Earth Syst. Dynam. Discuss., 4, C269–C274, 2013
www.earth-syst-dynam-discuss.net/4/C269/2013/

Earth System Dynamics Discussions
Open Access

ESDD
4, C269–C274, 2013

Interactive Comment

Interactive comment on "Agnotology: learning from mistakes" *by* R. E. Benestad et al.

R.E. Benestad

rasmus.benestad@met.no

Received and published: 16 June 2013

I'm grateful for the efforts Ross McKitrick has made in reading and commenting our paper. I think his views provide some interesting input to discussions concerning agnotology.

It is true that our paper was first submitted to two other papers, but got very mixed reviews. In both cases, there was one reviewer who vehemently opposed our paper, and McKitrick was obviously one of them as he himself admits. It comes as no surprise that he'd recommend rejection, since we criticize his own work – and we do not consider his review as being objective as a result.

It surprises us that the editors chose to get swayed by his negative comments, even after our response and explanation for why we were in disagreement.

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

McKitrick thinks that our paper is a “bait-and-switch”, but this is his own subjective interpretation – such labeling and hand wavy argument hardly has any scientific substance. We provide a set of samples of papers where we have found substantial flaws at various levels. To label it “scattershot of shallow commentary” with failing reasons is not very convincing. In fact, we do provide an R-package with computer code to replicate the past analyses. In many cases, the flaws are so simple and glaring that it does not require much work to uncover them.

The term “agnotology” is defined in our paper, and there is no presumption about it being “absence of information and a lack of basis for knowing”. We argue that we do know how we got to the different answers, and by tracing the line of work, we can learn why different analyses get different answers. This is the sensible thing to do, but is rarely done. Here we make an attempt. We did not intent our paper to go into a special philosophical issue.

Like the previous review, the first page of McKitrick response is merely personal views and lacking of substance. We argue that part of the reason why there are uncertainties concerning some questions is due to the lack of sharing methods and data. McKitrick is wrong to say that “evidence that non-disclosure was an issue for the papers they study” - we do indeed cite a reference in NewScientists (Le Page, 2009), describing how Scafetta maintained that the replication of his analysis was incorrect, but simultaneously refused to disclose his own code to get to the bottom of the matter. Humlum et al too have refused to show the code used in their analysis (this is not disclosed in our paper, however):we agreed to meet and sit down and work through the cases together. I sent some of my own code which I felt was relevant, but they never let me see their analysis.

The references Dewald et al. (1986) and Anderson et al. (1994) are not relevant for this case – they appear in the Journal of Money, Credit and Banking, which has little to do with our paper.

ESDD

4, C269–C274, 2013

McKitrick misinterpreted our reference to Le Page, 2009 – which concerned Scafetta and not MM04, and MM07. It is hard to see how he makes the connection to his own work.

Case 7. There is a real disagreement in our positions concerning the withholding/prediction test, and to refer to the test that I did as "an extreme version of the withholding test" is not very enlightening. To call it "extreme" is absurd, and a way to try to discard the results. This is not the way of objective scientific debate. In fact, I argued that the way they did this test was invalid because there are spatial correlations and that the data points they had in the different samples were not independent. McKitrick further misrepresents the test in Benestad (2004) stating " The Benestad (2004) comment only showed that it is possible to devise an extreme version of the withholding test, namely trying to predict the Northern Hemisphere data from the (smaller) Southern Hemisphere subset,...". If he had read the paper properly, he'd know that this assertion is untrue – quote: "In order to reduce the possible effect of inter-station dependency, the data were sorted according to latitude. Then, half of the data (latitudes from 75.5° S to 35.2° N) was used to calibrate the statistical models and the remaining data were used for evaluation (latitudes 35.3° to 80.0° N)". Furthermore, the failure of this test did have general implications, contrary to McKitrick's view. Even randomly picking data will fail to split the data sample into samples of independent data, since the temperatures and economic indexes were either correlated or even common within single nations.

It is well-known that spatial correlation reduces the number of real degrees of freedom and may enhance the estimates of statistical significance. It is surprising that McKitrick disagree on this issue, or at least that it is sufficient to undermine the significance of the conclusions. The question of the conclusion's veracity does of course depend on taking the real degrees of freedom into account. I will argue that the a split sample test that ensures independence (by a latitudinal stratification) is one of the best ways to do so.

McKitrick cites our failure to cite a paper in Journal of Economic and Social Measurement. This is a journal that journal that does not focus on climatological issues, and it is likely that the editors do not have a sufficient overview of the climatology community to find appropriate reviewers. Hence, the risk of lacking proper peer review. Furthermore, he expects us to follow journals specializing in completely different disciplines –when it's hard enough to keep up to date with all the new publication in our own field.

