Fig. 1, x_1 indicates the measured radiocarbon age surrounded by a Gaussian-shaped probability function, that gives the uncertainty of the radiocarbon date x_1 due to the measurement error. The function $\mu(\Theta)$ is the calibration curve that gives the relation between the uncalibrated – the so-called conventional – radiocarbon age based on the measured $^{14}C/^{12}C$ ratio of the sample, and the calibrated (true) age Θ. The calibration curve used in this example is an artificial one, chosen to support our example. (It is also assumed that this calibration curve has no uncertainty attached to it, while in reality it does have a finite width due to the procedure to establish the calibration curve itself.) The units for the numbers of both radiocarbon age and calibrated age are given in years Before Present (yr BP; present corresponds to 1950 AD). Note, that the calibrated age on the horizontal axis of Fig. 1 – and also in all analogue figures – increases from the left to the right. This presentation of calibrated ages is not very common, but for the following discussion it is better to envision increasing ages on both axes.

Age calibration of a single sample means to find the so-called likelihood function L_1, that shows how well a particular true age Θ fits to the measured radiocarbon age x_1. The procedure to

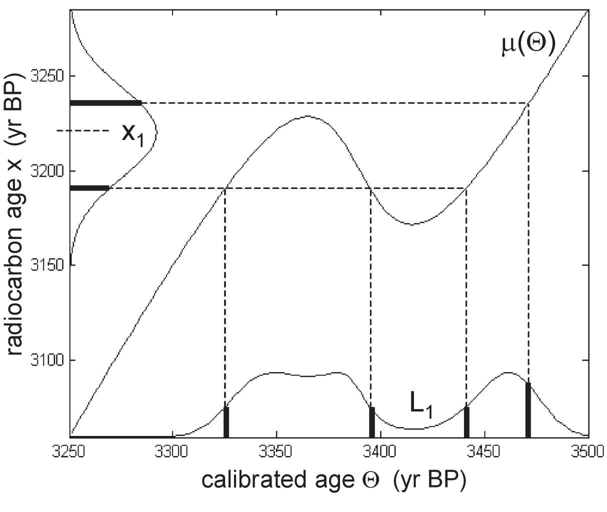

Fig. 1 The principle technique to calibrate a single radiocarbon age. The single sample calibration or 'likelihood function' L_1 shows how well a particular calendar age Θ fits to the measured value x_1. Technically this function is produced by transforming the Gaussian shaped probability distribution of the radiocarbon measurement to the axis of the calibrated age, by using the calibration curve $\mu(\Theta)$. Note that an artificial calibration curve is used here to produce an obvious example

* VERA Laboratory, Institut für Isotopenforschung und Kernphysik, Universität Wien, Währinger Straße 17, A-1090 Wien.

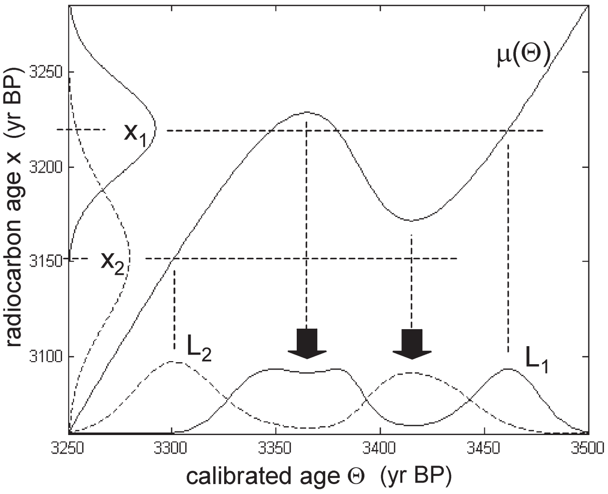

Fig. 2 *First example:* There are two samples with radiocarbon ages x_1 and x_2 and the additional information that sample 1 is younger than sample 2. By independent calibration, we get two likelihood functions L_1 and L_2. Obviously, only the peaks marked with arrows agree with the assumed chronological order

get this likelihood function is very simple. One only has to take the value of the Gaussian probability function around x_1 for each position on the radiocarbon age axis and to apply this value to the corresponding position on the calibrated age axis, as given by the calibration curve. This procedure is illustrated by the dashed lines for two positions in Fig. 1. (Please note that only the shape of the probability function carries the information relevant to our discussion. Therefore, we don't use units on axes characterising probabilities.) The shape of the likelihood function L_1 clearly shows the main problem of radiocarbon calibration. Due to the wiggle in the calibration curve, the likelihood function is dispersed and produces two regions on the true or calibrated age scale that match the measured radiocarbon age. It cannot be clarified from which of the two time ranges the sample originates by investigating one sample only.

