General.

We appreciate the reviews from the two anonymous reviewers and the comments posted through the discussion. It is important that the authors of the criticised papers to have their say. The reviewers pointed out that the paper was too philosophical, and we have tried to revise it so that it's more of a documentation of the current situation for the climate sciences - to describe science in practise. The reviewers were mostly concerned with the main paper, whereas the comments posted in the discussion mostly aimed at the appendix. Two of the commentators provided some good points for the main paper, which we have tried to bring into the paper.

The paper has undergone a major revision and re-structuring. We also add the R-package 'replicationDemos' as supporting material, as two versions: one for Linux/Mac and one for Windows. This package will allow the scrutiny of our work by fellow experts, whom we assume have hands-on experience with such computer code and similar analysis. We make a note that none of the reviews nor comments discussed our code, which was available on the R official site (CRAN); although there was rather a complaint that we assumed the reader should be able to run the R-code. The lack of appreciation of the source code and data is a central topic in our paper, and we assume such code would be of value to fellow experts. It is assumed that peers with hands-on experience with coding and data analysis will, from our description, realise that installing R-packages and checking the codes is simple, and this R-code comes with the necessary data, and together they make up 'hard evidence' regarding analytical strategies and methods. We also expect that it will be sufficient for the general science-literate reader to read the main paper and the appendix. The paper was not really aimed at non-experts outside the climate research community. The computer code and data also make the paper scientific rather than philosophical. We have tried to clarify the distinction between the levels further in the revised manuscript (the paper has three levels), so that readers will be able to understand our arguments, or learn from the examples in the appendix.

Reviewer #1

Review Report of "Agnotology: learning from mistakes"

I have no time to redo all the analysis to judge which side is right, but based on what I read I found there might be some critiques for which the authors did not provide enough justification. For example, based on what fact, the authors claimed that the data in Humlum et al. (2011a) violates the Dirichlet condition? From the definition, a function f (x) satisfies Dirichlet condition if

1. f (x) is absolutely integrable over a period.
2. f (x) has a finite number of extrema in any given interval, i.e. there must be a finite number of maxima and minima in the interval.
3. f (x) has a finite number of discontinuities in any given interval, however the discontinuity cannot be infinite.
4. f (x) is bounded

It is not obvious to me which one of those four conditions is violated, and I am curious how the authors verified those conditions and found at least one of the four is not ture.

The reference to the Dirichlet condition has been removed - the important point is the comparison between out-of-sample predicted and observed values.

The paper has a too long philosophical Introduction for a scientific paper, and I do not think it is necessary.
The philosophical aspect has been toned down, and the focus now is more on the documentation of climate science in practise.

In my point of view, the authors just need to present critiques based on the results, and not need to be so philosophical. I would prefer the papers just focus more on science. The title of the paper should be accordingly changed. Some published papers may not use the mostly appropriate methods, but they became the basis for the development of the most precise results, so I feel the tone in the paper is a bit too strong.
The discussion of some of the cases have been expanded, although all are now in the appendix. It is appreciated that some published papers which have been the basis for the development of most precise results may not have used the most appropriate methods; however, we will argue that this is exactly a point in case where we have learned from mistakes. The paper has been strongly revised, has a revised title, and we take on board these points.

Review #2
While the multiple bogus cases dissected by the authors all need to be formally rebuked in the peer-reviewed literature, I am not convinced that it should be done as in the present manuscript, under the somewhat pompous banner of “agnotology”, or under the premise described in the abstract.
New title is just ‘Learning from mistakes in climate research’.

Philosophizing The introduction is needlessly philosophical, obscuring fairly simple concepts under a verbose cloak of lofty assertions (with many typos and grammatical mistakes to boot!).
The philosophical aspect has been toned down, and moved towards the end of the paper. The focus is now on the different perception between the general public and climate scientists, with a reference to the review of the list of cases. We interpret the discussion of this paper and the posted comments as a manifestation of science in practise, where controversies are debated.

In my experience, there are three main reasons why scientific studies might be flawed:
1. One may start from a correct logical premise and execute an erroneous analysis
2. One may apply an correct analysis but start from the wrong logical premise
3. One may start from the right premise, and correctly apply the analysis, but overstate the significance of the conclusions (i.e. the analysis does not actually address the question)

The paper mainly deals with #1, leaving the other two possibilities somewhat buried within the laundry list of the appendix. The two cases selected for prime time do not seem to constitute enough material for a paper.

The revised paper brings in these three points, in addition to a fourth: starting from a wrong premise and applying wrong methods (two wrongs don't make a right). Now all the papers/cases have been moved to the appendix, and some of these have been expanded according to the comments during the discussion stage.

It seems to me that the authors face a choice: do they want to write a philosophical opus? If so, they should make sure to cover all logical possibilities for scientific ignorance or lack of consensus, and send it to the Transactions of the Royal Philosophical Society or a similarly-minded journal. If they want to write a paper on reproducibility, they should make this focus plainer. As it stands, the paper is an unwieldy hybrid of philosophy of knowledge, replication studies, and scientific meta-critique.

We want to document the state of climate sciences and how it reaches the general public. At the same time, we want to demonstrate the merit of replication, open-source methods, and sharing of code and data. In our paper we try to set a good example on how this can be done through the use of R-packages. We know of few other other scientific publications that provide the methods in this fashion, in a package containing the code itself, the necessary data, and documentation.

Overfitting I think the main problem with the two cases presented here is the logical fallacy of fitting a statistical model to a set of observations, using a the calibration skill as an implausibly rosy measure of prediction skill, and going on to make predictions about the real system without making sure that the model has adequate out-of-sample skill. This is simply known as overfitting, an elementary statistical notion that is explained in any good statistics textbook. It is sadly common in science, usually because non-statisticians are unaware of the problem, and happens usually because of ignorance rather than an intention to deceive. Perhaps the authors could use this framework to kill several birds with one stone, as it is a mistake common to many of the papers they scrutinize?

