Comment on npg-2022-15
Tamas Bodai (Referee)

This paper is very nicely written, with abundant information on its subject.

I would like to raise a caution about taking it for granted that any ensemble-statistics (E-stat') (even with a hypothetical infinite initial condition ensemble of a model) represents a forced response of the model climate. The following lines of the paper seem to have no such concern:

Because the temporal sequences of internal variability unfold differently in the
96 various ensemble members once the memory of the initial conditions is lost, one can estimate the
97 forced component at each time step (at each location) by averaging the members together, provided
98 the ensemble size is sufficiently large. The internal component in each ensemble member is then
99 obtained as a residual from the ensemble-mean.

However, Gabor Drotos and I (see reference below) worked out the conditions when we can regard an E-stat' (change) a sound quantifier of climate (change). This is a conditional definition of climate and it requires — for one thing — a time scale separation bw. certain fast and slow processes. Such a sound conditional climate change, however, might not be entirely forced, but the evolution of the slow system could introduce an unforced component. Thus, the concepts of climate change and forced change decouple. The following lines from the paper could be interpreted in our sense (not considering the citations), but then they would contradict the above quotation (l65-99):
In some areas, climate trends driven by internal processes may even outweigh those due to anthropogenic influences over the past 30-60 years (Deser et al., 2012, 2016 and 2017; Wallace et al., 2013; Swart et al. 2015; Lehner et al. 2017).

On the subject of “...a larger ensemble may be needed for some aspects of the forced response than others” (l99-100), maybe you wish to cite my papers that advocate that cross-correlations of TWO quantities as opposed to some E-stat’ of ONE observable quantity tend to show much smaller SNR. So, forced changes of teleconnections can be difficult to detect even in an ensemble. Also, typically the higher the statistical momentum/quantile, the smaller the SNR.

I didn’t quite understand how the obs ensemble can help in re-assessing the detectability of trends in single observed realisations (paragraph starting with line 376). The model ensemble can have a bias in the forced trend because of model error (or changes of the slow system, as mentioned above, or numerical model drift, being an artefact). Isn’t the forced trend of the OBS ensemble the same as that of the model ensemble by construction, i.e., possibly biased? I don’t see a solution for this problem.

I553 this is the first time?

I581 “wet” and “dry” — check for the consistency of the directionality of double quotation marks

I605 combined the internal variability?

Note: I do not make recommendation to editors for or against publishing a paper. I selected “minor revision” only to be able to submit my review; please consider it void.

Tamas Bodai

References
(2022) https://doi.org/10.1002/essoar.10510833.2