Now we assume a second sample yielding the uncalibrated radiocarbon age x_2. In Fig. 2 one can see both radiocarbon ages x_1 and x_2 and the likelihood functions of the two samples L_1 and L_2. Furthermore, we assume that sample 1 is younger than sample 2, which is a common type of additional information provided by a stratigraphic situation of an archaeological excavation. Obviously, only the marked peaks of L_1 and L_2 in Fig. 2 agree with this given chronological order and the other parts of the likelihood functions should be suppressed. What we need, is a method that can

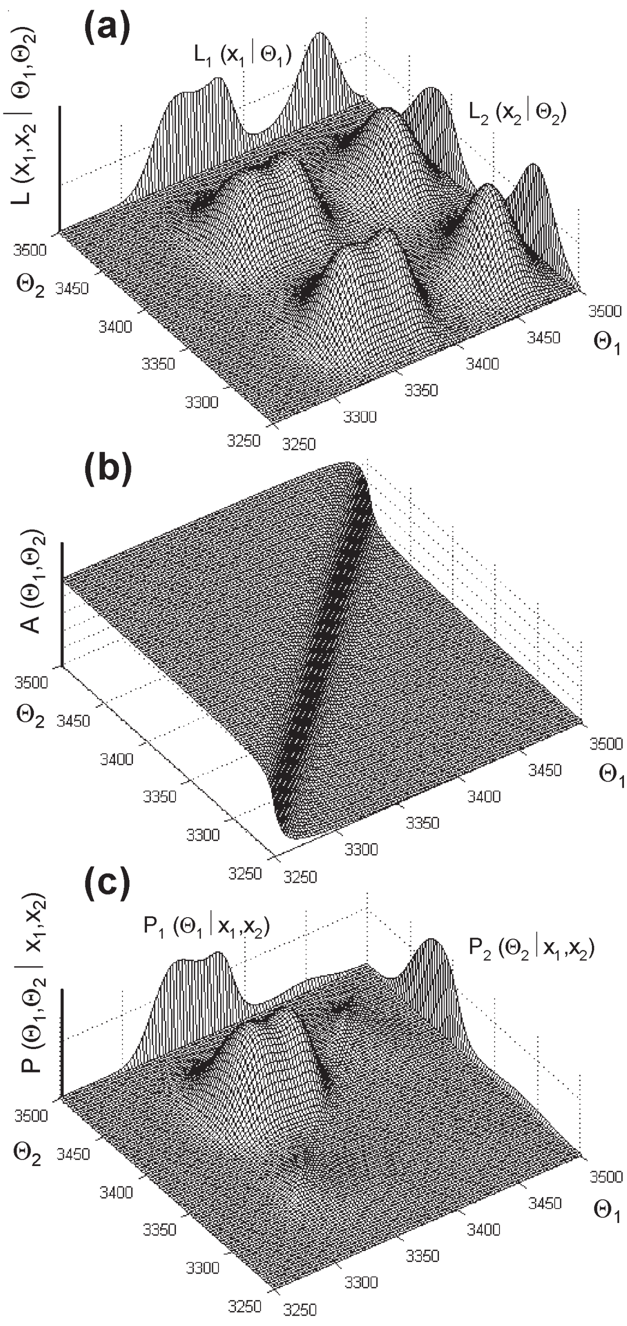

Fig. 3 *First example (continued):* **a)** The two-dimensional likelihood function L, that is the product of both single-sample likelihood functions L_1 and L_2. **b)** The prior (or 'a priori') probability A, containing the available additional information on the true sample ages Θ_1 and Θ_2. In our example, sample 1 is known to be younger then sample 2. So, the prior probability is high when $\Theta_1 < \Theta_2$ (left of the diagonal), otherwise it is low (right side). **c)** The posterior probability P is the product of likelihood function L and prior probability A. Consequently, all regions where the prior probability is low are suppressed, which can clearly be seen. Finally, the marginal posterior probabilities P_1 and P_2 are the probabilities for the age of each single sample. They are the projections of the two-dimensional posterior probability P to the sample age co-ordinates. (For details see text.)

deal with such additional information in a general mathematical way. The Bayesian statistics is exactly the mathematical formalism which is needed to utilise this additional information.