We hope that our paper will generate some interest by sparking discussions through replicating many different papers which we think are clearly flawed. We agree that overfitting is an elementary statistical notion that should be explained in any statistical textbook, however, we also see that some publications do not seem to appreciate this notion: yes, it seems to be a common mistake. This is now emphasised in the revised manuscript.

It would seem aptly philosophical to subsume many studies under one common framework rather than segmenting the issue into many small pieces.

Statistics vs logic The basic premise of the paper is that many of the errors in the denialist literature indicted here stem from methodological flaws. However, in the case of the multiple bogus studies by the prolific Dr Scafetta (whose work has been debunked in many places, including Benestad and Schmidt [2009] and http://agwobserver.wordpress.com/anti-agw-papers-debunked/), the issue is less the analysis itself (though there are certainly plenty of problems there), than the underlying logic. However, nowhere in the main text is

the word “logic” ever used, which I find odd.
The revised paper is now emphasising the logical aspects of both Dr Scafetta’s work and several other papers.

Scafetta [2012] purports to test climate models over the instrumental record, ignoring the fact that no GCM is ever expected to match observed temperature, in light of natural climate variability. In doing so the author sets up a strawman, an impossible task that models are not designed to achieve. In contrast, overfitting a sets of sinusoids will, by construction, match the training data, but may radically mis-predict out of sample observations. The logical fallacy here is that a model with no physics will outperform a physically-based model, not understanding which aspects of the climate system the performance should be judged on. In this case, it is fair to ask the models to reproduce various statistics of temperature observations (e.g. mean, power spectrum, higher order moments), but not the phase of these variations, which may be viewed as random.
This point is appreciated and brought onboard.

In many of the cases they investigate, the authors allude to these logical flaws as “unclear physical basis”, which is a little too vague. I think “agnotology” would be better served by precisely identifying logical fallacies.
Point taken - in the revised papers the logical fallacies are emphasised. Also it is mentioned that unclear physical basis does not rule out a connection.

Minor Comments
• The English is an unholy mess, which is shameful given that there are several native English speakers amongst the authors. This is just not serious and makes the reader feel like their time was wasted. I normally point out all the typos and grammatical mistakes, but in this case there are too many to type in a review, and given the comments above I don’t think it would be a productive use of my time. I do have a manuscript covered in red ink, if the authors are interested!
The paper has been revised, and we have been conscious of the grammar.

• The wikipedia page on agnotology claims that the first occurrence of the term dates back to Proctor, Robert N. (1995). Cancer Wars: How politics shapes what we know and don’t know about cancer. New York: Basic Books. ISBN 978-0-465-00859-9. I have no access to this book but I recommend that the authors check this, as it would change the date of the terms first use (L20) by 13 years.
This is appreciated - we now don’t claim that the term was first used, but argue that the term was coined in Proctor and Schiebinger (2008) who do spend some lines defining and explaining agnotology.

There is every sign that climate ignorance amongst the public is (like the tobacco scandal) due to a malicious intent to deceive, perpetrated by scientists who knowingly produce junk science in

the hopes of confusing public [Oreskes and Conway, 2010]. Exposing those logical and methodological flaws is a valuable endeavor aiming at dissipating this cloud of doubt. While the cases investigated here are quite interesting, and deserve to be published in an open scientific journal like ESD, I cannot recommend publication under current form. Indeed, the title, introduction and discussion of the paper appear to be a hastily-written wrapper slapped together to package these examples for publication. The fact that those sections are so ill-written only serves to weaken the arguments made in the 17 cases presented here. I recommend that the authors restructure their paper around common themes (e.g. logical fallacies or common methodological mistakes), write it clearly and concisely, avoid snarky comments against denialists (irritating though these characters might be!) and work on a coherent presentation, instead of publishing a laundry list of replication studies and wrap it in dubiously written philosophical verbiage. This will make for a much stronger contribution to the scientific literature.
Our intention is not to be snarky, and without examples of what are the reviewer snarky passages, it is hard to know where to start. Maybe the reviewer thinks that the issues that we discuss are just too elementary? Reading the discussion comments from the authors it is clear that they seem to take these arguments seriously. We have kept this point in mind when revising the paper, however. We also hope that the revised paper has less of a character of being "hastily-written wrapper slapped together to package these examples for publication" (we can assure that the paper was not hastily-written wrapper slapped together, as it is the result of a long process with previous reviews and where each reviewer has a different opinion. Furthermore, extensive time and efforts have been devoted to developing and evaluating the the accompanying R-package).

We do not understand why the reviewer takes the view of our paper as ill-written and full of grammar mistakes, but we've kept this point in mind through the revision of our paper. There has been a major revision, and now the introduction is different. We also understand that there may be a range of different expectations from a paper like this. Our manuscript is perhaps not a perfect piece of work that lives up to the very best of what science has produced ever, but we think it's a good start on a necessary debate. The comments to our paper indicates it's important to bring forth these matters, and we hope someone will follow suit and write an even better paper. We do not think our paper contains any flaws or fallacies, and our perception is that most of the objections to our paper are based on more subjective judgement. We have included some hard facts (replicationDemos), which has not really been discussed by commentators or the reviewers.

