

Nonlin. Processes Geophys. Discuss., community comment CC1
<https://doi.org/10.5194/npg-2022-15-CC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on npg-2022-15

Gabor Drotos

Community comment on "The Role of Internal Variability in Regional Climate Change" by Clara Deser and Adam S. Phillips, Nonlin. Processes Geophys. Discuss., <https://doi.org/10.5194/npg-2022-15-CC1>, 2022

The preprint undoubtedly contains interesting and important information regarding the detectability and properties of regional climate change. The result that thermodynamic-residual trends match closely the corresponding ensemble-mean trends is particularly remarkable. Even though most of the conclusions are well established or are presumably sufficiently robust, there is a number of issues, as I see, that require correction or further thought before final publication.

The most important one is a factual error in the text, which has a direct implication for the interpretation of the results on the signal-to-noise ratio. In lines 362-365, it is stated that "For a normal distribution, a signal-to-noise ratio greater than two indicates that the ensemble-mean (forced) trend is significantly different from zero at the 95% confidence level: that is, there is less than a 5% chance that the ensemble-mean trend could have been a result of random internal variability." It is easy to demonstrate that this statement is wrong in the sense that the actual confidence level is higher under the assumption of a Gaussian distribution. The standard deviation of the trends, which appears in the denominator of the signal-to-noise ratio, is (at least approximately and apart from sampling uncertainty) uniquely determined by the so-called natural probability measure (see Drótos et al., 2015; Tél et al., 2020) and has a finite value, irrespective of the number of ensemble members used. On the other hand, by increasing the number of ensemble members, the ensemble mean of the trends can be determined with arbitrary precision: that is, the chance that a nonzero ensemble mean is obtained while the true expectation value of the trends is zero can be arbitrarily reduced. In the particular case when the ensemble mean happens to be twice the standard deviation, the chance that the true expectation value of the trends is zero can thus be less than 5%: actually, it can be arbitrarily small if the number of ensemble members is sufficiently large (irrespective of whether the trends are distributed according to a Gaussian).

In fact, if the number N of ensemble members is sufficiently large, then the sampling distribution of the ensemble mean of the trends (as an estimator of the true expectation value of the trends) is a Gaussian that is centered on the true expectation value and has a standard deviation that scales as $1/\sqrt{N}$, according to the central limit theorem. Under the null hypothesis that the true expectation value is zero, the task is to find the value above which (in an absolute sense) this Gaussian integrates to the desired significance level (one minus confidence level). For instance, the value sought is twice the standard deviation for a 5% significance level. If we furthermore assume that the parent

distribution (that of the trends observable in the individual ensemble members) is a Gaussian with a standard deviation estimated precisely by the sample standard deviation σ of the trends computed over the ensemble, the sampling distribution of the ensemble mean will have a standard deviation of σ/\sqrt{N} . In this case, the signal-to-noise ratio, which is associated with a given significance or confidence level and is defined by dividing the actual ensemble mean by the standard deviation of the sampling distribution, will be \sqrt{N} times higher than what is presented in the preprint. For the 100-member CESM2 LE, it means a 10-fold (!) increase with respect to the presented results. It should be emphasized, however, that the assumptions about the shape and standard deviation of the parent distribution may not at all be justified, so that the results shown in the preprint may only provide a qualitative guidance in the absence of a dedicated investigation.

I would continue my comment with an overarching issue regarding the terminology. Already the title suggests that internal variability can have an effect on climate change, and this is explicitly confirmed by the first sentence of the short summary, as well as referring to accuracy in its last sentence. As Tamás Bódai has already pointed out in RC1 of the discussion about the preprint, this notion is only meaningful if climate is defined to be conditional on the state of slower system components which have their "own" internal variability and can thus possibly induce unforced changes in the probability distribution that defines climate, in the sense discussed in Drótos and Bódai (2022). But 50-year trends, analyzed in the preprint, or those of similar length, can hardly be dominated by such changes; instead, they mostly originate from processes having decadal time scales and sufficiently rapid forced changes. Even if the effect of variations in slower system components is not negligible, it may (and hopefully does) remain unique for some time; in any case, the differences in the mentioned trends between the individual members of the ensemble are mostly due to faster processes of the Earth system. Therefore, these differences should definitely not be interpreted as differences in the pace of climate change, at least if the particular study targets the time scale of a century (still see Drótos and Bódai, 2022). On the contrary, if slower system components do not deteriorate uniqueness on the time scale in question, these differences should be regarded as an inherent property of a single (but changing) climate. As a consequence, writing about internally driven or non-unique "climate trends" in this context (lines 58 and 137), as if the pace of climate change were (substantially) dependent on the particular realization, would be safer to avoid (I use the word 'substantially' to refer to a potential non-unique effect of slower system components). The sentence in line 166 seems to be problematic from the same point of view; and even though they are widely used, the expressions "anomalous climate event" (line 62), "climate anomaly" (lines 182, 192 and 594) and "climate extreme" (line 590) appear to suffer from a similar conceptual issue. (These latter expressions sound as if climate could be anomalous or extreme at a given time within a single realization mostly due to internal variability --- this would only be meaningful if internal variability in slower processes, with time scales beyond the targeted one, induced these anomalies and extremes.) Also, I wouldn't advise writing that internal variability (substantially) "limits the accuracy of climate model projections" on time scales longer than a decade but not longer than a century [line 18; climate projections are usually meant to be "uninitialized" (section 11.1 of Kirtman et al., 2013) and thus fully encompass the statistics of the internal variability of the faster processes at least] or generates (substantial) "uncertainty" in them (lines 50 and 53); instead, internal variability (of the faster processes at least) represents an inherent property of climate and thus its projections.

