Statistical Modelling for Dating Ice Cores

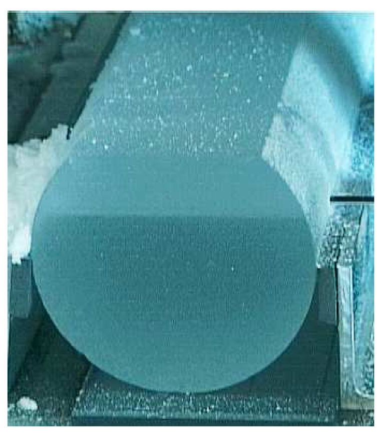
Katy Klauenberg, Paul G. Blackwell and Caitlin E. Buck
Department of Probability and Statistics, University of Sheffield

Regine Röthlisberger
British Antarctic Survey, Cambridge, UK

MOTIVATION

http://www.photo.antarctica.ac.uk/external/guest/detail/search/10002531/9/8

Ice cores preserve valuable information about past environment and climate. To interpret the information within them, it is pivotal to date the ice core, i.e. to relate time to depth. Existing dating methods can be categorised as follows: (1) layer counting using the seasonality in signals, (2) glaciological modelling describing processes such as snow accumulation and plastic deformation of ice, (3) comparison with other dated records, or (4) any combination of these. Currently, none of these methods benefit from statistical modelling.



Source: BAS image database

We are pioneering the combined use of glaciological and statistical models within a Bayesian framework. By formalising the uncertainty in glaciological relations, the dating uncertainty can be derived. This approach is applied to the dating of Antarctic ice cores and then compared to an independent dating derived from layer counting. For the first time, the effects of uncertainty implied by the dating method are investigated for ice core chronologies. This provides valuable insights for the applied community.

MODEL SET-UP AND ASSUMPTIONS

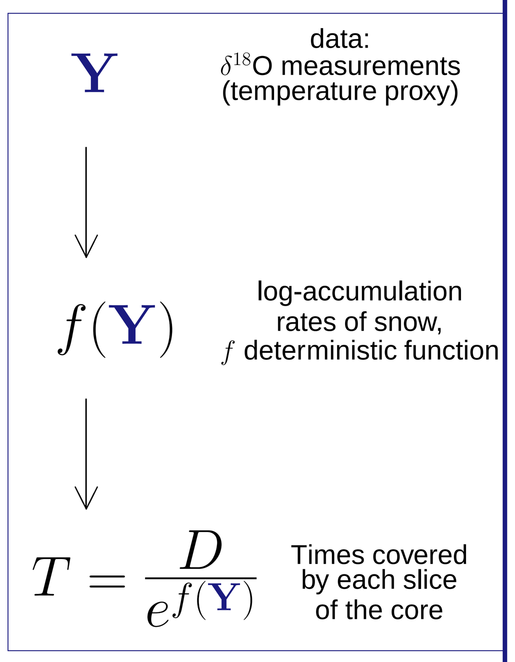
Ice cores are typically cut into slices and their chemical and physical properties analysed. We focus on measurements of the oxygen isotope ratio δ^{18}O, which is a temperature proxy used to help date ice cores.

A glaciological model

- It is assumed that a function f exists which transforms the measurements $\mathbf{Y} = (Y_1, Y_2, \ldots)$ of the temperature proxy δ^{18}O at depths $d_1, d_2, \ldots$ to observed log-accumulation rates of snow $f(\mathbf{Y})$.
- In particular

$$f(Y_i) = A_0 + a(Y_i - \delta_1) + \frac{1}{2}b(Y_i^2 - \delta_1^2)$$

shall be used here (as in[1]), where $\theta = (A_0, c_1, c_2, \delta_1, \delta_2)$ is conventionally fixed and $b = \frac{c_1 - c_2}{\delta_1 - \delta_2}, a = c_1 - \delta_1 b$.
- For each slice i the measurement Y_i prevails from depth d_i down to depth d_{i+1}, such that the elapsed time T_i is inversely proportional to the accumulation rate. (Let $D = (d_2 - d_1, d_3 - d_2, \ldots)$.)



