**Climate
of the Past
Discussions**

Interactive comment on "Identification of climatic state with limited proxy data" *by* J. D. Annan and J. C. Hargreaves

Anonymous Referee #2

Received and published: 26 March 2012

This is a well-written and developed paper describing a pseudoproxy study investigating the application of a data assimilation approach using the particle filtering process. The results are well presented and logically discussed. It could be published almost as is, but I include a list of minor revisions and improvements below for the authors to consider.

Pg. 483, Ln. 10: Tingley et al. (QSR, 2012) should be included here as a recent and comprehensive methodological discussion. I believe the authors also mean Smerdon et al. (GRL, 2011). The Smerdon (WIRES, 2012) review would also be applicable here. Note that a general confusion between these latter two papers exists throughout the manuscript and the authors should correct the citations based on which ones they mean throughout.

Pg. 484, First Paragraph: Another drawback of data assimilation methods is the fact that the derived reconstructions cannot be used independently to assess model simulations. Inasmuch as this is an important motivation for proxy reconstructions, this drawback should be mentioned.

Pg. 484, Ln. 11: The Tingley and Huybers (JC, 2010) and Li et al. (JASA, 2010) papers should be cited here as formal examples of studies that have applied Bayesian methodologies to the paleo problem.

Section 2: More information could be provided on the model. It is well discussed in the literature, but a presentation of the relevant aspects of the model performance and configurations would be useful here. For instance, the horizontal resolution of the model and its leading spatial covariance patterns would be very helpful for interpreting later results.

Pg. 484, Ln. 20: Although it is not discussed beyond this point, the skill assessments are only over a century of LOVECLIM simulation. A useful characteristic of pseudoproxy studies is to extend the time scales of validation to multidecadal and centennial periods. The authors should more clearly state that their skill assessment is limited in its ability to robustly measure performance over these lower frequencies (which other studies have investigated).

Pg. 485, First Paragraph: The authors describe their focus on the NH and it appears that they have limited themselves to only NH proxies (even though they show global fields in some cases). It should be noted that there are some SH proxies in the Mann et al. (PNAS, 2008) screened database, but these presumably have been left out because of the focus on the NH. It should also be noted that teleconnections would justify using SH proxies, even in reconstructions that only target the NH. A clearer discussion of the authors' choices in this regard would be useful.

Pg. 485, Ln. 13: I think Figure 1 is discussed in text after Figure 2. Switch the order if this is true.

Pg. 489, First Paragraph: Discussions of the relative merit of skill statistics are common in much of the methodological statistical reconstruction literature. Some of the discussion here about correlation vs. RMSE parallels arguments about the RE and CE statistics. Ammann and Wahl (CC, 2007) is one such example, as is Christiansen et al. (JC, 2009) and subsequent exchanges. Smerdon et al. (GRL, 2011) also showed and reported multiple measures of merit to stress the importance of a more collective description of reconstruction performance. Some of this could be cited here. It is also worth considering the presentation of some additional skill fields later in the paper. The authors use normalized RMSE, but correlation fields, for instance, would be useful for further illustration.

Pg. 489, Second Paragraph: The lost of variance in the NH mean has also been the subject of much discussion. This now has been shown in multiple pseudoproxy studies including von Storch et al. (2004,2006), Smerdon and Kaplan (JC, 2007), Smerdon et al. (JC, 2008), Christiansen et al. (JC, 2009), Lee et al. (CD, 2008), Riedwyl et al. (CD, 2009), Ammann et al. (CP, 2010), Smerdon et al. (GRL, 2011), etc. There have also been several studies to show NH mean reconstructions that do not suffer from this variance loss: Mann et al. (JGR, 2007) and Hegerl et al. (JC, 2007). A more complete presentation of this discussion would be warranted here.

Pg. 489, Last Paragraph: The collapse of the ensemble is first discussed here and then appealed to throughout the manuscript as an explantation for the results in Figures 4 and 6. This discussion should be expanded because it will be counter intuitive to many readers. How can the results effectively get worse with more sampling? The authors should more clearly point out that this is fundamentally the result of the ensemble size being too small to constrain the true posterior estimate. In other words, more data restricts the number of ensemble members in a fixed ensemble that can actually constrain the posterior. It should be pointed out that a larger ensemble would better constrain the posterior, and the limiting case of an infinite ensemble would show improved skill for the increased number of predictors.

Pg. 492, Ln. 19: The dependence of the results on the specific spatiotemporal characteristics of the model is important. Although it has been suggested that NH or global mean results probably do not depend on these characteristics too heavily, the spatial performance of reconstruction methods is undoubtedly important in this regard. One of the few papers to discuss this in the spatial context is Smerdon et al. (GRL, 2011). Moreover, assimilation techniques like the one currently investigated are perhaps even more dependent on model formulations, because the assimilation process implicitly incorporates the model assumptions into the posterior generation. I would therefore encourage the authors to discuss this a bit more as a means of highlighting this important point moving forward.

Figs. 4 and 6: I am not particularly fond of the color scale choice for these figures. It is a bit hard to decipher the values of normalized RMSE over the range provided. The authors should consider an alternative with more color graduations.