A related remark is that the ensemble mean of the trends obtained in individual members is principally interesting for the purpose of comparison with instrumental observations (having an eye on detection and attribution). If the aim were to quantify the effects of forcing or slower system components on climate, I believe that it would be more useful to investigate the trends of the ensemble mean (or those of further statistical quantifiers evaluated with respect to the ensemble).

Having mentioned the possibility of unforced changes induced by slower system components, I would point out that climate can be easily defined only if these unforced changes remain unique during the time span of the study. As mentioned above, hopefully this is the case, but whether or not this is actually so, such unforced changes may appear in ensemble statistics with some weight, which is a problem already discussed by Tamás Bódai in RC1 and RC2. In such a case, variations in ensemble statistics do not entirely represent a forced response, as opposed to what is stated in line 97 and made use of throughout the text.

I list further substantial issues in the order as they appear in the preprint.

lines 97-98: The question of separating forced change and internal variability seems to be simplified here to determining the time evolution of the ensemble mean and taking the differences from the ensemble mean in individual ensemble members. However, internal variability is characterized by a full probability distribution the time evolution of which (as a forced response, or perhaps including an unforced component originating from slow processes, too) concerns all statistical quantifiers, as actually acknowledged in lines 102 and 108.

175-177: They are not only decadal shifts in regional anthropogenic aerosol emissions that violate the assumption of a slow forced change, but also greenhouse gas concentrations and solar activity can substantially vary on decadal time scales, and volcanic eruptions have an instantaneous and sometimes very strong impact.

207: It depends on the choice of variable if the memory of the initial state can become negligible by the time specified. There are system components, e.g., the deep ocean, for which the statement is not true.

215 and 554-556: The realizations of internal variability in the OBS LE were obtained in McKinnon and Deser (2018) under the assumption that the forced response is described by the CESM1 LE. In particular, the β coefficients and the ϵ residuals of Eq. (1) of McKinnon and Deser (2018) were obtained using ordinary least squares regression under this assumption. Therefore, it appears to me that the trends in the realizations of internal variability in the OBS LE are ensured to be consistent with observations only if the full trends are obtained by adding the forced trend of the CESM1 LE to the internal trend of each OBS LE member; using the forced trend of the CESM2 LE [or the CMIP5 ensemble, as in McKinnon and Deser (2018)] or the thermodynamic-residual trend (obtained from dynamical adjustment) for the same purpose might yield spurious results. This issue may affect the corresponding analyses throughout the preprint.

335-350 and Fig. 7: The respective results for σ could be interesting for the future (2022-2071) trends as well in the CESM2 LE, even if a comparison with observations is not possible.

355: "significantly" should be replaced by "significant". More importantly, the particular statistical test should be specified. Actually, it should also be demonstrated that the conditions for the applicability of the given test are met.

463: While the 5th-to-95th percentile range indeed narrows slightly, the 25th-to-75th percentile range appears to narrow even more in a relative sense, which would be worth mentioning in the text, I believe.

604: What is referred to as "the model's forced thermodynamic trend" is in fact not purely thermodynamic (by construction, it includes changes in circulation however minuscule they are), and its purely forced nature is also questionable (as discussed earlier in relation to the effect of slower system components). It would be a more cautious choice to simply

write "the model's ensemble mean trend".

Finally, I would mention two technicalities that could facilitate comprehension and reproducibility:

- It could be explicitly stated that panels (a) and (c) of Figs. 10 and 11 are identical to OBS and EM in Figs. 1 and 2, respectively, except that contours of SLP trends are also included.

- The serial number of the ensemble members used in Figs. 10b, 10d, 10f, 10h, 11b, 11d, 11f and 11h could be specified.

In spite of the several critical comments, I do believe that the results presented in the preprint are important and will be useful for future research.

References:

G. Drótos, T. Bódai and T. Tél (2015). "Probabilistic concepts in a changing climate: A snapshot attractor picture". *J. Climate* 28, 3275–3288. <https://doi.org/10.1175/JCLI-D-14-00459.1>

G. Drótos and T. Bódai (2022). "On defining climate by means of an ensemble". *ESSOAr* (preprint). <https://doi.org/10.1002/essoar.10510833.3> [Note the update on 2022-11-07.]

B. Kirtman, S.B. Power, J.A. Adedoyin, G.J. Boer, R. Bojariu, I. Camilloni, F.J. Doblas-Reyes, A.M. Fiore, M. Kimoto, G.A. Meehl, M. Prather, A. Sarr, C. Schär, R. Sutton, G.J. van Oldenborgh, G. Vecchi and H.J. Wang (2013). "Near-term Climate Change: Projections and Predictability". In: "Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change" [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter11_FINAL.pdf

T. Tél, T. Bódai, G. Drótos, T. Haszpra, M. Herein, B. Kaszás and M. Vincze (2020). "The theory of parallel climate realizations: A new framework of ensemble methods in a changing climate - an overview". *Journal of Statistical Physics* 179, 1496–1530. <https://doi.org/10.1007/s10955-019-02445-7>

K. A. McKinnon and C. Deser (2018). "Internal Variability and Regional Climate Trends in an Observational Large Ensemble". *Journal of Climate* 31, 6783-6802. <https://doi.org/10.1175/JCLI-D-17-0901.1>