EGU Sphere, Eichenseer et al., (2025):

"StratoBayes: A Bayesian method for automated stratigraphic correlation and age modelling"

Review by Andrew Curtis, University of Edinburgh

This manuscript proposes a Monte Carlo based method to assess quantitative Bayesian uncertainties in the statigraphic correlation between data transects recorded at different locations. It applies the method to both synthetic and real data, to highlight strengths and weaknesses of the method.

Overall I feel that this work will make a significant advance over manual correlation methods and some other algorithmic methods. I nevertheless have a few comments which all concern the methodology rather than the applications.

Main comments:

1. In principle, a hiatus may occur between any two measurement locations in a geochemical transect (Sadler, 1981), as the authors acknowledge in the discussion at the top of page 29. For example, hiatuses certainly occur in age records at grain scale, in any grainy depositional setting. The reason that age correlation still has some validity derives from an assumption that these types of hiatus are insignificant, or can be represented in aggregate as smooth increases in age, when averaged over the time scales at which data sets are correlated.

Nevertheless, when we try to assess uncertainty in correlations, any possibility that hiatuses exist which are significant even at these longer time scales should produce a variation in the resulting age curves along the measurement axis or transect, and so should be considered in uncertainty estimates. In a Bayesian context, information (beliefs or constraints) on inter-datum hiatuses should therefore ideally be expressed explicitly as so-called prior information, embodied in prior probability distributions. Such prior information may come from, for example, the fact that at some scale of observation we have / have not observed a sedimentological sign of hiatus in the rock record along the transect or in neighbouring synchronously deposited sediments, a sedimentary process or statistical model that embodies such beliefs, or any other type of pertinent observation. In the current algorithm, the uncertainty in these sources of information is not included – the locations of hiatuses are defined definitively *a priori*. The authors do acknowledge the possibility that hiatuses may occur anywhere on page 29, but they leave that problem unsolved.

I feel that it is important to make this prior probability distribution (the *probability* that a hiatus exists at any point in the transect) explicit in, what the authors propose is, a general correlation method: this will force practitioners to consider properly the *quality* of information, and the *degree* of expert belief or received opinion, that are fed into the process. Ideally all of the number, locations/age and length of hiatuses would be subject to uncertainties, and practitioners would therefore be forced to decide not only whether/where they can definitively fix a hiatus, but also where they might have missed one or more, and how likely that was to have happened.

I would also admit that translating geological observations into prior probabilities that can be incorporated into the Bayesian inference is not at all straightforward! It requires a process of expert elicitation, which is itself a significant challenge and source of epistemic uncertainty. Polson and Curtis (2010), Bond et al., (2012), Curtis (2012) and Bond (2015) describe some of the biases that affect geological interpretations by experts, and techniques that might be adopted to minimise these. While these papers are concerned with different types of observations, similar types of biases and uncertainties can be expected in the cases described above – perhaps in addition to other biases that are particular to those cases.

2. This method, along with most others employed for stratigraphic correlation, contains the implicit assumption that signals in the data from each synchronous sedimentary package exhibit similar data values. However, this is clearly not always the case, for example when comparing records in shallow and deeper water settings, or any other settings in which one record is more prone to have missing sections than another. This is discussed by the authors at the top of p.29. One approach to address this is to incorporate explicit relationships between the patterns that one might expect in different contemporaneous settings, such as might be embodied in a geological process model (e.g., Bloem et al., 2024); the authors rightly say that the latter approach has not been tested on real data, but then leave this problem hanging, with no other suggested solution.

Yet this effect may not be minor, and similar to comment 1 above, might introduce significant epistemic uncertainty that is not currently accounted for in the authors' algorithm. If the authors aim to assess uncertainty then ideally they would think about how to embody, or at least make explicit, all sources of uncertainty that affect age correlations, so I suggest that some further discussion would be valuable, about how one might address this in future.

