

Interactive comment on “Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling” by Robert Tardif et al.

Anonymous Referee #2

Received and published: 18 November 2018

General comments: The study seeks to perform the climate reanalysis for the past millennium. Although the topic is of great interest to a broad readership of this journal, this reviewer believes the methodology, analysis and the final LMR product presented herein are too premature to be acceptable for formal publication, let alone for its stated purpose to serve as the basis for the first publicly released NOAA last millennium reanalysis. My specific comments are given below.

Major comments: 1. It is misleading for this study (and its prototype in H16) to call the DA method used in this as an ensemble Kalman filter (EnKF). As in Evensen (1994) and subsequent studies, the primary promise of the EnKF is the use of flow dependent background error covariance represented by the forecasting ensemble. The current so-called “offline” DA method has none of that: the ensemble perturbations are randomly

C1

sampled from a past-millennium climate simulation that has no relation to the prior estimate, and the same set of sampled perturbations were used at all analysis times. This method used in this study is similar to the commonly used 3D-Var method for numerical weather prediction with static background error covariance, and is arguably less advanced than 3D-Var since 3D-Var in NWP used the dynamic model to propagate the previous cycle’s analysis as the prior before the analysis. The current so-called “offline” DA method neither cycles the analysis nor the ensemble perturbations, with the stated reason that the forecast model is not good enough to do either.

2. If the forecast model is not good enough to cycle the mean analysis or the analysis uncertainties to provide the best estimate of the prior estimate and related prior uncertainties, why would this model(s) be good at all for use as the prior estimate that the LMR reanalysis depends critically on? In this regards, it is premature to state (line 10) that the “LMR employs the ensemble data assimilation to optimally blend the information from the proxies and the climate model data”. The current method is more like an objective analysis method.

3. It is not clear whether the authors are aware that the traditional static 3D-Var methods also derive the background covariance from an ensemble of perturbations, as is traditionally called “the NMC method” using the sampled forecast divergence between different lead times from many realizations. The Kalman filter update in this case is equivalent to the variational update using the 3D-Var algorithm, though again the 3D-Var in NWP cycles the analysis and forecast during data assimilation, which is the most basic function in combining the model and data.

4. The validation performed in this study for the prototype and updated LMR “reanalysis” with several existing 20th-century reanalysis is misleading at best. The quality of the LMR reanalysis for the 20th century is the least issue given the availability of the modern much more advanced reanalysis and given the exponentially increased number of proxies or model instrumental observations. The validation currently focuses exclusively on the 20th century says little on the quality and performance of the LMR

C2

products, in particular over the early period when the proxy data are scarce. A more appropriate validation can potentially be done in two objective methods: (1) perform the 20th century “reanalysis” through thinning the observation density and maybe also degrading the observation accuracy to those representation of different periods of the past millennium; and/or (2) performing observing system experiments in which a certain number of observations are not assimilated but reserved for independent validation (or all of them in cross validation).

5. The use of a 2,5000-km covariance localization is highly questionable for the use of a 100 sets of fixed ensemble perturbations. At midlatitudes, this is amount to the observation impacts across the entire global latitude belt. The use of a fixed set of 100 sample perturbations also means a high rank deficiency over such a large area with this large localization distance.

6. On a related note, the current final LMR reanalysis derives from the mean of 51 such 100-member analyses, should it be the same if the 5100 samples of perturbations are used simultaneously in the Kalman filter update given the Kalman filter used is largely a linear operation? How much is the result sensitive to the choice of this arbitrary number of sample perturbations? It is also worth noting the the NMC method used for 3D-Var uses singular value decomposition to make it full rank. Such a approach is different from (and likely more advantageous over) the current Kalman filter update using purely non-envolving static ensemble covariances.

7. More generally, it is unclear what is the purpose of such as hastily done LMR reanalysis products with such ad-hoc DA approaches and the not-good-enough forecast models? The so-derived climate trend is almost certainly depending too much on the climate models used as a prior and ensemble sampled perturbations (and maybe the assumed climate forcings used in these models), as well as the density of observations over different periods. It could do more harm if such a premature reanalysis product is used or misused and if it were publicly released through NOAA, unfortunately. A more careful vetting of the products, and a more concerned effort in refined DA methodology

C3

are warranted before NOAA sanctioned such a product as reanalysis, in this reviewer’s opinion.

Interactive comment on Clim. Past Discuss., <https://doi.org/10.5194/cp-2018-120>, 2018.

C4