

Interactive comment on “Multi-time scale data assimilation for atmosphere–ocean state estimates” by N. Steiger and G. Hakim

Anonymous Referee #3

Received and published: 26 September 2015

Steiger and Hakim show the application of data assimilation (DA) on different time scales for paleo climate reconstructions. The method is in general quite flexible and allows for the inclusion of different proxy types through the $H(x_b)$ function. I personally have great hopes in this method, similar to Tingley and Huyber's BARCAST (JClim 2010). However, while development on the latter has stopped at the annual time scale, the DA method of this contribution actually targets also low resolution proxies.

Some caveats (as I understand the method) do however apply: The calculations for eq. (2) and also hinted at in p. 3730 l. 13, and later in the text make it clear (to me) that multivariate normality of the joint distribution of the target x_b and the estimator for the observations $H(x_b)$ needs to be given. Otherwise the covariance matrix that enters the Kalman gain is a dangerous tool. As most proxy scientists select their records by

C1807

linear regression and simple Pearson correlation, this is not too much of a problem - unless one should target precipitation, for that x_b will be very non-normally distributed.

In general, I like the article and the results presented. I do however see some shortcomings: The high frequency proxies (annual proxies) are often criticized for having no (r not enough) variability on the decadal and longer time scale. That is, they are likely a high-pass filtered, noisy realisation of the climate variable(s). This method uses the raw, annual data and simply adds white noise to it. While this seems to be the current standard, it seems to be accepted that "real" noise does indeed change the spectral characteristics of the proxy.

Since the reconstruction is repeated several times, the authors do get an ensemble of (equally likely) reconstructions. In fact, the distributional nature of the method (they liken it to Bayesian inference and use the same vocabulary) should be used. Nowhere the uncertainties are shown or even discussed. Worse, still, the "measure" used to evaluate the "skill" of the reconstruction, the coefficient of efficiency, is not applicable. It is not a proper scoring rule and should not be used in the evaluation of probabilistic forecasts (or reconstructions), see Gneiting and Raftery (2007) for this and for possible alternatives.

I strongly urge the authors to respect the probabilistic nature of the method and 1) show uncertainty estimates where possible 2) use proper scoring rules for their method

Minor comments: I do like the use of the cross power spectra. The authors should still note that they do not contain different information than the cross correlation function, of which it merely is the fourier transform. The "noise" on the cross power spectra the authors refer to is mostly related to the finite sampling of the data. The (correct) use of Hamming, Welch or other filters does reduce this (if my memory is right).

Regarding the statement that no other methods have been able to target the different frequency domains in a spatially explicit single reconstruction, the authors are likely not familiar with the work of Joel Guiot (see ref. below), or (cannot check currently, no

C1808

access to it) the neural network method developed by Carro-Calvo.

Literature mentioned above: Carro-Calvo, Leopoldo, Sancho Salcedo-Sanz, and J. Luterbacher. "Neural computation in paleoclimatology: General methodology and a case study." *Neurocomputing* 113 (2013): 262-268.

Gneiting, T. and Raftery, A. E.: Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association*, 102, 359–378, 2007

Guiot J, Corona C, ESCARSEL members (2010) Growing Season Temperatures in Europe and Climate Forcings Over the Past 1400 Years. *PLoS ONE* 5(4): e9972. doi:10.1371/journal.pone.0009972

Interactive comment on *Clim. Past Discuss.*, 11, 3729, 2015.