Combining glaciological and statistical modelling

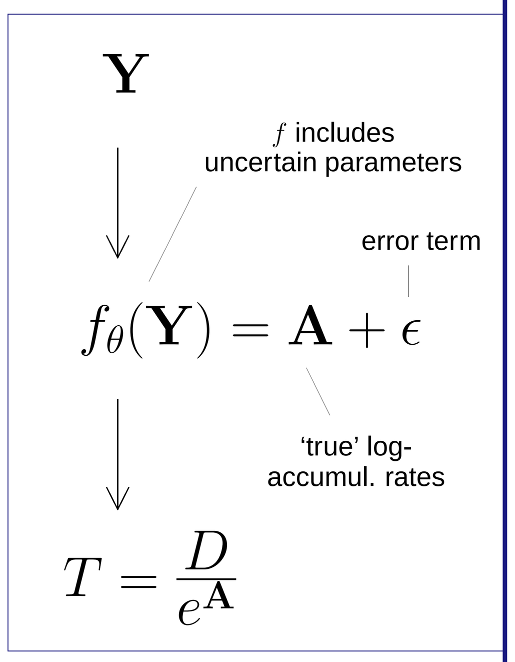
- To quantify the uncertainty in dating ice cores we assume that the observed log-accumulation rates $f(\mathbf{Y})$ are additively related to true log-accumulation rates $\mathbf{A}$.
- As a starting point these are assumed to be mutually independent

$$P(\mathbf{A}) = \prod_i P(A_i). \quad (1)$$

- The errors are assumed to be independent and normally distributed:

$$f(Y_i) = A_i + \varepsilon_i, \text{ with } \varepsilon_i \sim N(0, \sigma^2). \quad (2)$$

- Additionally, uncertainty in the parameters θ of the relationship f is incorporated.



Quantities of Interest

- The posterior distribution of the estimated times covered by each slice $P(\mathbf{T}|\mathbf{Y})$ derived from

$$P(\mathbf{A}|\mathbf{Y}) \propto P(\mathbf{A})P(\mathbf{Y}|\mathbf{A}) \propto P(\mathbf{A}) \int_\theta \int_\sigma P(f_\theta(\mathbf{Y})|\mathbf{A}, \theta, \sigma) P(\sigma) P(\theta) d\theta d\sigma. \quad (3)$$

- More interesting is the aggregated posterior distribution of the number of years covered by an ice core $P(\sum T_i|\mathbf{Y})$ which is calculated from (3) using the central limit theorem.
- The effect of different plausible prior distributions on the posterior distributions $P(\mathbf{T}|\mathbf{Y})$ and $P(\sum T_i|\mathbf{Y})$.

RESULTS

Data

As an example the Dyer core shall be used, but general statements apply to other cores analogously. The Dyer core is located at 70°39'S, 65°01'W in Antarctica and contains roughly monthly δ^{18}O measurements. It covers approximately 140 years starting in 1985 and was dated by identifying minima in the δ^{18}O signal as annual winters over the past 75 years. (The data was kindly provided by the BAS.)

Prior Uncertainty in Accumulation and the Error

In order to make the calculation of $P(\sum T_i|\mathbf{Y})$ analytically tractable and to describe a range of possible prior assumptions on the log-accumulation rates A_i, we tested different normal distributions as well as an improper uniform distribution. The respective distributions for the accumulation rates e^{A_i} are depicted in Figure 1 (a).

For the error term, different Gamma and log-normal distributions for σ were analysed to cover a plausible range of prior assumptions (see Figure 1 (b)).

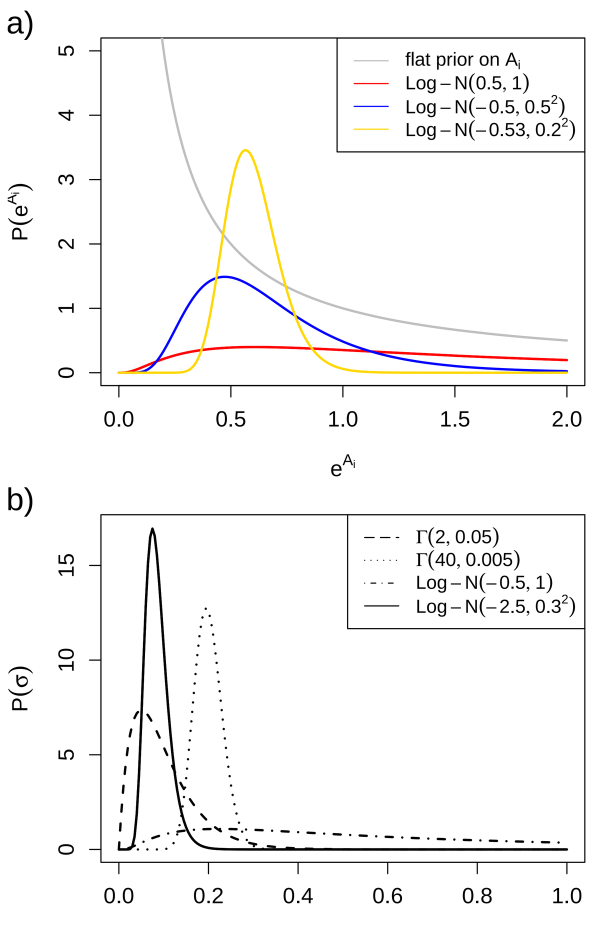


Figure 1: Plausible prior assumptions (a) on the accumulation rates e^{A_i} and (b) on the error parameter σ.

