A sequential Bayesian approach for the estimation of the age-depth relationship of Dome Fuji ice core

Disclaimer! The following review deals with the statistical aspects of the article, not on geophysics nor glaciology.

Summary The article presents a new model for the relationship between the age and the depth of the Dome Fuji ice core. The model is a state space model, with observations $(y_z)_{z\geq 1}$ and hidden states $(x_z)_{z\geq 0}$. The index z represents the depth (surface is $z = 0$). The hidden states are made of the age ξ_z (in years, the variable of interest) and of the annual rate A_z of accumulation of snow, at depth z. Conditional on a state $x_z = (\xi_z, A_z)$, the next state $x_{z+1} = (\xi_{z+1}, A_{z+1})$ is given by a non-linear transformation of x_z plus some random noise (Gaussian and log-Gaussian). This transition distribution is parametrized by θ, assumed unknown. Inference about the states $(x_z)_{z\geq 0}$ is performed using observations $(y_z)_{z\geq 1}$, through the measurement distributions $p(y_z \mid x_z, \theta)$ assumed Gaussians, also parametrized by θ. The model deals with missing data for some depths z. Joint inference of θ and $(x_z)_{z\geq 0}$ given $(y_z)_{z\geq 1}$ is achieved using particle Markov chain Monte Carlo methods, which constitute the state of the art in the context of non-linear state space models. Numerical results show the approximated marginal distributions of the parameter θ given $y_{1:z}$, as well as the inference on the hidden states $(x_z)_{z\geq 0}$ given $(y_z)_{z\geq 1}$ under parameter uncertainty, i.e. the distribution $p(x_{0:z} \mid y_{1:z}) = \int p(x_{0:z} \mid y_{1:z}, \theta)p(\theta \mid y_{1:z})d\theta$.

General comments The proposed model seems novel. The formulation as a generic state space model (referred to as sequential Bayesian model in the article, see remark below) is powerful, in that it readily allows various modifications of the model to be envisioned: other forms of the thinning factor Θ_z, or of the accumulation rate A_z, or of the distributions for the noises ν_z, η_z, ε_z, w_z. The inference method described in the article (particle MCMC) would be able to deal with such modifications. Therefore, the article proposes a rather generic and scalable approach. Note that Bayesian methods and Monte Carlo approaches have already been applied to the context of glaciology, according to a quick google search (see [4, 2]), but state space models seem to have been so far restricted to linear Gaussian models (and thus inferred with Kalman filters, see [5]). Thus, the present article constitute a significant step forward, catching up with modern statistical methodology.

The description of the model is fairly clear and self-contained (although see remarks below). The article successfully demonstrates that the method can be applied to obtain various outputs from the model: the posterior distribution of the parameters, and estimation of the age-depth relationship under parameter uncertainty.

A slight concern about the paper: little validation of the results is presented. Some of the marginal distributions of the parameters are compared to results previously obtained in the literature, noting that some are consistent with the literature, and some are different, with the comment (page 954, lines 17-19):

> *It should be noted that these two results were based on different modeling for the accumulation rate, and thus it should not be expected that they would necessarily provide similar results.*

Indeed different results are to be expected, but it would be good to think of some criterion to evaluate the performance of the models in an objective way. In which way is the proposed model better than previous ones? External validation of the results would perhaps be a convincing way to promote the approach, and here are a few things that come to mind.

- The prior distribution on the parameters should be very explicitly described (all the results depend on it). Then, the posterior distribution could be compared to the prior, for instance using overlaid kernel density plots. This would give a visualization of how much information is gathered on the parameters from the data. Perhaps some parameters are easier to estimate than others?

- A simulation study on synthetic data generated from the model would also be informative. How precisely can we identify the model parameters using synthetic dataset (using the same number of observations as in the real dataset)?

- The model quality could be evaluated based on its predictive performance: for instance, the parameters could be inferred using the first 80% data points, and the remaining 20% data points could be predicted. Certainly other criteria could be envisioned, such as Bayes factors. In fact the statistical literature is quite rich on this topic (see [6]).

- Some aspects of the model seem more arbitrary than others: Gaussian distributions for the noise distributions, the accumulation rate is a random walk process (why not an autoregressive process?). The article could either justify these modelling assumptions in more details, or test various model modifications in practice. The resulting models could be compared, again, using predictive criteria or Bayes factors.

To conclude, the article proposes an interesting model with an application to a real dataset, and rely on modern computational methods to perform inference. The prior distribution should be written explicitly, along with enough implementation details for the results to be reproducible (see some remarks below). Some objective way of assessing the quality of the model would be welcome. Other than that, my remarks are just suggestions.

Remarks

- The language is clear. The general descriptions of the model and of the methods are fine, but the article should allow readers to reproduce the results; it is not the case here by lack of implementation details (lack of details on the prior distribution), details on the proposal distribution $q(\theta' \mid \theta)$, etc). Perhaps an appendix could give all the values used in the implementation that are not specified in the main text.

- Inconsistent notation: $\delta^{18}O$ or $\delta^{18}O_z$ or $\delta^{18}O(z)$ data.