Response to the comments from the discussion-phase: several of the people who wrote their comments had had their papers criticised in our discussion paper, and we think that the opportunity for them to explain their interpretation is indeed valuable. Of

course, we disagree on many topics, which are explained in detail below. We do not feel it's our call to respond to the comments that Drs Scafetta and Rypdal made to reviewer #2's comments or to each other's comments. Those are somewhat off-topic, but illustrate that the issues brought up in our paper provoke discussions. This is good. Some of the comments are rather extensive (particularly Dr Scafetta's, who posted several comments), despite the instructions to be concise, and we feel that many of these comments are not directly relevant to our paper.

Loehle

This paper makes two fundamental assumptions that are false. The first is that the science of climate change is "settled" and that this consensus cannot be questioned.

This is a position that our paper is not adopting, and is a misrepresentation of our message. The revised paper now spells out that the science is never settled.

The second, which therefore follows, is that anyone this consensus does so willfully and malevolently (intentionally promoting).

This is also a misreading of our paper. In our revisions, we have tried to phrase the text to make this misunderstanding less likely.

We can distinguish between weak consensus and strong consensus. Weak consensus is "humans are having some impact on the climate," a statement with which every one of those critiqued in this paper doubtless agrees, but it is a meaningless statement in terms of either science or policy. A small amount of warming would be "some" effect but would both contradict the models and not be alarming. The models assume that essentially all of the warming since 1950 is due to human action. If half or more of this is natural then the models are seriously wrong. Confusing "some" effect with an effect as large as IPCC claims is how Cook got the 97% figure in his recent study.

The climate models make simulations according to the mathematical equation programmed into the code, and these are based on physics and empirical data - they do not 'assume'. Cook et al. (2013) also included categories explicitly quantifying the anthropogenic contribution, and 96% of author self-rated papers in these categories placed the human contribution at greater than 50% since 1950, consistent with the IPCC. In any case, the main point of this discussion is that the perception on climate change within the general public and the climate sciences are different.

Strong consensus implies that the IPCC documents are as fixed in stone, as scientifically solid, as Maxwell's equations or Einstein's laws. Yet no two models produce the same output for global temperature histories or future warming, ocean currents, polar ice, cloud patterns, ENSO, the polar vortex, the Gulf Stream, or global or regional pre-cipitation (to name just a few), nor do any of the models match actual historic data for these items in anything like a precise manner. If this is a consensus that must not be questioned, it is a very curious and sloppy one.

It is important to question the consensus, and the best for scientific progress is to provide open-source code used for analysis and share the data. We are emphasising this in our paper.

The second assumption is pernicious.It asserts that anyone who disagrees with this (sloppy) consensus in any particular is not merely wrong, but willfully wrong; that is, is engaged in disinformation or propaganda.

This is not a logical implication of our conclusion; however, we mention some work suggesting that in some quarters there may be a degree of that. We are not adopting a position on this, but merely state that openness and replication can reduce the risk of this taking place. We also note that this point address the concerns that Hanekamp raises: non-epistemic consensus (more on this in the response to his comments).

“Agnotology” is thus just a fancy way of saying “denier.”

This is incorrect and does not reflect the content of our paper.

Overall, the manuscript is simply a litany of complaints about papers the authors don’t like. Yet in no case do the authors bother to truly refute anything, they simply argue that this or that “might” have a problem or uses a method they disapprove of. They are essentially requiring the reader to take their word for it that these papers are wrong.

No, it is exactly the opposite - we provide the general society with our source code and data so that anyone with some analytical and programming skills can check our results. We argue that replication is important, and others should not take our word for our results. We do, however, try to explain why we think the cases that we refer to are flawed.

James Hansen insists that the whole scientific community (including the IPCC) and all the models are wrong about future sea level rise–why don’t the authors criticize him?

James Hansen suggests that faster sea level rise than predicted is a possibility. However, his hypothesis is based on paleoclimate data and cannot yet be tested, unless there is a published paper we are unaware of.

In the general case, disputes in science can and do go on for many decades.

We agree, and we continue this dispute with our paper. We are entitled to voice our opinion about the cases that we have selected. We hope that Loehle and the other commentators don’t try to stifle this debate by suggesting a rejection of our paper.

To prematurely claim consensus is a clever shortcut simply designed to shut up one’s opponents. That is, it is a political ploy.

We agree - and we are not shutting up anybody. We are doing exactly the opposite by shedding light on papers which we find unpersuasive.

Because they are dismissing papers that have been published, they are further setting themselves as higher authorities than the peer-review system.

We are just questioning some results - as one should question a consensus if one doubts its basis. We think that Dr Loehle is being inconsistent here, and that he is trying to redefine our role.

As an experienced scientist, I have more publications (143) than most (maybe all) of the authors. This only matters if one wants to make claims of authority, as these authors do, but I do point out that ALL of the works being criticized here are authored by well-published scientists, not hacks.

We have not suggested that the papers whose results we attempt to replicate are "hacks."

The standard in science is that refutation of a point requires a coherent argument, not merely handwaving that something "might" be confounded (as their arguments about spectral analysis).

We agree, and hence the accompanying R-package 'replicationDemos'. We do not think our work is handwaving, but in the revised version of our paper, we have tried to more carefully explain our points.

The authors repeatedly object to results in papers for which "a clear physical basis is lacking" or "not based on physics." Science proceeds by first attempting to detect regularities in nature. The "physical explanation" usually follows.

This is now revised and we explain carefully the implications of this. In some notable cases, there have indeed been ground-breaking discoveries based on just empirical evidence, although there may also be a vast number of undocumented cases where empirical patterns just turned out to be a coincidence. The point is that the evidence for a connection is still not very strongly established, and that it is only speculative at the point where the physical mechanisms are not understood.

On page 455 the authors mention the resignation of editors as proof of bad science getting published, but the ClimateGate emails showed that these events resulted from pressure from a handful of activists such as Mann. This only shows how political the subject has become and how much power the Hockey Team has.

