The study uses one metric to evaluate the quality of the reconstruction methods : the correlation between observed and reconstructed index over a test period. However, other properties of the reconstructed indices may also be relevant, for instance, the variance. Many regression-based reconstruction methods underestimate past variability. This can be illustrated in a simple one-dimensional set up. Considering one proxy record P that reacts to variations of the NAO index:

$P(t) = \alpha NAO(t) + \varepsilon(t)$

where ε is random noise.

A simple, but widely used, reconstruction method is the statistical regression model:

$\widehat{NAO}(t) = \beta P(t) + \eta(t)$

where η represents the variability not captured by the regression model. Using Ordinary Least Squares regression to estimate β leads to underestimation of the true value of β and, therefore, of the true NAO variance (see for instance Isobe et al 1990 Linear regression in astronomy for a review of different regression flavours and their properties).

This problem may or not be present in the methods used in this study. It would be useful if the authors could report in Table 4 also the variance of the reconstructed NAO index in the test period wrt. to the observations and also the variance of the reconstructed index over the full period.

We thank Eduardo Zorita for this constructive and useful comment. In this study we are performing multiple ensemble reconstructions, thereby bringing this result tedious to be presented efficiently for every reconstructions. We decided to add a table and a figure giving this result for the best reconstructions from each method (i.e. the reconstructions presented in figure 11). The variance of the reconstructions is presented for the whole instrumental period, the testing period, the training period, the full reconstruction period and its portion before instrumental observations of the Jones et al. (1997) NAO index (the years before 1856 being excluded). We also add discussions in the main text of the manuscript about this well-known problem in paleoclimate reconstructions.

Also, it would be informative if the time series in figure 11 were not normalized to unit variance (?), but showed the actual reconstructed variability.

Normalizing to unit variance is a useful way to easily quantify NAO variability using standard deviations as unit. Nonetheless, as Eduardo Zorita is mentioning, it is actually hiding important informations about the reconstruction we performed. Thus, we decided to modify figure 11 in order to keep the actual reconstructed variability by our code. +1 and -1 standard deviation levels for each reconstruction have also been added in this figure in addition of their regression uncertainties (see response to 1.3 comment of Anonymous reviewer 2).