This paper presents new reconstructions methods and applies them to reconstruct the NAO using data primarily from the PAGES2k database. I think this is a good study that introduces some potentially useful new paleoclimate reconstruction methodologies.
We thank the reviewer for this overall positive evaluation of our work.

I have a number of comments, corrections, and requests for clarification below:

p.1 l.7-9, p.4 l.18, p.20 l.10 These statements are too strongly worded. Not every mode of variability is reconstructable, some occur on too short of time scales to be captured in the paleoclimate record (e.g., monthly time scales) and some modes are in locations where there are poor covariances with available proxy records (e.g., the Southern Ocean).
We agree with the reviewer that this claim was too strong. This statement is modified in the corrected manuscript to clarify that our method is not able to reconstruct every climate index but only the ones for which sufficient covariances between large-scale modes and proxy records are found and for which proxy records exhibit fine enough time resolution to resolve the main time scale of the considered variability mode. Furthermore, we will also highlight that our approach can be used to reconstruct other kind of climate variable time-series such as temperatures or precipitations for a given location.

p.2 l.9-11 This sentence is unclearly worded, for example, "non-stationary variability" doesn't "ask" questions, people ask questions.
We agree with the reviewer on this statement. We replaced "asks the questions of" by "highlights".

Introduction: In general, the introduction takes a long time to get to the main points of the study. The authors might consider revising the introduction to cut down the length.
The introduction has been largely cut down by only keeping the most important informations relative to the topic of the manuscript.

p.5 l.4-5 Linear interpolation of low resolution proxies artificially increases the influence of these records and introduces spectral artifacts in the proxy time series (e.g., Hanhijarvi, Tingley, Korhola 2013, doi: 10.1007/s00382-013-1701-4). This process also ignores dating uncertainty in such low-resolution proxies, which can be a significant source of reconstruction error. Have you accounted for these factors, particularly the dating uncertainty? What is the influence of using only annually resolved data?
Reviewer 2 also highlights this issue. He also highlights that the database of proxy records that we use (the 2014 version of the Pages 2k database plus 69 additional proxy records) has been recently updated in 2017. Following this comment we have updated our code, manuscript and data with the use of the 2017 version of the Pages 2k database. Then, using this new proxy database, and in order to address this comment, we decided to remove the proxy records that are not annually resolved. Indeed, we found that using interpolated low resolution proxy records results in overestimating their weights in our reconstruction because

of the falsely high correlations they have with the NAO index. This is largely due to their respective high auto-correlations at the annual time-scale. Hence, as mentioned by the reviewer, using this kind of proxy record indeed brings a lot of reconstruction errors due to overestimated weights, dating uncertainties, but also, because they induce erroneous validation scores as the link between these proxy records and the NAO index is overestimated.
Concerning the dating uncertainty, it is also present in annually-resolved proxy records and this aspect is not accounted for in the present version of the reconstruction toolbox. Nevertheless, this is certainly something to be considered in the next version of the code. We thus add a short discussion on this aspect in the discussion section, concerning potential outlooks for the next versions.

Section 2.2 Do the methods estimate uncertainty in the reconstruction or just provide a single reconstruction? Are the ensembles of reconstructions discussed elsewhere a kind of uncertainty estimate of the mean reconstruction? These, or something like them, would be essential to use and display because without reliable uncertainty estimates, paleoclimate reconstructions are not useful.
This was actually a major omission in the former version of the paper and we thank the reviewer to report it. The uncertainties we now provide are calculated as in Ortega et al. (2015) using the residuals calculated over the 50 training periods. These uncertainties are represented by the standard errors (s.e.) of the regression, calculated as the root of the sum of the squared residuals divided by the degree of freedom over the training periods divided by the degree of freedom:

$$s.e = \sqrt{\frac{\sum_{i=i}^{n_{train}} (Y_{train} - \widehat{Y}_{train})}{n_{train} - 2}}$$

Where n_{train} is the length of the training sample, Y_{train} the true values of the NAO index over the training period, and $\widehat{Y}_{train}$ the fitted NAO by the regression model over the training period.
An uncertainty band 2*s.e. is calculated for each of the 50 individual reconstructions and the envelope of this 2*s.e. uncertainty bands is our estimate of the total uncertainty range of the final reconstruction.
We added regression uncertainties in a table and on the figures where the reconstructions are shown. Also, the code we deliver provide standard errors for each member of a given final reconstruction.

p.7 l.16-19 Using correlation as the only validation metric is problematic, especially when it comes to comparing reconstruction methodologies. You really must include additional metrics that account not just for the correlation, but the variance and bias as well. If the

approaches provide uncertainty estimates, then the skill metrics need to also account for those (using, for example, the continuous ranked probability score).

This comment was also highlighted by the other reviewer as well as in the short comment of Eduardo Zorita. We totally agree with this comment and we decided to add both the root mean squared errors and the Nash-Sutcliffe Coefficient of Efficiency (NSCE) as additional metrics. The NSCE calculates the ratio of the averaged quadratic distance between the reconstruction and the observations and the quadratic distance between the mean of the observations and the observations. This metric, defined between $-\infty$ and 1 indicates that the reconstruction is robust when NSCE>0. Otherwise, lower values mean that using the mean of the testing series is more robust than performing a reconstruction using the statistical model.

We thus believe that these two metrics adequately account for the bias and variance in the reconstruction, which should then improve the conservation of these properties in our reconstruction. The whole new manuscript now accounts for these two metrics and use the NSCE as main decision metric.

p.16 l.19-20 This statement is incorrect. Previous reconstructions almost never overlook this issue, but rather proxy network selection is integral to the reconstruction process. It is very rare to have a reconstruction approach, especially one that is regression-based, that does not remove proxies because of insufficient correlation with the target climate variable.

For climate index reconstructions we found at least two major studies that have not used proxy network selection to perform their reconstruction : Cook et al 2002 (NAO reconstruction) and Wang et al 2017 (AMV reconstruction). For the latter, a table of the proxy records used is presented in supplementary information. According to this table, we found that this study has used proxy records with correlations close to 0 and non-significant between some of the proxy records and the targeted AMV index. Nevertheless, we indeed found that these studies are particular case and we modified this statement to clarify that we were referring mainly to these two studies.

p.18 l.1-2 Or the "significant" correlation with the NAO could be spurious. Also note that non-stationarity violates one of the fundamental assumptions of these (and nearly all) reconstruction approaches.

Indeed, we also ask ourselves if the significant correlations we found could be spurious but it is relatively difficult to determine whether they are or not. An indirect way to “verify” this significance of correlation is the location of the proxy records that have high correlations with the NAO. Indeed, the proxy records we use are located in the well-known center of actions of the NAO, which, in a sense, shows that the corresponding correlations are not fully spurious but may be related with well-known climatological fingerprints of the NAO (e.g. Casado et al. 2013). The second comment about non-stationarity indeed highlights a problem that not only questions our study, but also all of the proxy based reconstructions studies. We believe that this sentence was not at the right place in the submitted manuscript, since this type of caveat has to be included in the discussion section. This has been done in the revised version.

p.19 l.12-15 I think this statement is too strongly worded given that you've only validated the reconstructions using correlation and haven't validated reconstruction uncertainties. How do the reconstructions compare given the uncertainties?

As mentioned above, in the revised version we use the coefficient of efficiency to validate our reconstructions.