When it comes to the the differences in the opinion between Ross McKitrick and Gavin Schmidt, I think the proper thing is to let them discuss the matter between themselves. It suffices here to focus on the McKitrick and Michaels (2004; MM2004) and Benestad (2004) papers, and I still maintain that the analysis in is flawed, and the R-package replicationDemos contain a function that demonstrated this.

Based on previous dialogue with McKitrick (and the discussion above), I have learned that we cannot take his word as face value – we need to repeat the analysis and evaluate the models. We need to look at proper documentation and evidence. So far, McKitrick's arguments have not impressed.

Case 6. It is curious that McKitrick tries to defend a clear-cut flawed study such as Douglass et al (2007), where inconsistencies are revealed even with the simplest tests. The mix-up between the sample size and the error bars of the mean estimate reveals a serious misunderstanding of basic statistics. Douglass et al (2007) claim that since the observed trends are not similar to the mean of the climate model ensemble within the error bars of the mean estimate, the models do not correspond to the observations. This is nonsense as described in our paper - simple as that.

It is true that Foster and Rahmstorf (2011) don't look at the climate model results, but they do compare satellite-based trends with those based on surface measurements (thermometers). The Foster and Rahmstorf is not very central to our discussion, and our paper has neglected McKitrick et al. (2010) because we only selected a sample of papers and that we could not find space to everything. We will keep this paper in mind

ESDD
4, C269–C274, 2013

if we were to expand our replication exercises with further cases. At the present, I do not think that including this paper will add much substance, other than citing another paper of McKitricks'. Longer trends do not rectify the problems of mixing up the error bar of the estimate of the mean and the sample size. The objective of our paper is to show that there is a number of flawed papers that have influenced the public opinion, and provide an interesting case of agnotology.

Case 10. We can show that the presence of forcing influences the autocorrelation function (ACF), and there is one function in replicationDemos that shows this. We do not need a "statistical rebuttal" (whatever that means) to demonstrate that the trend testing involved circular logic when ever-present forcings affect long-term persistence. There has recently been a discussion on this issue on ClimateDialogue.org, as McKitrick too refers to. We are not saying that the proper null model should be AR1, and it is surprising that McKitrick thinks that we propose the use of AR1. It is important to keep in mind that we are not the IPCC, so mixing us up with the IPCC leads to more confusion.

We agree that LTP can also emerge from the internal dynamics, but that the process has to obey the laws of physics. There are some examples where LTP appears naturally, such as river levels, but this does not mean that it is present in the global mean temperature. Transferring the characteristics from one physical situation to another is logically flawed.

When it comes to Bunde and his view, our differences may not prove one way or the other, especially as these comments were on a blog. If it really is the case that GHGs do not lead to changes in the ACF, that subsequently may influence the FARIMA models used in Cohn and Lihns, we should be able to repeat this result over again. The replicationDemos package demonstrates that with climate models, GHGs have a clear influence on the ACF. I did not see that Bunde responded to the figure showing that.

Case 12. When it comes to PCA, I believe that McKitrick does not understand its use when focusing on the shape of the leading PC – the relevant question is how many PCs are included ah variance do they explain? Furthermore, it's the subsequent analysis that weights the different PCs according to their ability to describe past variations that determines the final answer. The argument proposed by McKitrick is therefore a red herring.

Contrary to McKitricks' claim, The reference to Wikipedia is not a "main source". It is merely provided to show that there is a long debate about the issue - quote: "These criticisms, however, did not convince McIntyre and McKitric, and further exchange followed in the literature (Mann et al., 2009; McIntyre, 2005a, b; McIntyre and McKitrick, 2009); there is also a long Wikipedia entry on this topic: "Hockey stick controversy" http://en.wikipedia.org/wiki/Hockey stick controversy. ". In fact, this reference was included to the benefit of McKitrick himself – in addition to show that this has been a controversy that has made a public impact.

The response provided by McKitrick suggests that he has misrepresented the split sample test described in Benestad (2004), AR1, and the reference to Wikipedia. McKitrick's response demonstrates that he is not understanding the difference between the error bars of the mean estimate and the sample range, and that he does not understand how PCA is used in reconstruction of temperatures.

McKitrick provides a nice input to the discussion on agnotology and the anonymous peer review process by giving everybody insight into the type of comments that can lead to a rejection of a paper. Furthermore, McKitric is one of the Heartland Institute's experts (http://heartland.org/ross-mckitrick). Our paper takes a critical position to this think tank, which should disqualify him as objective reviewer – he should have disclosed a conflict of interest (which he obviously did not). This case demonstrated the fact that the reason for rejection may not always have much substance to it, but may be more due to personal dislike.

Interactive comment on Earth Syst. Dynam. Discuss., 4, 451, 2013.