Posterior Distribution of the Number of Years Covered

The effects of these particular prior distributions on the posterior distribution of the time covered by the core, given the current measurements, are depicted in Figures 2 and 3. In particular, the coloured lines for the posterior distributions $P(\sum T_i|\mathbf{Y})$ in Figure 2 were calculated by using the prior distributions on the accumulation rates of the same colour in Figure 1 (a) together with the log-normal prior distribution $\sigma \sim$Log-N$(-2.5, 0.3^2)$ (depicted in Figure 1 (b)). Similarly, the different line types of the posterior distributions $P(\sum T_i|\mathbf{Y})$ in Figure 3 were calculated by using the corresponding prior distributions on σ in Figure 1 (b) together with the log-normal prior distribution Log-N$(-0.7, 0.2^2)$ on the accumulation rates (depicted in Figure 1 (a)).

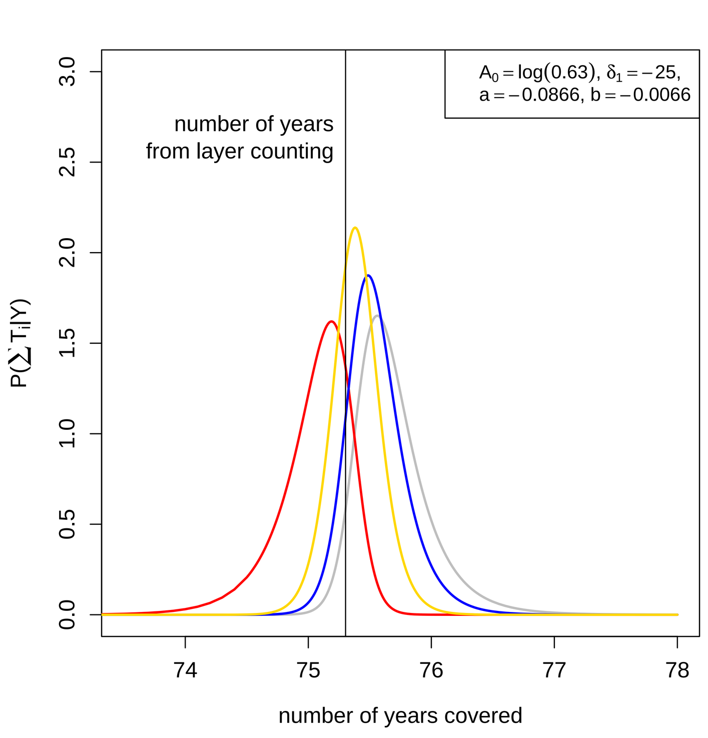


Figure 2: Approximate posterior distributions for the number of years covered by the Dyer core given the δ^{18}O measurements and assuming the prior distribution coloured correspondingly in Figure 1 (a) together with the solid-lined prior in Figure 1 (b).

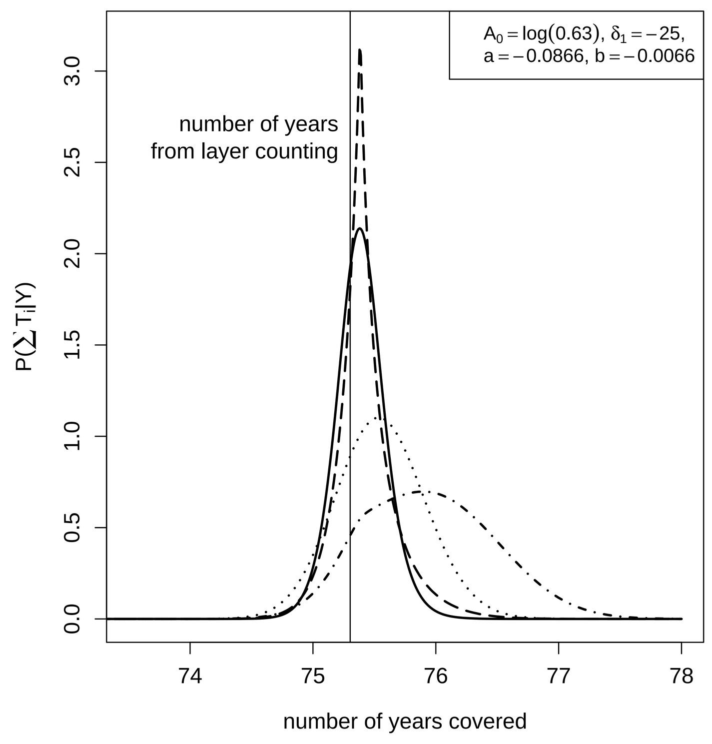


Figure 3: Approximate posterior distributions for the number of years covered by the Dyer core given the δ^{18}O measurements and assuming the prior distribution of corresponding line type in Figure 1 (b) together with the yellow-coloured prior in Figure 1 (a).