In the first step, the two likelihood functions are multiplied point by point, resulting in the two-dimensional likelihood function shown in Fig. 3a. On the edges of the plane spanned by the two axis of true or calibrated age Θ_1 and Θ_2, the normal likelihood functions L_1 and L_2 as explained above (Fig. 2) are shown again. Each value of the two-dimensional function is the product of the corresponding values of L_1 and L_2. The meaning of this two-dimensional likelihood function is now a probability of age combinations. The function value at a particular point in the plane, e.g. with age co-ordinates Θ_1 = 3350 yr BP and Θ_2 = 3400 yr BP, indicates the degree of agreement of this particular age combination with the set of measured radiocarbon ages (x_1 and x_2). In this example we get a two-dimensional function. In general the dimensionality of the likelihood function – and also the prior and posterior function discussed below – is equal to the number of radiocarbon samples, and can be quite high. High dimensional functions are definitively a problem for our imagination, nevertheless the principal mechanism of the Bayesian method remains the same. Therefore we will continue with our two-dimensional example for better understanding.

The reason for using multi-dimensional functions lies in the fact that it is a convenient way to introduce the additional information mentioned above into the procedure. In Fig. 3b the prior ('a priori') probability function A is shown. This function is the two-dimensional representation of the additional information about the sample ages derived from the stratigraphy. In this example, sample 1 is known to be younger than sample 2. Therefore the prior probability is high for all combinations of age Θ_1 and age Θ_2 where Θ_1 is less than Θ_2 (left of the diagonal in Fig. 3b) and otherwise it is low (right side of Fig. 3b).

Having found the prior probability, the next step is to combine it with the two-dimensional likelihood function established above, which contains the information from the radiocarbon measurements. This is easily done by taking the product of the two functions, that means to multiply the values of the likelihood function and the prior function for each position in the two-dimensional co-ordinate plane. The result is the posterior function P, shown in Fig. 3c. This function gives now the probability for any particular combination of the sample ages Θ_1 and Θ_2 to be the true one, including both, the information from measurement and the additional information from stratigraphy. In our example, three of the four peaks of the two-dimensional likelihood function are located in the region with low prior probability and therefore are strongly suppressed within the posterior function.

The last step in our procedure is to go back from the multi-dimensional co-ordinate space representing age combinations to usual probabilities for the single samples again. This is done by projecting the multi-dimensional posterior function P to the sample age co-ordinates, getting the so-called marginal posterior probabilities P_1 and P_2 of the single samples, also shown in Fig. 3c. They give the probability that a considered sample has a particular true age based on both, measurements and stratigraphic information. This is the result we intended to achieve. In Fig. 4 we finally compare the initial single calibration likelihood functions L_1 and L_2 with the resulting marginal posterior probabilities P_1 and P_2. It is shown very clearly, that only these peaks remain within the posterior probability, that fulfil the condition that sample 1 is younger than sample 2, according to our prior knowledge. This is exactly what we claimed ini-

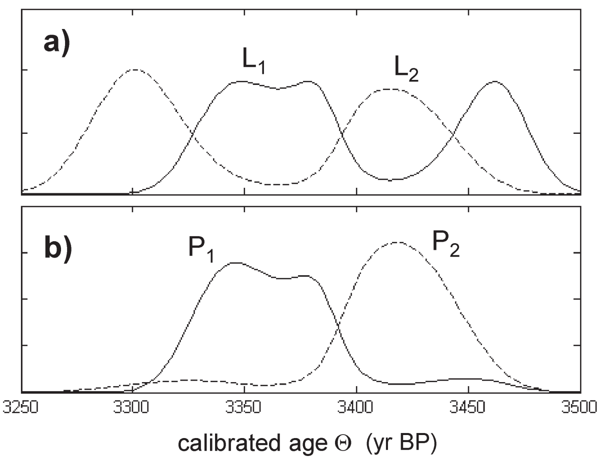

Fig. 4 *First example (continued):* Comparison of the single sample likelihood functions L_1 and L_2 that contain no prior information **(a)** with the marginal posterior probabilities P_1 and P_2 including the prior information **(b)**. Only these peaks remain within the posterior probability fulfilling the condition that sample 1 is younger than sample 2, according to the existing prior knowledge

