Earth Syst. Dynam. Discuss., 4, C262–C268, 2013
www.earth-syst-dynam-discuss.net/4/C262/2013/

Earth System Dynamics Discussions

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 review comments provided by J. E. Solheim (JES), Kjell Stordahl (KS), Ole Humlum (OH), and Harald Yndestad (HY), which provide an indication of points which need clarification. We take their remarks into consideration in our revision of the paper.

I maintain that there has been a culture neglecting replication, not sharing methods and data, or not testing the methods in general, and our paper provides a set of cases demonstrating this point. We also provide an R-package to illustrate how replication can be carried out, in line with Pebesma et al., (2012).

The important question here is not only whether the data are from open sources, but also whether the methods are robust and support the final conclusion. Since some of the cases in our paper replicate their analyses, we are not questioning whether the data

are open nor that the methods are not described. We merely say that the methods are inappropriate and lead to incorrect conclusions. We are repeating their investigation on the basis of the information given in the paper, but we are also saying that it would be even better if the source code for their analysis were open so that we could see exactly how the analysis was made.

Case 1. It is true that there is no mention of the giant planets and their possible climatic influence anywhere in the paper, and we have corrected the text to solar cycles and the moon. However, Humlum and Solheim do explain these variations in terms of the giant planets in the Norwegian paper in FFV 11/1, and they promote this idea in Scafetta et al., (n.d.). Hence, we are not making this up.

Curve-fitting can involve different techniques, such as regression, Fourier truncation, or wavelets. In this case, the specific method is not as important as how the results are interpreted. For such curve-fitting, it is impossible to make any statement about physical causes. We have revised paper, and now refer to the moon and solar cycles as in the paper. We have added information about the fact that the curve-fit was wavelet-based.

I still don't think that the statement "the main thrust of our investigation is on climatic variations in the recent past and their potential for forecasting the near future" explains why the discarded data was not used for evaluation purposes. If they assume forcing by permanent factors such as the moon and the sun, they must use out-of-sample data for testing their model. If they assume that their model cannot do this, then they implicitly say that there are describing a curve-fit, which cannot be used to make any causal statement. There would be little point in a paper like that, and I would ask if the paper had been accepted if they had shown the entire GISP2 data since the last ice age and a comparison between the model extrapolations and observed data for the entire length.

The wavelet-based curve-fit was based on long periods, so there is some degree of persistence. Hence, it does not validate a model by saying that it has (limited) skill (not defined in the paper!) for modest lead times, and this statement by itself implies the curve-fitting nature of the model. Furthermore, such curve-fits are by nature less reliable near the end-points of time series, and this is why wavelets often are expressed with uncertainty regions near the ends; they suffer from edge effects is known as the cone of influence (COI). There may be different boundary conditions used in wavelets (e.g. periodic or reflecting) that have different effects near the end-points and beyond. Extrapolations by wavelet-based curve-fit will often depend on the chosen strategy for handling the end-points. We can use the data before those of 4000 years ago, and we can see by eye that the interval they selected was quite different to the rest of the data, and that the fitted cycles quickly break down. They have still not explained why they did not show this in their paper.

Solheim et al. argue that BHDCN have to understand the forecasting process! That is a silly accusation, and if their objective is scientific progress and reducing uncertainty, they could help by sharing the code they used for their work – to avoid potential misunderstanding. This brings us back to the topic of agnotology and the first point in their response. Some more information about their work is provided in their response here, and we are grateful for this. The Humlum et al paper made no mention of stationarity nor transform to stationarity, and why their used the most non-stationary part of the data for their model training. Did they apply a basic form for detrending (e.g. regression-based?).

Astronomical objects behave very differently to geophysical processes, where the former tends to involve precise periodicities whereas the latter tends to involve complex chaotic processes. Our replication and Fig 1 in our paper illustrates this point nicely. This point is not a strawman argument, but illustrate the flaw in their analysis – the reference to astronomy is not relevant for this case. Again, the evaluation of the models must involve independent data, and persistence (long periodicities) does preclude the points near the end-points of the calibration interval. By stating that it was "safe"

(whatever that means!) for the forecast about 10-25% of the series length does not provide much information when the autocorrelation is not discussed.

There has been an upward trend in temperature since 1900 that coincides with an upward component of the fitted curve. Because the curve-fit involves time scales greater than 500 years, one would expect to see a trend, but trends are not ideal for validation of models. We have plenty examples of that, e.g. child births and nesting of storks in Scotland.

Case 4. It is true that research often starts with an empirical relation, but often such speculations turn out to be incorrect. I do not have the statistics on how often empirical patterns turn out to lead to new understanding or whether it was a fluke – but without this comparison, the argument must be taken as hand waving.

What we now know is that up-dated temperature measurements on Svalbard indicate that the Solheim et al mode fails in a big way. Unless there is an unprecedented (since 1898) big freeze in the Arctic.

By selecting the lagged results and winter temperatures, they inflated the significance (here, not in the statistical meaning but in terms of interpretation) of the correlations when they cannot provide a good explanation for why there is little correlation with the corresponding epochs and for the other seasons.