This is an inaccurate allegation.

Case 3 in A2 (A2.1, p. 466+) examines Loehle and Scafetta (2011), my paper. The authors' statement that our results "are at variance" with most of the climate science community is false (and if no one could publish anything that disagrees with dogma I don't think science would progress). We cite multiple attribution studies that attribute only part of the recent rise in temperature to human activities.

We take the IPCC and the NAS reports to provide a reflection of the consensus within the climate science community, and Dr Loehle has already implied that his paper is among the selection that disagrees with the consensus.

The bulk of empirical sensitivity studies also give lower sensitivity than the models. On p. 466 l. 25 they find it “difficult to conceive” what could cause the forcings we propose. The fact that there is a ~60 year solar cycle and a ~60 year PDO makes our hypothesis far from absurd, and in fact represents movement toward a mechanistic model. That they lack imagination does not invalidate the putting forth of a hypothesis that the sun is having an influence on Earth’s climate. Do they think it does not? Solar forcing can act via electromagnetic currents, ultraviolet effects at the poles, the cosmic ray mechanism, and differential heating of parts of the ocean which change clouds, atmospheric movements, and ocean currents this helps their imagination.
The effect on cosmic rays must be verified through the low-level cloud record, but there are no clear indication of a sustained correlation nor that the clouds vary at this frequency.

They claim it is not valid to fit to 160 years of data and that we did not validate our method. We in fact show a validation test in the appendix.
The supporting material in the L&S11 paper misses the point. It states: “*For obtaining a very precise estimate of signal properties this may be true, but for estimating periodicity to within plus or minus a few years, this is not true*”. This is not the issue here - the issue is whether these cycles are persistent and characteristic of the earth system. We show in the discussion of how a short interval fails to capture the true long-term character of ENSO in case 2.

Their attempt to demonstrate that our approach is wrong using a synthetic series is itself wrong. They did not replicate our method and fitting periodic data is not simple. In fact, a free fit of cycle length to the data gave results visually almost indistinguishable from the results we got. We use the solar cycle lengths found in Scafetta’s earlier work to filter out the solar effect on the post-1950 period warming and attribute the residual to human activity. It is an attribution study.
Our demonstration clearly shows that sinusoids with a range of frequencies provide reasonable fit for the calibration period, and that short series compared to periodicity (only a few cycles) tend to yield misleading results. Based on these facts, we argue that the analysis does not provide a very convincing attribution study.

There is no dependence on being able to properly find 60 and 20 year cycles in 100 years of data except for the lag (timing) and amplitude because the existence of these cycles (based on a solar connection) is our HYPOTHESIS. We think the quality of the fit and the fact that our data 1850-1950 enabled us to predict the post-2000 flat temperatures speaks for itself. Their critique shows only that working with time series is tricky, not that we made any mistakes. In fact, they clearly did not read our paper very closely.
There is no logical connection between ~20 and ~60 year cycles and solar forcing - this is merely assumed. Furthermore, we argue that the strategy is not suitable for addressing the scientific question concerning a sustained presence of these cycles and the connection to solar variability.

Solheim et al.

The accusation made in the BHDCN-paper that we belong to a .. culture neglecting replication, not sharing methods and data, or not testing the methods..is completely wrong and fails totally if our papers are read correctly.

This is not exactly what we say. The manuscript has been revised, and we try to explain even more carefully our position.

The claim that we have hidden data, and give incomplete description of methods is wrong. The data used are all from open sources, and our methods are described in our papers.

We have not claimed this either. One point though, is that sharing the source code for the methods and the data will make it easier to probe the analysis, test the assumptions, and shed light on controversies - this rarely happens. The paper has been revised to emphasis this.

Other researchers may use other methods and reach other conclusions. New data may appear which give different results. We may have done mistakes, and that will of cause be corrected by repeating the investigations by ourselves or other scientists – and other methods may of cause give different results.

We agree on this, and our paper is trying to do exactly that.

In the following we will comment on some of the cases. To us, the BHDCN paper is full of strawman arguments, which may emerge from poor reading or understanding of the papers they criticize.

We try to look at the core of the analysis and the important bits on which the conclusions hinge.

Case 1: ignoring data which do not agree with the conclusions.

BHDCN: Humlum et al. (2011a) suggested that the moon and the giant planets in the solar system play a role a role in climate change on Earth, and that their influence is more important than changes in the GHG.

OH: There is no mention of the giant planets and their possible climatic influence anywhere in the paper, and the above statement appears to have been made up by BHDCN, and it appears that they have not even read the paper referred to thoroughly. The possible influence of the Moon is however mentioned, but this can hardly be considered controversial, as this association has since long been documented by fishing and tide records. Several references to this are provided in our paper.

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., (2013). Hence this is not ‘made up.’

reference: Scafetta, N.; Humlum, Ole; Solheim, J.-E. & Stordahl, K. (2013). Comment on “The influence of planetary attractions on the solar tachocline” by Callebaut, de Jager and Duhau . *Journal of Atmospheric and Solar-Terrestrial Physics*. ISSN 1364-6826.

BHDCN: The core of the analysis carried out by Humlum et al. (2011a) involved curve-fitting and tenuous physics, with a vague idea that the gravity of solar system objects somehow can affect the Earth’s climate.
OH: The analysis is not based on curve fitting, but on wavelet analysis, which is something else. There is no mention of the gravity of the solar system anywhere in the paper, and this point of criticism is apparently also made up by BHDCN.Page 459:
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.

