Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-117-RC2, 2020

Natural Hazards and Earth System Sciences Discussions
Open Access

Interactive comment on "Does the AO index have predictive power regarding extreme cold temperatures in Europe?" *by* Tamás Bódai and Torben Schmith

Anonymous Referee #2

Received and published: 23 June 2020

I have read with interest the paper : " Does the AO index have predictive power regarding extreme cold temperatures in Europe? ". The authors claim, not surprisingly, that what they define a " native " covariate (the Temperature) is better at forecasting extreme cold temperature in winter than a dynamical co-variate, i.e. the Arctic Oscillation index. Although I appreciate the use of the non-stationary GEV model for this kind of studies, that in my opinion this paper is yet too far from being publishable in NHESS. My recommendation is to reject the paper in its current form and I encourage the authors to completely rethink their study. The two most important arguments for rejection are the following :

Printer-friendly version

Discussion paper

1) The results of the paper are trivial : a good representation of cold extreme temperature distribution in winter is obtained if a model which include temperature is used. Of course, another index that does not include temperature information performs poorly. How is this results useful at all for the community of seasonal forecast ? Why do the authors analyse only the AO index, instead of focusing on SST, NAO, or an index based on weather regimes computations ?

2) In the paper, the authors fail to evaluate what in the title and the abstract they announce as " seasonal predicting power " of their model. They limit their analysis to fits of distribution and tests for distributions (Lilliefors and Chi2). Standard forecast skills metrics are not applied (e.g. CRPS, RMSE). The scientific question claimed in the title, about predictive power, should be answered by : i) training the model on just a part of data (training data set), 2) testing the quality of the forecast in a testing data set (the remaining of the data set). If they really want to say something about seasonal forecasts, then the question is how to evaluate the forecast made for DJF if when the model is initialized in November or October. They should then try to run the model over ensembles of possible AO and T values and finally get a CRPS (or equivalent foreacst skills metrics).

Other major comments :

1) The abstract is totally non-informative. I would recommend to review the abstract with the following suggestions : - " extreme value statistics " of what ? Are the authors talking about weather extreme events ? Which ones ? - " large scale quantities " ? Which scales ? Large with respect to what ? I think the authors mean synoptic scales here. And what are the " quantities " ? Are the authors referring to climate indices ? - " nonlocal " with respect to what ?

2) The presentation of the data is completely displaced : The authors evaluate M1 (model 1) in lines 80-100 before presenting the data sets used (lines 140-145) and even before presenting the evaluation metrics (section 2.2). Furthermore, the authors

mix up different time scales in their analysis : daily, monthly, seasonal. It is not clear whether the data for AO are daily or Monthly ? Do you use ECAD or EOBS ? What is the time scale of the model? Daily or Monthly ?

3) The evaluation of the model M1 is just qualitative. No CRPS or other standard forecast evaluation metrics are provided, or they appear later in the text, generating great confusion.

4) The authors recognize that seasonality and non-stationarity are important and should be taken into account. They say that their methodology " does not (and probably cannot) " take into account this issue. However they did not even try to apply the standard procedure to take into account this issue : e.g. repeating the analyses by removing a linear trend on the temperature, removing the seasonal cycle and repeating the analysis. These are very elementary tests and for such a simple basic analyses they should have been implemented.

5) Another point about the evaluation metrics : Lilliefors and chi2 tests are good to assess the adherence of distributions to the targeted ones. However, they do not say anything about the dynamics and therefore the predictive power of the model in terms of forecasts. I guess the authors are well aware that reshuffling the data destroys all the dynamical features (and therefore the predictability) but preserves the distribution so the results for both tests will be identical, even with a completely random dynamics.

6) The results section consist only of few lines of figure descriptions. No attempt of explaining the geographical differences on model performence is provided. Speculative sentences with no justifications appear here and there

Minor (but important) remarks

Introduction : -L9 : you cite Tel and Gruiz but the discovery of chaos in weather forecasts dates back to Lorenz 1963 -L12 : " ignoring completely some physical quantities " which ones ? References ? -L20-25 : in the GEV definition, mu is undefined, xi is undefined, only sigma is defined. These equations should be numbered.

Methodology :

LL80-100 : as said before, this evaluation is completely subjective and no quantitative analysis are provided. Note also that Yiou et al. 200 GRL, Ferranti et al. 2015 QJRMS, and Faranda et al. 2015 (Clim Dyn) have suggested that there is a non-trivial relationship between the AO patterns and time series and the predictability of the weather. They suggest that a simple time-series analysis should be discarded in favor of more comprehensive dynamical systems approaches.

LL105 " Seasonality: the different months of the winter should have different climatologies (Bódai and Tél, 2012). " I do not think this (auto)citation is pertinent here.

LL114-115 : " Furthermore, beside the climate-change-type nonstationarity, playing out on multidecadal time scales, there should be considerable internal variability on multidecadal time scales, too. ". this statement is too qualitative, how do the authors support it ?

Results :

LL171-176 Would the authors expect anything different ?

Discussion :

LL 220-222 I would be more careful and respectful citing the work of other colleagues as " ironic "

LL223-226 The authors mix up again (partially admitting) their confusion about predicting power, predictability, seasonal forecasts. . . I want to stress once again that distributions (and all the metrics associated) are totally insensitive too reshuffling. Metrics for predicting power and/or predictability are based on time evolution.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-

2020-117, 2020.