Earth System
Dynamics
Discussions

Open Access

Interactive comment on "Statistical significance of rising and oscillatory trends in global ocean and land temperature in the past 160 years" *by* L. Østvand et al.

dr. Rypdal

kristoffer.rypdal@uit.no

Received and published: 8 August 2014

Response to comments by reviewer #3.

This response is organized as follows. The reviewer's comments are separated into three themes I-III and contains comments labeled (a)-(c). Under each theme we first list the reviewer's comments (a)-(c) and then our general response to the theme followed by responses to the points (a)-(c). Finally we respond to the list of review details.

Theme I: Novelty, physical insight, methods and structure

Reviewer:

(a) The paper does not offer sufficiently novel conclusions, or new insight into the Earth system.

(b) The detection of a trend is not recoverable without reference to the physical processes involved.

(c) The techniques seem unnecessary complex, the paper has a poor structure and many unnecessary digressions.

Response:

The paper deals with two issues that are still debated in public forums and in certain domains of the scientific community. The questions we want to address are:

Is it possible that (i) the rising global temperature during the instrumental era, and (ii) the apparent 70 year oscillation observed in these records, are part of a natural variability described by a simple stochastic process specified by parameters that can be derived from the short time-scale statistical properties of the observation data?

(a) In the mainstream climate science community (as reviewed in IPCC AR5, WG1, Chap. 10) the rising trend (i) has been attributed mainly to anthropogenic influence by multiple regression or "fingerprinting" techniques. The implications of considering LRM-processes as models for the natural variability are barely mentioned. On the other hand there are a considerable number of papers (as discussed in our Sect. 2.2) which perform trend testing under LRM null models, and with ambiguous results. In the literature the oscillatory trend has been attributed to astronomical forcing (Scafetta) or to specific internal modes like the AMO (Ramankutty et al.), but no attempt has been made to assess it's statistical significance under an LRM null hypothesis. Hence, we can't see that the reviewer's claim of lack of novelty is justified.

(b) In Sect. 2.2 we emphasize that absence of a significant trend (equivalent to acceptance of the null hypothesis) does not imply absence of a global warming signal. It just means that the particular record under study exhibits so large and auto-correlated natural variability that this signal is masked. This also implies that establishment of the global warming signal from physical modeling does not necessarily imply statistical significance of this signal. This statistical significance is of practical importance in it's own right. For ecosystems and human societies the existence of a global warming signal may (in the short term) be less relevant than its detectability in the LRM-noise, i.e., it's statistical significance.

(c) Common for the papers we have cited in Section 2.2 are that the details of their hypothesis testing techniques are not explained. In our paper we have tried to avoid using prefabricated statistical techniques and jargon. This means that it can be read by anyone with some knowledge about elementary probability. But not without some effort, of course, given the subtleties of hypothesis testing. Considering that we have built the framework from the bottom the techniques are not more complex than they have to be, and the "digressions" have emerged from questions and comments from individuals who have read early drafts. The reviewer offers few suggestions for improving the structure except a suggestion to move a relatively long paragraph on estimation of memory parameters to supplementary material. We are, however, receptive to suggestions that can improve accessibility and readability, and will therefore in the revision consider to move some of the more technical material to a separate section preceding the concluding section. Most of this material is essential for fully understanding the methodology and the results, so moving it to a supplement is not an alternative.

Theme II: Circular reasoning, rejection of null=acceptance of alternative

The reviewer makes some serious claims of inconsistencies in the basic logic of our approach:

(a) The analysis seems circular. Specifically by choosing an arbitrary fixed frequency f of the oscillation the hypothesis test simply gets out what is put in. The alternative hypothesis seems arbitrarily defined.

(b) The authors make a classic mistake in assuming that the rejection of the null is good evidence for the alternative hypothesis.