BHDCN: They claimed that they could “produce testable forecasts of future climate” by extending their statistical fit, and in fact, they did produce a testable forecast of the past climate by leaving out the period between the end of the last ice age and up to 4000 yr before present. However, they did not state why the discarded data was not used for evaluation purposes, and the problem with their model becomes apparent once their fit is extended to the part of data that they left out. We extended their analysis back to the end of the last ice age. Figure 1 shows our replication of part of their results (Table 1). It clearly shows that the curve-fit for the selected 4000 yr does not provide a good description for the rest of the Holocene.
OH: On page 145-46 in the paper discussed we state the following: “The second series, the GISP2 data has merit because of the long time range represented (back to the Eemian interglacial), and because the Greenland air temperature appears to vary in overall concert with the temperature of much of the planet (Chylek and Lohmann, 2005; Brox et al., 2009). Here we chose to focus on the most recent 4000 years of the GISP2 series, as the main thrust of our investigation is on climatic variations in the recent past and their potential for forecasting the near future. In addition, this part of the GISP2 series shows an overall linear temperature trend, which simplifies the following analysis”. Here we clearly state that the GISP2 record is much longer than just 4000 years, and explain why we make use of only the last 4000 years.
We 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. We would ask if the journal

would have accepted the paper if Humlum et al had shown the entire GISP2 data since the last ice age and the comparison between the model extrapolations and observed data for the entire length.

On page 152 we define limitation of our approach by writing “This suggests that natural variations which are strong now are likely to continue without major changes at least some time into the future, and therefore likely to influence also the future climate. This knowledge on persistent and strong natural variations may be used for an attempt of forecasting, at least some time into the near future”.
On page 155 we write “Our empirical experience suggests a realistic forecasting time range of about 10–25% of the total record length. In the case of Greenland, such forecasting suggests that the present post LIA warm period is likely to continue for most of the 21st century, before the overall Late Holocene cooling may again dominate, but this being dependant on the magnitude of the anthropogenic greenhouse enhancement”. By this the limited forecasting range is clearly defined, and a ‘hindcasting of more than 6000 years based on the 4000 year record used as indicated by Benestad is highly inappropriate and is not in any way reflecting what we are writing in the paper.
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 authors have still not explained why they did not show this in their paper.

BHDCN: Furthermore, they failed to acknowledge well-known shortcomings associated with curve-fitting, but rather based their analysis on unjustified fit to a set of Fourier series.
OH: Once again, we did not employ curve fitting, but wavelet analysis, which is something quite else.
We maintain that their analysis is a fancy form for curve-fitting, based on wavelets. One can also call it an overfit.

KS: Forecasting is a challenging process. It seems like BHDCN think that the forecast period should be of the same length as the length of the observed data. However, the important part of the observed time series of data is the data closest to the intended forecasting period - obviously, because these data gives more information for future

predictions.
This applied to forecasts based on a present state (e.g. initial value problems), but in this case, we are discussing cycles which may be forced by variations in the sun (e.g. boundary value problems). For the latter, there is no reason to expect that the model results will diverge with the true outcome if one knows the history of the forcing.

BHDCN have to understand the forecasting process!There are certain conditions which have to be satisfied, to make a forecasting model. One example: the data (time series) have to be transformed to stationarity in the forecasting process. (When the analyses are finished, the analyzed data are retransformed). This is the case for ARIMA models, wavelet analysis (which is used in Humlum et al (2011a)) etc. Hence, the ability of the time series will in many cases influence of the selection of the time period used for the analyses and model building.
Solheim et al. argue that BHDCN have to understand the forecasting process. They have used the most non-stationary part of the data to calibrate and evaluate the model, and say that they probably should use the more stable part if stationarity is a big issue. Anyhow, we have accounted for trends in our analysis, as discussed in our paper. Our replication is based on the harmonics and the frequencies given in the paper, and our results are to our best knowledge identical with the results in Humlum et al. (2011a). The fact that the model results cannot be extended to the part of the data that was ignored, indicates that their analysis is a mere curve fitting exercise.

BHDCN use the word “curve fitting” for modeling the temperature evolution. “Curve fitting” is a misleading expression. It is important to underline that the forecasting method, wavelet analysis, is a rather advanced forecasting method.
Curve fitting can be based on advanced methods.

Another point: The long-term temperature time series reflect temperature the last glacial period and the last interglacial period. For a forecaster point of view, it is important to analyse the temperature evolution the last part of the interglacial and NOT the last glacial period when temperature forecasts for the last part of the interglacial are developed.
We too have excluded the last part of the interglacial because of altered conditions, and our replication is merely done for the holocene.

JES: Another strawman argument used by BHDCN appears when they explain that a function can be fitted by a sum of Fourier series having no predictability at all (outside of the data interval), hereby indicating that the periods we find cannot be used for predictions.
This is a well-known fact, as documented by the reference to several textbooks on Fourier methods.

This is correct if the series contain just noise, but it is our experience from research on light variations from stars and planets, that it is often possible to find stationary, or nearly stationary periodic signals, which can be predicted to continue at least in the near future. We only used the

3 most significant periods, and we tested the predictive power by deleting a part of the series. This is shown for the Svalbard data as a proof of the validity of the method. Our conclusion was that it was safe for the forecast about 10-25 per cent (of the series length) in the future. So we have done exactly what BHDCN claim we have not. Extensive Monte Carlo tests (S.O.Kepler 1993, Baltic Astronomy 2, 515) have proved the validity of FT periods with amplitudes above the FAP (false alarm probability) as developed by J.D. Scargle (1982, Astrophysical Journal, 263, 835). We used only the strongest periods fulfilling this criterion. The ice-core analysis was done for 4000 year input, and we made a forecast for the period 1850-2800. This forecast shows a temperature increase 1900-2000, which is observed. Our model is then clearly verified outside the range analysed, and may be a good predictor for the next couple of hundred years. Our model shows that the temperature may reach a maxium during this century, and that we are close to that maximum now (figure 8).