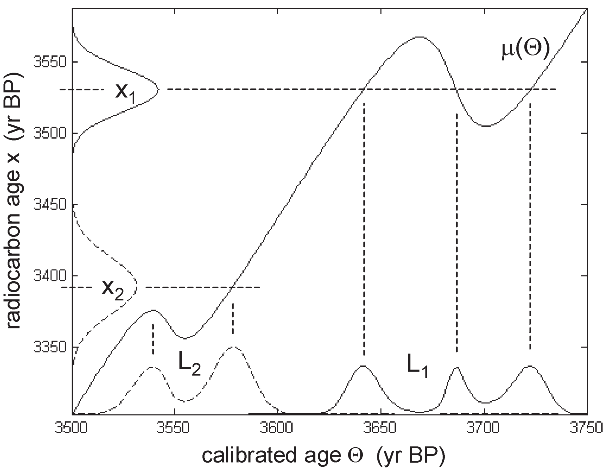

Fig. 5 *Second example:* Again we take two samples with radiocarbon ages x_1 and x_2. L_1 and L_2 are their likelihood functions obtained by independent calibration

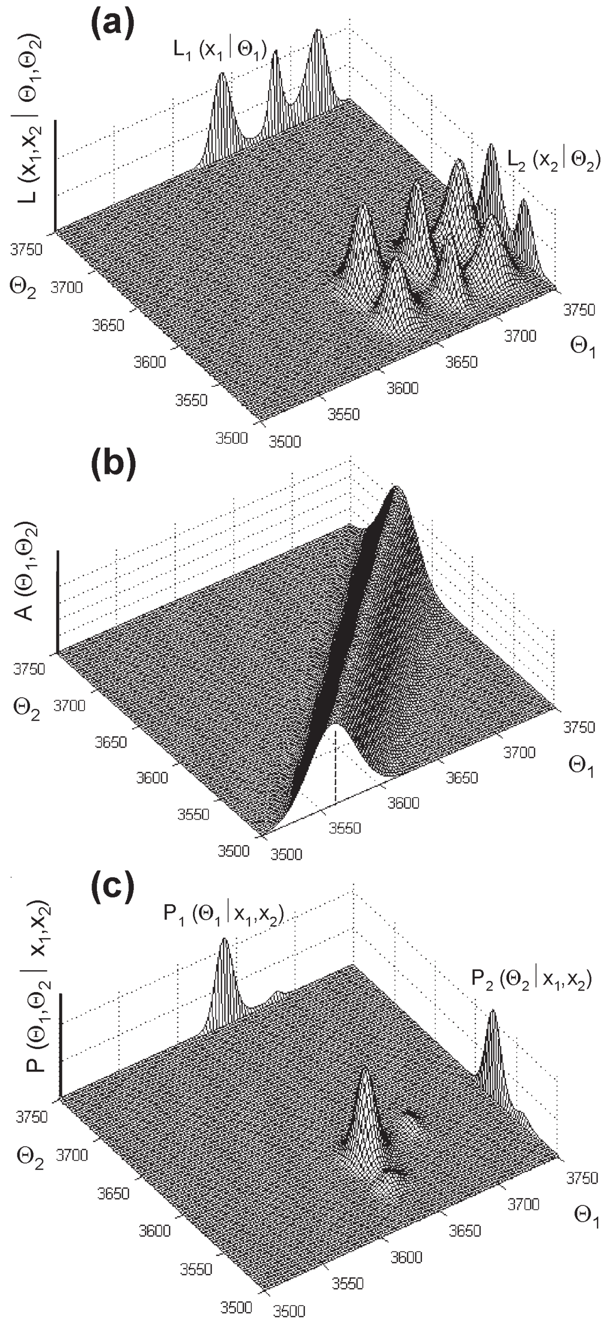

Fig. 6 *Second example (continued):* **a)** Two-dimensional likelihood function L, and single-sample likelihood functions L_1 and L_2. **b)** Prior probability A, containing the additional information on the true sample ages Θ_1 and Θ_2. Unlike the first example, we now assume that sample 1 is older than sample 2 by a particular known value with a given uncertainty (60 ± 20 yr). This leads to the form shown of the prior probability, which looks like a wall with Gaussian cross section. **c)** Two-dimensional posterior probability P and marginal posterior probabilities P_1 and P_2. (For details see text.)

tially based on a qualitative consideration that the Bayesian method should produce. So, in our simple example, the method works very well, and it can be assumed, that the Bayesian method is applicable in other situations as well.

Second Example

Let us discuss shortly a second – also artificially constructed – example, similar to the one described above, but now using a different kind of prior information. Again we restrict ourselves to only two radiocarbon measurements for the reason of simplified visualisation. Fig. 5 shows the determination of the single sample likelihood functions, as already explained above. Due to all possible combinations of the three peaks of L_1 with the two peaks of L_2 there six peaks arise in the multi-dimensional likelihood function, shown in Fig. 6a. Unlike the first example, we now assume that sample 1 is older than sample 2 by a particular known value with a given uncertainty. This leads to the shape of the prior probability shown in Fig. 6b, looking like a wall with Gaussian cross section. Analogous to the first example, building the product of the multi-dimensional likelihood function L and the prior function A lets only this peak remain within the posterior function P that fulfils the prior condition. Or in other words, only this peak in the multi-dimensional co-ordinate plane remains, that originates from the combination of regions in the single calibration likelihood functions L_1 and L_2 having the required age difference. Building the marginal posterior probabilities as explained above, it is again clearly shown that all regions are suppressed that do not match the required condition (Fig. 7).