The point about SCL not being estimated but is taken from NGDC is corrected in the revised paper. SCL is hard to pin point accurately, which is illustrated by the fact that different estimates exists. To say that one version is a far better estimate just because it is the "official" version is not very scientific (The Benestad (2005) SCL is derived from an objective method). The same logic is saying that the conclusions from the IPCC is much better that any of the cases described in our paper because the IPCC is the official line. I don't think that Solheim et all really agree on that, and our paper tries to look into why there are differences – hence the reference to agnotology.

Even though “solar cycle 24 started officially in 2009” [sic] there are uncertainties around SCL, and the table in Solheim et al did set 2008 as the end of cycle 23 on which it based it forecast. Hence, I’m a bit surprised by this comment – nature and science do not care about what is “official” or not. Furthermore, the period 2008 – 2011 provides an illustration of the situation. The temperatures from 2010 are also very high, and doesn’t change the picture much.

I appreciate the new information provided about the analysis concerning SCL and its error bars. Again, it would be better if they had disclosed this together with the publication. The replication was based on their table, which listed the numbers. We compared the results, and got similar values, as described in our paper. The assumption that the SCL from NGDC is calculated with the precision of 0.05 yrs is also new information – it would be good with more documentation.

I agree that the estimated confidence interval according to R’s ’cor.test’ is not exact, also for the reasons explained in our paper. We also argue that the bootstrap approach in Solheim et al is wrong, and we demonstrate this through our replication of their analysis. They may disagree, but to say that “bootstrap methodology has been used of statisticians for many years now, and is accepted as a methodology for giving deeper insight in statistical data analysis” is not very substantial – it’s mer hand waving. I know that bootstrapping does a good job in many cases. Here it was misapplied.

Solheim et al assume that there is weakness in Benestad (2005) SCL estimates, but does not provide documentation. It seems just because the values do not fit their results. Solheim must also document the statement that “significant uncertainties in the estimates of the mean temperature and in the standard deviations” . Their conclusion lacks any scientific substance.

Case 4. It is well-known that short-term variations such as ENSO drives CO2-changes, but Humlum et al’s logic fails when taking that to the long-term trends. This is now explained in Richardson (2013) and Masters and Benestad (2013), and these citations

are now included in the revised paper. I would also like to point out that the source code for carrying out the demonstrations are a part of replicationDemos.

Humlum et al also failed to contact the providers of the CO2 data that they used. My impression is that they would have advised Humlum et al against the way they used the data (I have asked them for their view on their analysis, based on description made on Realclimate.org). They may not have read the conditions for using these data (asking for advice prior to publications), but having used the global CO2, they should themselves make their code and data available for others. This is also the point we want to make in our paper.

Case 13. CO2 measurements may be contaminated by local sources, and not show the background levels. Modern satellite-based measurements suggest substantially higher CO2 concentrations over the North American, European and Asian continents, close to CO2 sources. Humlum et al argue in case 4 that they chose not to use local CO2 measurements made at Mauna Loa, but argue here that it’s OK to use measurements since 1800 in Europe which were likely to have been made near human CO2 sources.

If these CO2 measurements have any merit, it is important to write-up a paper with proper documentation that passes peer review. I do not think that E&E is a journal with credibility. At this stage, the E&E paper by Beck is not convincing.

Case 14. Many others too calculate the effect of CO2 through line-by-line code of the IR. Miskolzi presents a very simple model in the E&E journal, where he neglects important aspects such as latent and sensible heat fluxes. This is explained in our paper.

The Lu paper is curiously published in a journal for Condensed Matter Physics; Statistical Physics; Applied Physics. It would be better to publish such papers in climatological journals where the editors tend to have a better knowledge about appropriate reviewers. Furthermore, the Lu paper just appeared, so we need time to consider its substance – perhaps it is another case to replicate? Nevertheless, a paper by Lu does not imply that the paper by Miskolzi – it is faulty logic to imply so.

Case 15. I don't see the connection here.

Case 17. OK – Arctic climate it is. And wavelet rather than harmonic analysis, although a harmonic analysis would also do for any permanent periodic response. The periodicities are probably correct, but we question the interpretation of the causes. The title is revised to 'Misinterpretation of periodicities', which what the section is about.

A final observation which brings us back to agnotology is that Solheim, Stordahl, Humlum, and Yndestad are a part of "klimarealistene" mentioned in our paper. They are also senior scholars who only started publishing work concerning global warming and causes for climate change in scientific climatology journals in 2011, years after they presented their first "skeptical" views in the media (e.g. forskning.no: Humlum started in 2006; Solheim, 2008; Stordahl, 2007; and many more! – this is documented through the search engine on www.forskning.no). They had already taken a stance which they now defend by submitting papers scientific journals in climatology, with characteristic prejudice against the main stream climate science. Furthermore, I expect that they are not objective to any criticism of their own work – as presented here in our paper. I welcome a debate and further investigations into these questions, but I will also reiterate my point that the debate should be based on openness and transparency. We need to test their methods and evaluate their skill, as we have done in replicationDemos. I hope they will start to disclose the source code and data, and will recommend R-packages which provide a nice frame for doing so (Pebesma et al., 2012).

References Pebesma, E., Nüst, D., Bivand, and R., 2012. The R Software Environment in Reproducible Geoscientific Research. Eos 93, 163–164. Scafetta, N., Humlum, O., Solheim, J.-E., Stordahl, K., n.d. Comment on "The influence of planetary attractions on the solar tachocline" by Callebaut, de Jager and Duhau. J. Atmospheric Sol.-Terr. Phys.