Astronomical objects behave very differently to geophysical processes, where the former tend to involve precise periodicities whereas the latter tend 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). This precludes the points near the endpoints 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. These points are now discussed in our paper.

Case 4: ignoring negative tests

BHDCN: The conclusion of the paper lacked clear physical basis, as the chain of processes linking the solar cycle length and temperatures in the Arctic over the subsequent decade is not understood.

JES: Research often starts with an empirical relation. This relation may have predictive power, even if the physical reason is not understood. In this case we have suggested a simple model: Heat transported with the Gulf Stream make the North Atlantic warmer. A heat pulse from the Caribia takes about 10 yrs to reach the coastof Norway. Maybe it needs longer time to reach Svalbard. It is reasonable to speculate if this delay can be used for forecasting. If a heat pulse is created at each solar maximum, longer SCLs may mean longer time between heat pulses, which means fewer heat pulses per time unit, and less warming.

Case 4. It is true that research often starts with an empirical relation, but often such speculations turn out to be incorrect.

BHDCN: Furthermore, the analysis was not objective, inflating the significance of the results. A more subtle aspect of this study was the number of attempts to find a correlation, and the lack of accounting for all the tests in the evaluation of the significance of the results. There is a good chance of seeing false fortuitous correlations if one examines enough local temperature records.
JES: We have not inflated anything, just reported the result of our investigation and the methods used. The claim that we selected only local series where we did find correlations is wrong. We selected temperature series from places with long records, away from (large) cities, preferably stations at light houses on the coast. The results from all tests are reported in our paper II (SSH12). In the Svalbard paper (SSH11) the analysis was done in more detail for one location, as we demonstrated that the correlation, which for all places were found for yearly data, was significant only for the winter temperatures.
We have rephrased 'significance' to 'importance' - by conducting many trials and only reporting the ones with a good match, one does implicitly inflate the importance of the results. It's sometimes called 'cherry picking'. This is a common error, and we explain why this tends to happen and argue that one may use e.g. the Walker test as a remedy.

BHDCN: When we reconstructed their Table 1 we got nearly the same results, albeit not identical. SSH2011 stated that they based their method for estimating SCL on a publication from 1939 (Waldmeier, 1961), however, more recent work on the estimation of SCLaccount for uncertainties in estimating the true SCL as the sunspot record exhibits stochastic variations around the slow Schwabe cycle. Rather than estimating the SCL from the few data points around the solar minima, Benestad (2005) proposed to use a Fourier truncation to fit the sunspot record and hence use the entire data sample to estimate the SCL. In particular, SSH2011's estimate of the SCL for cycle 23 (12.2 yr) was substantially longer than the estimate of 10.5 yr reported by the Danish Meteorological Institute (based on Friis-Christensen and Lassen (1991) and follow-up studies) and 10.8 yr estimatedby Benestad (2005) (Table 1). Such a long cycle is the basis for their projected cooling (a decrease from −11.2 to −17.2 _C with a 95% confidence interval of −20.5 to −14 _C) at Svalbard over solar cycle 24 (starting 2008).
JES: Several errors here: We have not estimated the SCL, but used the official SCL determined by an international committee based on several parameters and reported by NGDC (National Geophysical Data Center). Their SCL is based on the Waldmeier (1961) definition with certain additions, and is specified as follows: " When observations permit, a date selected as either a cycle minimum or maximum is based in part on an average of the times extremes are reached in the monthly mean sunspot number, in the smoothed monthly mean sunspot number, and in the monthly mean of spot groups alone. Two more measures are used at time of sunspot minimum: the number of spotless days and the frequency of occurrence of old and new cycle spot groups".
The length of 12.2 yrs for cycle 23 is simply the length calculated by NGDC two years after the cycle was finished. We consider this as the OBSERVED CYCLELENGTH. This must be a far better estimate than the length estimated by DMI and Benestad (2005), who estimated the length before the cycle had ended. It should also be remarked that the solar cycle 24 started officially in 2009.

The errors are rather minor and easy to correct. The SCL will always be an estimate, and we have rephrased the text to make it clear that they themselves did not estimate the SCL: "SSH2011 stated based their analysis of SCL on data from National Geophysical Data Center, which uses a method based on a publication from 1939 (Waldmeier, 1961)". The point about SCL not being estimated but is taken from NGDC is corrected in the revised paper. SCL is hard to pinpoint accurately, which is illustrated by the fact that different estimates exist. 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).

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.