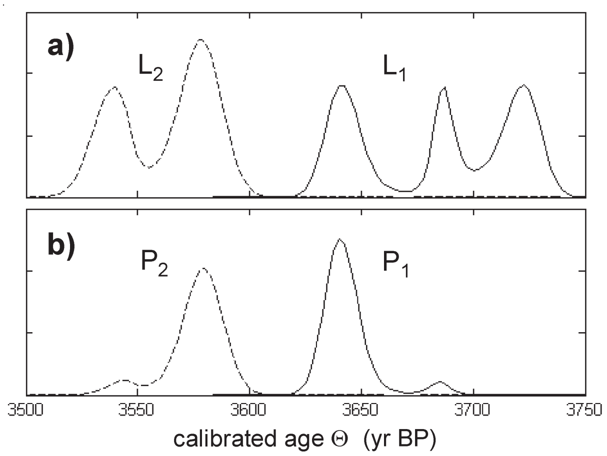

Fig. 7 *Second example (continued):* Comparison of the single sample likelihood functions L_1 and L_2 **(a)** with the marginal posterior probabilities P_1 and P_2 **(b)**. Analogous to the first example, again only the peaks that are in agreement with the prior knowledge remain within the posterior probability which means that sample 1 has to be 60 ± 20 years older than sample 2

Third Example – a 'Real-World' One

An early investigation, where Bayesian statistics were used in conjunction with radiocarbon dating was performed by Buck *et al.* 1994 for an excavation of an Austrian Bronze Age village. We use here the radiocarbon dates and the stratigraphic relations investigated in this work to show a realistic application of the Bayesian method.

Naturally in a real-world application, the situation is more complex then in the artificial examples previously shown. Now we have ten-dimensional functions according to a sample number of ten, a lot of wiggles in the real calibration curve, and a more complex prior information (see Fig. 8). However, the method works just the same way as explained above. Fig. 8a shows the pattern of the overlapping uncertainty ranges of the radiocarbon dates of ten measured samples, indicated by their Gaussian probability distributions, as well as the relevant section of the radiocarbon calibration curve (IntCal 04 by Reimer *et al.* 2004). Part (b) in Fig. 8 gives a symbolic representation of the stratigraphic relations, which must be read in the following way: Sample 10 is known to be younger than sample 9 and 9 is known to be younger than 8. The samples 3, 4, 6 and 7 are all known to be older than 8, but the age relation between them is unknown. Again, sample 3 is