Response:

(a) The alternative hypothesis is by no means arbitrarily defined. The rising trend (as pointed out by the reviewer) is predicted as an outcome of anthropogenic forcing in e.g., the CMIP5 ensemble run with all forcings and natural forcing only (see AR5, WG1, Chapter 10). The 70 yr oscillation has been modeled in several papers (for references see Chapter 9 in the monograph of H. A. Dijkstra: Nonlinear Climate Dynamics, Cambridge, 2013). The reasoning would have been circular if we were testing for the frequency f. But we are not. We are testing for the linear trend coefficient A_1 and the oscillation amplitude A_2. Contrary to the reviewer's perception, our reasoning is really very simple: We observe that a smoothed version (e.g., by a 20 yr moving average) of the record is very well approximated by our trend model Eq. (3) with the coefficients estimated by a least square fit. As discussed above, this model is based on physical modeling and mechanisms that have been described in the literature as anthropogenic warming and AMO, respectively. The question we want to address is NOT whether this model with the estimated frequency and phase parameters is a correct model for the global temperature, but whether the null model is likely to produce a record that will return amplitude coefficients A_1 and A_2 at least as large as those estimated from the trend model. The null model should reflect the short-time scale correlation structure of the record, so its parameters are estimated by MLE, which emphasizes the short scales. In plain language; we test if our null model of natural variability can produce the rising trend and the oscillation which appear to be present in the data.

(b) From what we have written above it should be clear that we do not assume that

rejection of the null proves the alternative hypothesis in the sense that the alternative hypothesis is the only statistical model that gives an adequate description of the large-scale structure of the data. What rejection of the null proves is that the alternative model (which gives a good description of the large-scale structure) is preferred to the null model, but of course not necessarily to any other model. We agree with the reviewer that we should be more careful with our wording on this point.

Theme III: Alternative methods

The reviewer makes sketchy suggestions of approaches alternative to ours, such as:

(a) A supplementary approach to considering physical understanding might be to use a standard statistical model selection procedure penalising complexity, instead of simply prescribing a model structure with little or no justification.

(b) A good way to demonstrate the efficacy of the methods would be a blind test, where the time series might be a draw from a long memory process. . ..

(c) . . ..it would simplify matters if it is pointed out that the trend estimate is simply an extension of the stochastic model. In this case, the hypothesis test takes on the structure; are the coefficients (A) significantly different from zero?

Response:

(a) As described above the alternative model is well justified, and so are the choices of null-models (AR(1) and fGn). Otherwise it not clear whether the reviewer thinks about selection between different alternative models, between different null models, or use model selection as a hypothesis testing procedure between the alternative and null models. In any case the models to be selected between have to be specified. Penalising complexity through some information-theoretic criterion seems rather arcane here, where both alternative and null models are very simple, and it is not clear to us how such procedures could improve physical understanding.

(b) This suggestion seems meaningless, since the null hypothesis already involves

building a Monte Carlo ensemble by drawing long-memory processes. The suggested test must, by definition, give acceptance of the null in 95% of the cases.

(c) It seems that the reviewer here suggests to consider the trend coefficients in the alternative hypothesis as random variables whose distribution can be determined by the data. This implies a Bayesian approach where the trend coefficients (A) in the alternative model Eq. (1) are interpreted as random variables and the noise w(t) as an AR(1)/fGn characterized by variance and memory exponent which are also random variables. If the frequency and phase are not prescribed, the trend model will contain 5 and the noise model 2 parameters/random variables. In principle it is possible to compute the 7-dimensional likelihood function, but in order to use Bayes' theorem to compute the conditional PDF p(A|data) we need to choose a prior 7-dimensional PDF (uniform?) and then to integrate over all other parameters than the two trend coefficients (A). If one succeeds in computing p(A|data) then the trend could be considered significant if the origin (A)=0 lies outside the 95% confidence region. If this is at all feasible it suffers from the ambiguities in defining a meaningful prior, and one can hardly claim that it "simplifies matters."

Response to the review details

Page 328 L13. The consistency test with NH multiproxy reconstruction records before 1750 was done in Rypdal and Rypdal (2014). The discussion of that result was omitted from a previous version of the text. We shall remove this phrase from the abstract.