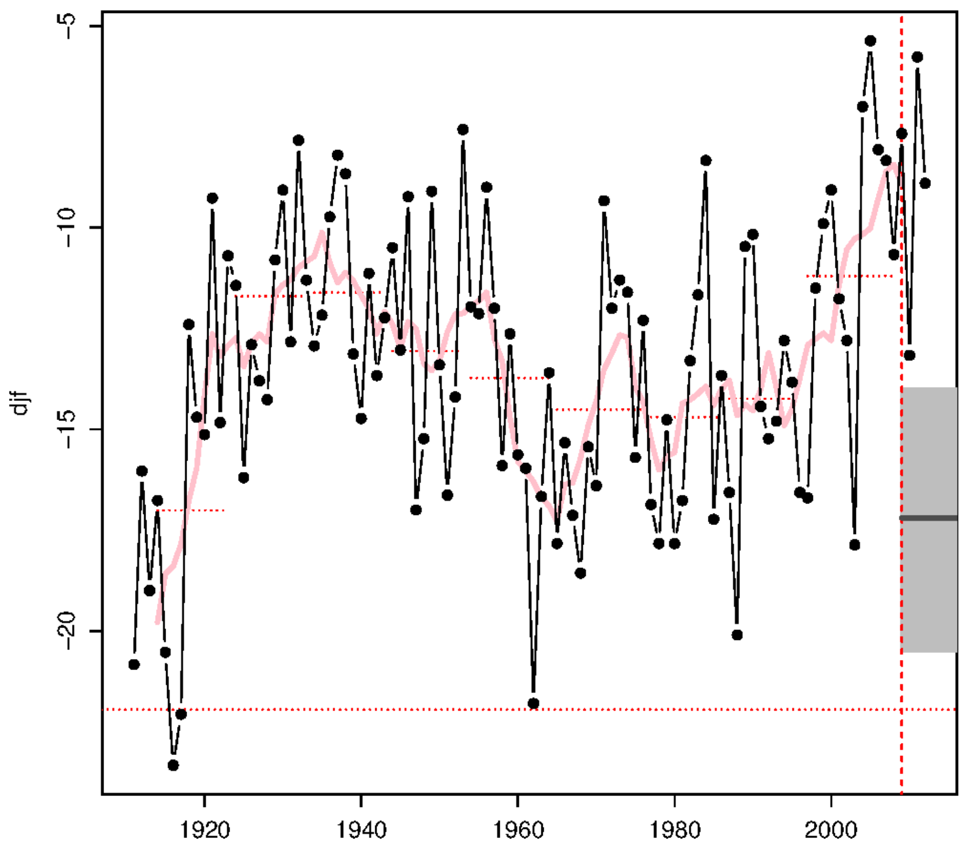
It is not possible to check the estimates of Benestad (2005) for SCL, since table 1 in that paper is nonexistent.

True, the paper has no Table 1, and the confusion is caused by reference to Table 1 in our paper, providing the recipe for how to repeat the analysis in 'replicationDemos' (this reference has now been removed to avoid confusion - there is still several sentences describing how the replication can be implemented). Hence, with computational or analytical skills, anyone should be able to check the number. In other words, experts with hands-on experience with computer codes and data analysis would be able to check the calculations and the results from the R-package.

BHDCN: The observed mean over 2008–2011 suggests a continued warming that reached −9.17 _C as an average for the 4 yr, which means that the mean winter temperature of 2012–2018 (the next 7 winter seasons) must be −21.8 _C for a good prediction. An analysis of 7-season running mean values of the Svalbard temperature reveals that it is rarely below −15 _C and has never been as low as −21 _C since the measurements began.

JES: Two errors here: The year 2008 is included in the sunspot cycle 23 used for prediction – (this cycle ended in December 2008). The temperatures for 2009 -12 should be used for the first 4 yrs observation. The length of cycle 24 is not known. An estimate based on solar parameters indicates that it may be an extremely long cycle of 15-17 yrs. If so – it can be as much as 13 years left, not 7 as written above.

The discussion has been revised and updated for the 2009-2012 period. This does not change the picture. We have also included some lines showing how the results are repeated in replicationDemos (e.g. see figure below:)



The DJF temperature for Svalbard. The dashed horizontal line indicates the mean winter temperature over the next 7-year period needed to fulfill the prediction of a mean temperature of -17.2°C over 2009-2019. The thick pink line marks 7-year winter averages, and the black line the individual winter mean temperature.

BHDCN: SSH2011 used a weighted regression to account for errors of the mean temperature estimates over the periods corresponding to solar cycles. Hence they accounted for errors in the mean estimate, but neglected the errors associated with the SCL, which are more substantial than the errors in the mean seasonal or annual temperature over 10 yr segments.
JES: There are two possible errors in SCL that affects the analysis. The first is in the temperature averages. The definition of winter is the months DJF. If the SCL starts in Jan or Feb, the month of Dec (prev. yr) should belong to the previous cycle. In these cases there will be one monthly temperature associated with the wrong cycle. This happened only once for the series, since cycle 17 started 1944.2. Similar for the year average temperature. We did not split years, so temperatures for cycles starting .5 or later was calculated from the next year, similar if it ended after .5: temperatures for the rest of the year was included in the cycle. We assume that these errors are included in the sigma calculated for the mean temperature. The temperature

averages were taken over 10, 11 or 12 years, as SCL to the nearest year, not in 10 yrs segments.

The second error is in the temperature forecast, where an error in SCL, should be added to the estimated error. We assume that the SCL by NGDC is calculated with the precision of 0.05 yrs, which means an error 1/20 of the temperature change for one year longer/shorter SCL.
We 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. We find it hard to believe.

For Svalbard this would be ±0.05C for the yearly average and ±0.1C for the winter temperature. This is small if we compare with the estimated errors (sigma) in the forecast which for the year average is σ = 0.7C and the winter σ = 0.9C. We disagree with BHDCN that the error in SCL is substantial and larger than the error in the mean and seasonal temperatures.

BHDCN: They also applied a bootstrapping approach to estimate the errors in the correlation coefficients (between −0.52 and −0.97), as they argued that there is no analytical expression to do so...the claim made by SSH2011 that there is no analytical expression for estimating confidence intervals for correlation is false.
KS: The statement by BHDCN that there is no analytical expression for estimating confidence intervals for correlation is false, is wrong. An approach for estimation of the confidence interval for the correlation coefficient which do not give exact values, is the following: The probability distribution function f(r) of the correlation coefficient ,r, is given in: http://mathworld.wolfram.com/CorrelationCoefficientBivariateNormalDistribution.html.

A 95% confidence interval of r consist of the 2.5% and 97.5% percentiles. Here, the percentiles are found by:

2.5% = f(r(low))

97.5% = f(r(high))

where r(high) and r(low) are upper and lower bar of the confidence interval. Because the complexity of the probability function it is difficult to find analytical expressions for r(high) and r(low).