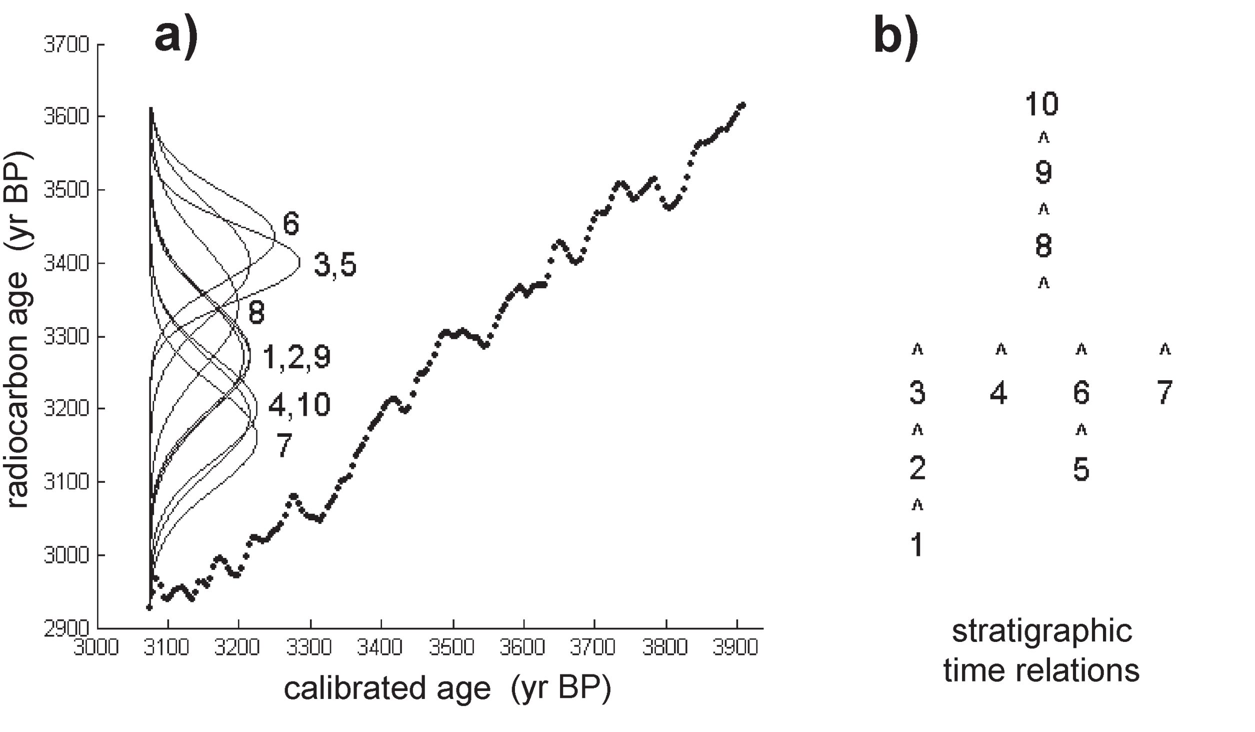

Fig. 8 *Third example:* In a real-world application the situation is more complex then in the artificial examples shown in the previous figures. Now we have functions within a ten-dimensional mathematical space, many wiggles in the real calibration curve and more complex prior relations. The time relations between the individual samples based on the stratigraphic information are given in **(b)**: sample 10 is known to be younger than sample 9, sample 9 younger than 8. Samples 3, 4, 6 and 7 are all known to be older than 8, but the relationship between them is unknown. Again, 3 is younger than 2 and so on. The data are taken from Buck *et al.* 1994 from an excavation of an Austrian bronze age village

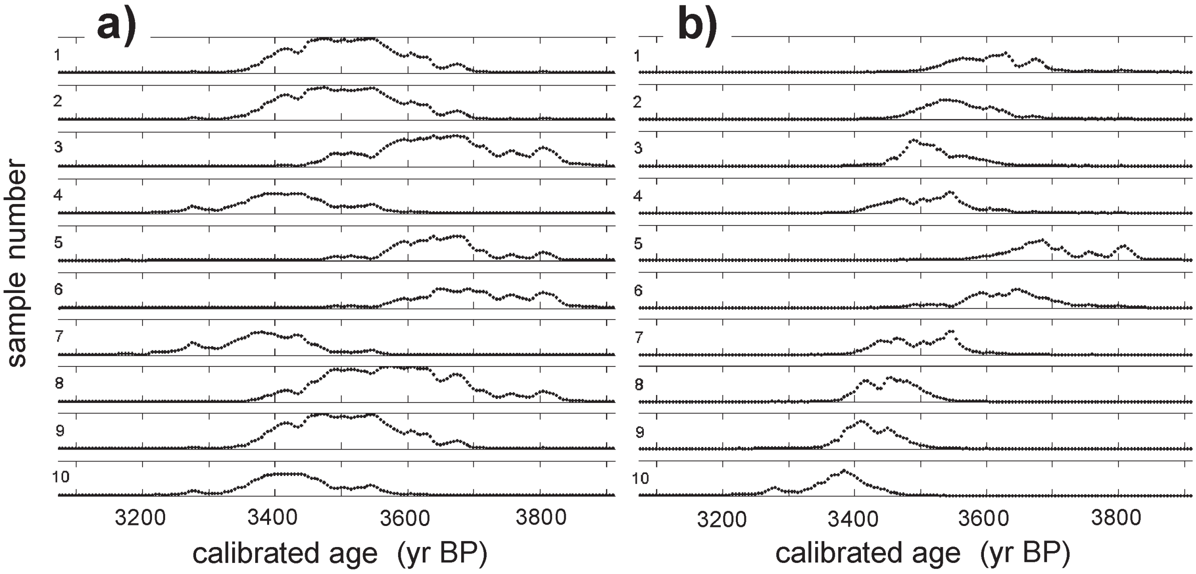

Fig. 9 *Third example (continued):* Single sample calibration without prior information **(a)**, compared with the outcome of the Bayesian method, including the prior information **(b)**. Once again parts of the probability distributions that are not in agreement with the prior knowledge are suppressed

younger than 2 and so on. Nearer to archaeological terms, the stratigraphy contains two phases. The younger one contains the sequence 8-9-10, the older one the sequences 1-2-3 and 5-6 and additionally the samples 4 and 7. Across these sequences (e.g. sample 3 vs. 5) no relation is known.