Generally, the effects can be summarised as follows:

- Assuming small variance and a 'correct' mean for the prior distribution on accumulation rates results in a narrow, normal-like posterior distribution $P(\sum T_i|\mathbf{Y})$ around $\sum_i \frac{d_{i+1} - d_i}{e^{f(Y_i)}}$.
- When the prior distribution concentrates considerable weight at low accumulation rates, the posterior $P(\sum T_i|\mathbf{Y})$ has a fat tail towards a larger number of years covered.
- Concentrating weight at high accumulation rates gives a fat tail towards a smaller number of years.
- Assuming large variance for the prior distribution on accumulation rates gives a wider posterior.
- When the prior distribution on σ concentrates considerable weight close to zero, the posterior distribution $P(\sum T_i|\mathbf{Y})$ is very peaked.
- Assuming that weight in the prior distribution concentrates at larger values of σ, results in a wider posterior $P(\sum T_i|\mathbf{Y})$.
- The larger the prior variance we allow on σ, the more positively skewed the posterior distribution $P(\sum T_i|\mathbf{Y})$ becomes.

Additionally, these results show the magnitude of the effect of uncertainty in the relationship f and in the accumulation rates e^{A_i}, and underline the importance of properly assessing these elements to make sensible statements about the dating itself as well as its uncertainty.

Effects of Uncertainty in the Parameters of f

The effect of different prior distributions for the parameters of the relation f on the posterior distribution of the number of years covered is multifaceted. Let us examine uncertainty in A_0 (representing present-day log-accumulation rate) and in δ_1 (representing present-day δ^{18}O value):

- The examined prior distributions for e^{A_0} and δ_1 (normal distributions with standard deviations of 1–10%) lead to uncertainty in the dating of an ice core which is one order of magnitude larger than the effect of uncertainty in accumulation rate e^{A_i} and error variance σ^2 examined above.
- If the posterior distribution of the number of years covered is relatively peaked for a function f with fixed parameters, then uncertainty identified in
  - the parameter A_0 directly relates to uncertainty in the posterior. In particular, if the present-day accumulation rate e^{A_0} is uncertain with standard deviation $x\%$, then the posterior distribution is uncertain with standard deviation close to $x\%$.
  - the parameter δ_1 relates to uncertainty in the posterior via the remaining parameters of f. For the examined setting, a prior distribution for δ_1 with a standard deviation of x results in a posterior distribution with standard deviation of about $100c_1x\%$ (as long as x is small), where $a = c_1 - \delta_1 b$.
- If the prior distribution for the error variance σ concentrates weight at large values, the effect of uncertainty in the parameters of f on uncertainty in the posterior distribution of the number of years covered is less pronounced, compared to weight concentrated closer to $\sigma = 0$.

DISCUSSION AND FUTURE RESEARCH

The current research has examined the effect of candidate prior distributions for uncertain quantities of a glaciological model on the posterior distribution of the number of years covered by an ice core. Initially simplifying assumptions had to be made. Future research will use recent snow accumulation data from Automatic Weather Stations as well as recent δ^{18}O measurements of snow to

- revise assumption (1) of independent accumulation rates and assumption (2) on the structure of the error and develop more realistic assumptions, and
- substitute each candidate prior distribution by an informative prior distribution.

In order to date ice cores of increased depths, the deformation of ice under the strain of ice which accumulated above it has to be taken into account. Future research will therefore modify the relationship between accumulation and time to $T = \frac{t(D)}{e^{\mathbf{A}}}$ in order to take ice sheet thinning into account. Glaciological models can be supported by other dated records. Uncertainty in the posterior distribution of the number of years covered can thus be reduced by including extra prior knowledge on e.g. volcanic eruptions which leave traces in the ice core.

Contact: Dr. Katy Klauenberg; mail to: k.klauenberg@sheffield.ac.uk; Tel: +44 114 222 3824
Department of Probability and Statistics, University of Sheffield, Hicks Building, Sheffield, S3 7RH, UK

References: [1] Johnsen, S. J., Dahl-Jensen, D., Dansgaard, W., and Gundestrup, N. (1995). Greenland palaeotemperatures derived from GRIP bore hole temperature and ice core isotope profiles. *Tellus B*, 47(5):624–629.