3. In a number of previous studies, each geochemical proxy data set is often scaled in magnitude, in order to better match one transect to another. Was this not done in the current study (or did I miss it)? Differences in depositional environment, as mentioned in comment 2, may result in signals having different magnitudes, so is there a need to include such a scaling (perhaps *a priori*) to match the magnitudes of signals from one transect to another?

4. There is almost no way that a geologist can assess suitable values, or even ranges of values, for some of the parameters employed in the authors' algorithm (e.g., lambda) *a priori* – without looking at *any* outputs of the process). It is therefore not possible to define the corresponding prior information. I would guess that anyone applying this algorithm will use a trial and error approach to vary such parameters, running the algorithm each time, looking at the results until they get a good result, where 'good' then becomes entirely subjective. The parameter is then in fact defined *a posteriori* in a pseudo-hierarchical way – but without ever defining its prior distribution. I think that an example of this is even given by the authors themselves, in lines 281 to 285. As a result, while this method looks Bayesian mathematically, in practical use I fear that it might not be.

How can the authors change or differently embody these parameters to provide geologists with an intuitive way to define them *a priori*? One possibility might be to use the inverse approach to define prior information from Curtis and Wood (2004) or Walker and Curtis (2014), but are there others? Generally, it seems to me that if this methodology is to make a significant impact, in making the quantification of uncertainty in correlations more objective, then more research (perhaps in other papers) and discussion (in this paper) is needed to develop structured methods to define the prior distributions; otherwise this method may well be used in a similarly subjective manner to manual correlation, and while the results will be quantitative, they could end up being little more objective and well defined than previous results.

5. Lines 228-232 indicate that uncertainty may not increase with distance from absolute age constraints. I agree with the previous reviewer that this seems to indicate a significant flaw in the methodology. It may

again be due to the particular implementation – perhaps the density of spline knots should increase with distance away from the absolute age constraints (although it is not clear how quickly).

A similar issue arises close to sequence boundaries, around which time tends to be compressed in the stratigraphic record. Spline knots might be more densely distributed around such boundaries, but again it is not clear how dense they should be. This is another case where defining prior information is difficult, and requires more study (similar to the comment above).

Minor Comments:

Line 130-131: This sentence needs some explanation; the main text should be understandable without having to read the Appendices.

Fig. 1 caption: as far as I can tell, both alpha and gamma are used in the main text before they are defined, other than in this figure caption.

REFERENCES

C. E. Bond. Uncertainty in structural interpretation: Lessons to be learnt, Journal of Structural Geology, 74, 2015, https://doi.org/10.1016/j.jsg.2015.03.003

C. E. Bond, R.J. Lunn, Z.K. Shipton, A.D. Lunn; What makes an expert effective at interpreting seismic images?. Geology 2012;; 40 (1): 75–78. doi: https://doi.org/10.1130/G32375.1

A. Curtis, 2012. The science of subjectivity. Geology. 40, pp. 95-96. doi: 10.1130/focus012012.1

Andrew Curtis, Hugo Bloem, Rachel Wood, Fred Toby Bowyer, Graham Anthony Shields, Ying Zhou, Mariana Yilales, Daniel Tetzlaff, 2025. Natural sampling and aliasing of marine geochemical signals. Scientific Reports, 15:760, DOI: 10.1038/s41598-024-84871-6

A. Curtis and R. Wood 2004. Optimal elicitation of probabilistic information from experts. In, Geological Prior Information, A. Curtis and R. Wood ed's. Geol. Soc. Lond. Special Publication, Vol. 239; pp. 127-145; DOI 10.1144/GSL.SP.2004.239.01.09

D. Polson and A. Curtis, 2010. Dynamics of uncertainty in geological interpretation. Journal of the Geological Society, London, Vol. 167, pp. 5-10. doi: 10.1144/0016-76492009-055

M. Walker and A. Curtis, 2014. Expert elicitation of geological spatial statistics using genetic algorithms. Geophys. J. Int., 198, pp.342–356, doi: 10.1093/gji/ggu132