In Fig. 9 we show a comparison of the likelihood functions from single sample calibration and the marginal posterior functions produced by the Bayesian method for each sample, analogous to Fig. 4 and Fig. 7 in the previous examples. It is not necessary to discuss these functions in detail, there can easily be seen two qualitative aspects. First, the time ranges of possible ages turn out to be much smaller in the posterior functions then in the likelihood functions. Second, the regions are shifted in accordance to the stratigraphic relations. This can be seen when looking e.g. at sample 8. As mentioned above, sample 8 has to be younger then the samples 3, 4, 6 and 7. So, roughly spoken, all these samples try to shift the posterior function of sample 8 to the younger side of the age scale. This is what one sees when comparing the likelihood function (Fig. 9a) and posterior function (Fig. 9b) of sample 8. Of course, both of the effects described have again one basic reason: the suppression of all parts of the (multi-dimensional) posterior function, which do not agree with the prior probability, i.e. the stratigraphical information.

Although we have seen that a large number of samples does not change the method in principal, there is a non-negligible technical difference, which will be discussed below. For further reading about the application of Bayesian methods in radiocarbon dating see e.g. Buck *et al.* 1991, Bronk Ramsey 1995, 1998, Bronk Ramsey *et al.* 2004, Steier and Rom 2000.

Gibbs Sampling, the Key to Numerical Feasibility

For the simple examples with only two samples measured, we had to deal with two-dimensional functions. If we suppose to work with 100 points on each of the two age co-ordinates within the numerical calculations, this leads to 100^2 = 10 000 points within the co-ordinate plane spanned by two age co-ordinates. So, all functions have to be evaluated at 10 000 points. This is for sure no problem for a computer. But if we suppose now that there are 15 samples and we again use 100 points on each age axis, we get $100^{15} = 10^{30}$ points to be evaluated for our now 15-dimensional functions. Therefore, it is not possible any more to calculate these functions point by point. Fortunately, a very efficient Monte-Carlo method can solve this problem, the so-called Gibbs sampling.

The principle of doing the calculations with a Monte Carlo method is to evaluate not all points in the multi-dimensional co-ordinate space, but only a number of randomly drawn ones. It would not, however, be very efficient to draw the points really randomly out of the multi-dimensional space, because many points would lie in areas where the posterior function is nearly zero, as one can imagine when looking at Fig. 3c or Fig. 6c. These points would be useless, because they do not contribute to the resulting marginal posterior functions. In contrast, the Gibbs sampling finds a pattern of points in the multi-dimensional co-ordinate space where the density of the points reproduces the probability function investigated. These points can be found by evaluating only one-dimensional slices within the multi-dimensional probability function. These slices are conditional probabilities with all dimensions but one fixed. The procedure is the following: One starts with calculating a slice e.g. in the first dimension, choosing all other co-ordinates arbitrarily. This slice represents a probability distribution along the first dimension and out of this distribution a position on the first co-ordinate is randomly drawn. Next, a further slice is calculated along the second dimension, located at the position of the first draw. All other co-ordinates stay the same. This procedure is repeatedly performed. When reaching the last co-ordinate it jumps back to the first and iterates all over. With each change of any co-ordinate a new point is found. It can be shown theoretically, that the density of their pattern converges to the probability function processed.

Within the Bayesian method, Gibbs sampling is used to find the marginal posterior probabilities of the samples. As explained previously, the marginals are projections of the multi-dimensional posterior probability to the single sample co-ordinates. Mathematically, they are evaluated by an integration – or summation when done numerically – over the age co-ordinates of all other samples. With the Gibbs sampling method this summation can easily be done by projecting every selected point onto the corresponding positions on each sample co-ordinate and adding them up. This produces the correct marginal probability distributions, because the density of the points represents already the probability to be integrated. So we only have to run the Gibbs sampling on the multi-dimensional posterior probability. And the great advantage is, that there is no need to evaluate all points of the posterior function before, because the method uses only one-dimensional slices of the function.

Therefore, the use of Gibbs sampling makes Bayesian statistics numerically feasible, allowing it to become a powerful tool in radiocarbon calibration widely used today. For more details to Gibbs sampling and related numerical methods see e.g. GILKS *et al.* 1996.