- State space models are called "sequential Bayesian models" in the article, which is non-standard and a bit misleading, because nothing is really "Bayesian" about them (Bayes formula is just used to obtain the recursion formula for the filtering distributions). "Bayesian" usually refers to inference methods treating parameters as random variables, and does not refer to models. Hence, non-Bayesian approaches could have been applied to the model of the article. Another common term for state space models is "hidden Markov models".

- The model description is split into Section 2 & 3, starting in "continuous time" and with the description of Θ_z (Section 2), and then switches to discrete time and to the description of A_z and of the measurement distributions (Section 3). These sections could perhaps be combined in one section.

- page 945: Equation (14) should read

$$p(\xi_{z+1} \mid \xi_z, A_z, \theta) = \mathcal{N}\left(\xi_z + \frac{1}{A_z \Theta_z}, \sigma_\nu^2\right),$$

  according to Equation (12)...?

- page 946: when there are multiple observations $\delta^{18}O$ within an interval of one meter, a mean is used (presumably, without modifying the standard deviation σ_w^2). This seems unfair, as when there are more observations, the uncertainty should be reduced. One simple approach would be to use the mean of the observations at each meter, but with a variance σ_w^2 divided by the number of observations.

- Again, the prior distribution on the parameter θ should *absolutely* be specified somewhere.

- page 948: why isn't σ_ε included in the parameter θ? More details should be given on this. Does the method fail if this parameter was included in θ? What are the values given to it, in the end?

- In the proposed model, the transition is non-linear (because of the term $1/A_z\Theta_z$) but the noise distributions are Gaussian (if we use the parametrization $x_z = (\xi_z, \log A_z)$ instead of $x_z = (\xi_z, A_z)$). Thus, a "locally optimal" particle filter approach could be implemented, that is, instead of propagating the particles using $p(x_{z+1} \mid x_z, \theta)$ and weighting using $p(y_{z+1} \mid x_{z+1}, \theta)$, one could sample from $p(x_{z+1} \mid x_z, y_{z+1}, \theta)$ and weight the particles using $p(y_{z+1} \mid x_z, \theta)$; these two distributions are Gaussian. This is called the optimal proposal scheme in [1]; it could reduce the variance of the likelihood estimator.

- page 951, line 11: "using the SMC" + algorithm ?

- page 952, line ~5: perhaps give the formula for the likelihood estimator, since it is quite central in the particle MCMC method?

- page 952, not much details is given on the tuning of the proposal distribution $q(\theta' \mid \theta)$. How is the variance tuned? Using preliminary runs?

- page 952, Equation (35) and onwards: it is not very clear that only the likelihood estimator $\hat{p}(y_{1:Z} \mid \theta^\star)$ on the numerator is calculated at each step, and that the one in the denominator is kept fixed. The method would not be valid if both the numerator and the denominator estimators were drawn at each step.

- page 953, line 13: "this greatly reduces the computational cost": does this refer to the memory cost instead of the computational cost? Is the memory cost a problem here?

- page 953, line 13: "$p(y_{1:k} \mid \theta)$" should be $p(y_{1:z} \mid \theta)$?

- End of section 4: perhaps mention other particle MCMC methods. In particular, some variations such as particle Gibbs, and particle Gibbs with ancestor sampling (see [3]), would be applicable here and could significantly improve the performance.

- Section 5: some comments could be made on the correlations between the components of the posterior distribution. If they are not close to zero, perhaps some pairwise scatter plots would be informative.

- Section 5: some indication that the Markov chains have mixed would be appreciated, for instance using traceplots instead of histograms. There is no indication in the text that multiple chains, with the same tuning parameters and starting from various points, lead to similar results. By the way, how were the Markov chains initialized? And how long was the burn-in period? Why was a sample kept every fifth iteration and not at every iteration?

- On the dataset: how large is it? Can it be plotted in some way? Can it be downloaded somewhere? It seems that the maximum depth is $Z = 2,500m$, and that there are a few dozen age markers (from Figure 2); it should be described in the text.

- page 956, on the computational cost: there should be some mentions of parallel computing, which could make 250,000 iterations with 5,000 particles much faster to run than 1,250,000 iterations with 1,000 particles. There is a rich literature on how to implement particle filters on parallel computing hardware.

- Figure 7-9 could be replaced by traceplots of the chains, starting from a few initial points, and plots of the average acceptance rates against number of particles, for a fixed proposal $q(\theta' \mid \theta)$.

References

[1] Doucet, A., & Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering*, 12, 656-704.

[2] Eckert, N., Baya, H., Thibert, E., & Vincent, C. (2011). Extracting the temporal signal from a winter and summer mass-balance series: application to a six-decade record at Glacier de Sarennes, French Alps. *Journal of Glaciology*, 57(201), 134-150.

[3] Lindsten, F., Jordan, M. I., & Schön, T. B. (2014). Particle Gibbs with ancestor sampling. *The Journal of Machine Learning Research*, 15(1), 2145-2184.

[4] Rougier, J. (2013). Intractable and unsolved: some thoughts on statistical data assimilation with uncertain static parameters, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **371**(1991).

[5] Trudinger, C. M., Enting, I. G. & Rayner, P. J. (2002). Kalman filter analysis of ice core data, 1. Method development and testing the statistics. *J. Geophys. Res.*, 107(D20), 4422.

[6] Vehtari, A., & Ojanen, J. (2012). A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, 6, 142-228.