Page 328 L23: The reviewer does not cite the full sentence, which starts with "At the surface of things,. . .." It means that this conceptually simple, and very frequently used approach, is superficial and insufficient to establish a trend. By reading the full introduction there should be no doubt about what we intend to say.

Page 329 L5. We discussed this point in Theme I (a) above. The climate change debate encompasses more than the IPCC reports.

Page 329 L28. This was also discussed in Theme I (a) and Theme-II (a) above. We had, in fact, a more extensive discussion of the physical rationale behind the alternative model in an earlier version of the paper, but we omitted it due to concerns about the length of the paper. We can see that there may be reasons to put it back.

Page 330 L12. This was discussed in Theme II (b). We will rephrase this sentence in accordance with this discussion.

Page 332 L3 and Page 334 L14. In this paper we don't use information about forcing. The statistical uncertainty in the instrumental global mean temperature is very small, since it represents a mean over a very large number of observations, and it is even smaller on the multidecadal time scales, which can be considered as a moving average of observations over this scale. Systematic errors surely exist, but we don't know what they are, since if we did we would have corrected them. Such errors have presumably been reduced from HadCrut3 to HadCrut4, but the difference between the two records has no significant effect on the values of the parameters estimated for the trend model and the null model, and hence for the conclusions of our analysis.

Page 334 L13. Yes, these results emanate from our analysis! The fact that they also emanate from other approaches to the problem does not change this fact. Statistical and physical modelling are complementary, and if they yield consistent results we gain support of a hypothesis, and if they don't this support is weakened. The reviewer seems to favor approaches where physical and statistical modelling are mixed, or physics is used to constrain statistical models. That is fair enough, and is in accordance with what we like to call "a Bayesian spirit," but it is not the only way. Independent approaches have their advantages. That's way patients often like to have a second opinion from an independent doctor.

Page 334 L22. This point was responded to in Theme I (c).

P337 L27. It was not explicitly stated (we will do that in the revision), but should be clear from the exposition that the parameters of the alternative model are not treated

as random variables with a prior distribution that is turned into a posterior distribution. Thus the basic approach is frequentist. The possible exception is in the estimation of the parameters of the null model, where a Bayesian or frequentist approach may be used as described in the actual paragraph. The results here are indistinguishable.

P337 L9. This point is responded to in Theme III (c).

P339 L5. We have reserved the letter A for the two relevant trend parameters.

P339 L5. We responded to this in Theme II (a). We will revise the text to make our point clearer.

P341 L2. We responded to this in Theme – II. We will rewrite the quoted paragraph to something like: "...the land record analysis establishes beyond doubt that the linear temperature rise cannot be explained as a natural fluctuation and also that it is improbable that the apparent oscillation can be produced from the models for natural variability adopted as null models in this paper."

P 341 L17. We think it is necessary to distinguish between the technical discipline of Bayesian data analysis/hypothesis testing and Bayesianism as a school within the philosophy of science. The latter advocates a systematic approach to "plausible reasoning" where probabilities for the truth of assertions are reassessed in the light of new evidence. In this case the "new evidence" to be taken into account is the establishment of an "almost certain" linear trend, which has been established both from physical modeling and statistical analysis of observation data. We shall discuss this point more carefully in the revision.

P346 L27. We do not write that the climate model must be correct but that it must "correctly describe the relevant aspects of the pattern of natural variability." If it does not, how can the model results be useful as a null model?

The reviewer writes "ignoring information given in these models - along with underlying principles – is not a viable alternative." We discussed this point in our response to com-

ment on Page 334 L13 above. We agree that in an overall assessment, like the IPCC report etc., one cannot overlook physical modelling. But analysis of observation, unconstrained by our physical theories, can also be useful. The history of science is full of cases where discoveries have been delayed because flawed, but universally accepted, theories have constrained the analysis and interpretation of observation data.