FINAL REMARK

We hope that this work shed some light on a method which – although intrinsically complex – is very useful for the reduction of the well-known uncertainties of standard radiocarbon dating. Of course, there are many detailed aspects and specialised applications of the method, as well as various questions and problems that are not mentioned in this article.

APPENDIX: BASIC MATHEMATICAL RELATIONS

Here we present the mathematical formulation of the Bayesian method used in this article. The equations are particularly given for the second example described in the text.

$$L_1(x_1|\Theta_1) \sim e^{-\frac{(x_1-\mu(\Theta_1))^2}{2\sigma_1^2}}$$

$$L_2(x_2|\Theta_2) \sim e^{-\frac{(x_2-\mu(\Theta_2))^2}{2\sigma_2^2}}$$

$$L(x_1,x_2\,|\,\Theta_1,\Theta_2) \sim L_1(x_1\,|\,\Theta_1) \cdot L_2(x_2\,|\,\Theta_2)$$

$$A(\Theta_1,\Theta_2) \sim e^{-\frac{(\Theta_1-\Theta_2-d)^2}{2\sigma_d^2}}$$

$$P(\Theta_1,\Theta_2\,|\,x_1,x_2) \sim L(x_1,x_2\,|\,\Theta_1,\Theta_2) \cdot A(\Theta_1,\Theta_2)$$

$$P_1(\Theta_1\,|\,x_1,x_2) \sim \int_{-\infty}^{+\infty} P(\Theta_1,\Theta_2\,|\,x_1,x_2)\; d\Theta_2$$

$$P_2(\Theta_2\,|\,x_1,x_2) \sim \int_{-\infty}^{+\infty} P(\Theta_1,\Theta_2\,|\,x_1,x_2)\; d\Theta_1$$

The variables used are exactly the same as in the text: x_1 and x_2 are the uncalibrated radiocarbon ages of sample 1 and 2 with their measurement errors σ_1 and σ_2. Θ_1 and Θ_2 are the calibrated true ages of the to samples and $\mu(\Theta)$ is the radiocarbon calibration curve. $L_1(x_1|\Theta_1)$ (read: "the probability L_1 of x_1 given Θ_1") and $L_2(x_2|\Theta_2)$ are the one-dimensional likelihood functions for single sample calibration. The notation $L_1(x_1|\Theta_1)$ means, that L_1 is the conditional probability to measure the value x_1 when having a true age Θ_1. $L(x_1,x_2|\Theta_1,\Theta_2)$ is the multi-dimensional likelihood function. $A(\Theta_1,\Theta_2)$ is the prior probability for the special case, that Θ_1 is known to be older then Θ_2 by a particular age difference d with an uncertainty σ_d. Further $P(\Theta_1,\Theta_2|x_1,x_2)$ is the multi-dimensional posterior probability. This is the conditional probability, that the particular combination of age Θ_1 and age Θ_2 is the true one, when having measured x_1 and x_2. Finally, $P_1(\Theta_1|x_1,x_2)$ and $P_2(\Theta_2|x_1,x_2)$ are the marginal posterior probabilities of the true age of sample 1 and sample 2.

Bibliography

BOWMAN S.

1990 *Radiocarbon dating – (Interpreting the past).* British Museum Publications Ltd, London.

BRONK RAMSEY C.

1995 Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon* 37(2), 425–430.

1998 Probability and dating. *Radiocarbon* 40(1), 461–474.

BRONK RAMSEY C., MANNING S.W. and GALIMBERTI M.

2004 Dating the volcanic eruption at Thera. *Radiocarbon* 46(1), 325–344.

BUCK C.E., KENWORTHY J.B., LITTON C.D. and SMITH A.F.M.

1991 Combining archaeological and radiocarbon information: a Bayesian approach to calibration. *Antiquity* 65 (1991), 808–821.

BUCK C.E., LITTON C.D. and SHENNAN S.J.

1994 A case study in combining radiocarbon and archaeological information: the early Bronze Age settlement of St. Veit-Klinglberg, Land Salzburg, Austria. *Germania* 72(2), 427–447.

GILKS W.R., RICHARDSION S. and SPIEGEL-HALTER D.J. (eds.)

1996 *Markov Chain Monte Carlo in Practice*, London.

REIMER P.J. *et al.*

2004 IntCal04 Terrestrial Radiocarbon Age Calibration, 0–26 cal kyr BP. *Radiocarbon* 46(3), 1029–1058.

STEIER P. and ROM W.

2000 The use of Bayesian statistics for ^{14}C dates of chronologically ordered samples: a critical analysis. *Radiocarbon* 42(2), 183–198.