The calculations of the confidence interval based on the mentioned Z-transform is described in:

http://support.sas.com/documentation/cdl/en/procstat/63104/HTML/default/viewer.htm#procstat_corr_sect018.htm

It is important to underline that described bias expression caused by the Z-transform of the estimated confidence limits, have to be taken into account. The variance caused by the Z-transform is approximative.

Their estimate of the errors in the correlation involved 1000 picks of random paired sub-samples from the SCL and temperatures, where the same pair sometimes were picked more than once. A more appropriate strategy would be to carry out a set of Monte-Carlo simulations accounting for the errors due to the SCL (_S) and mean temperature estimates (_T).
KS: BHDCN raise a discussion about usage of the bootstrap methodology contra a special Monte Carlo analysis for estimating correlation statistics. This type of discussion should be directed to other statistical bodies and not to a specific paper where bootstrap methodology is used. The 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.
We are well aware that bootstrap methodologies have been used in the past, but it is also important to appreciate the fact that such bootstrap methods sometimes do not give a reliable result. Here we argue that the particular bootstrap approach in Solheim et al is wrong, and we demonstrate this through our replication of their analysis. We state in the paper that it is important to take into account uncertainties associated with the estimation of the SCL. 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. Bootstrapping does a good job in many cases. Here it was misapplied.

Page 471
BHDCN: The Monte-Carlo simulation also revealed that the SSH2011 correlation estimate was not centered in the simulated correlation error distribution, but was biased towards higher absolute values. ....From just 9 data points, we find it quite incredible that the magnitude of their lower confidence limit was higher than 0.5. These results therefore suggest that the choice made in SSH2011 of SCL was indeed "fortunate" within the bounds of error estimates by getting correlations in the high end of the spectrum. Since SSH2011 made at least 10 different tests (zero and one SCL lag and for 4 seasons plus the annual mean), the true significance can only be estimated by a field significance test, e.g. the Walker test: pW =1−(1−_global)1/K (Wilks, 2006).
KS: The BHDCN approach is to apply Monte Carlo analysis where the standard deviation of the difference between the SCL estimates from SSH2011 and Benestad (2005) are used to the SCL simulations. However, the different quality of the SCL

estimates – especially the weakness in Benestad (2005) SCL estimates, destroy the quality of these simulations.

Solheim et al assume that there is weakness in Benestad (2005) SCL estimates, but do not provide documentation. This assumption seems solely due to the fact that 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.

Another aspect is the quality of the Monte Carlo simulations itself. The Monte Carlo simulations are based on simulations of the temperature. For applying simulation of the temperature evolution, BHDCN calculates the temperature variation by estimation of the mean temperature and adding the estimated standard deviation for each observation in the period. BHDCN does not mention that there are significant uncertainties in the estimates of the mean temperature and in the standard deviations.

We used the same mean and uncertainties as Solheim et al, and if it is true that there are significant uncertainties in the estimates of the mean temperature and in the standard deviations, that strengthens our case and weakens their analysis.

Hence for each of the 80.000 simulations, these uncertainties are generated in every run. Compared to the bootstrap analyses, where each temperature sample is based on real observed data and not constructed observations based on uncertainties in estimated mean and standard deviations. The conclusion is rather simple, the bootstrap analyses presented in SSH2011 has to be recommended.

This argument is not convincing - we disagree.

JES: As mentioned above, we did not "choose" SCL in a fortunate way, we used the observed SCL as reported by NGDC.

The paper does not say they "choose" SCL in a fortunate way, to quote our manuscript: "the choice made in SSH2011 of SCL was indeed 'fortunate' within the bounds of error estimates by getting correlations in the high end of the spectrum."

The full set of tests of yearly averages is reported in SSH2012. In total 16 series were investigated, only one (Oksøy) did not give significant result on the 95% level. In SSH2011 we did a more detailed investigation on one of the series (the one with the largest solar effect, Svalbard) to investigate if the relation was related to seasons. It is not logic to us to use non-significance of 3 of the seasons, to discard the yearly and winter season significance. It may be correct if they are random chosen, but seasons are not random. It may be an odd (or unexpected) result to find that the Sun has most relevance in the winter season, but that may be supported by the fact that heat carried by ocean currents arrive also in the winter. Looking at the monthly temperatures, one get the impression that the low temperatures are mostly found in the period Jan-Apr, while Dec still is a warm month, so a different definition of the winter may give an even more significant result.

Searching for the best results is known as 'cherry picking', unless a proper evaluation is made. The proper way to analyse such a set of records is to apply a field significance test, as explained in our paper. These results suffer from the same methodological weaknesses as SSH2011.

BHDCN also mentions that the correlation estimate was not centered in the simulated correlation error distribution. This may have a simple explanation. Since the correlation was high (0.8) there is less room 0.8-1.0 than between 0.0 and 0.8 for a distribution.
We show in our paper that the explanation is not quite that simple after all.

BHDCN: Solheim et al. (2012) expanded the correlation exercises between SCL and temperature to include several locations in the North Atlantic region. The fact that several of these give similar results can be explained from the spatial correlation associated with temperature anomalies on time scales greater than one month. Their analysis involved 6–11 degrees of freedom, depending on the length of the available record, but since they applied their analysis to both SCL with zero and one-period lag, in addition to a number of locations, they would need to account for the problem of multiplicity and apply e.g. the Walker test. The failure to do so will give misleading results.
JES: We are aware of the spatial correlations, in fact that was one of our conclusions that the effect was strongest (with r2 > 0.5) for stations in or near the North Atlantic – see table 1). We reduced the degree of freedom with one, since we investigated twolags (0 and 1).
This does not take into account the interdependencies between the different stations.

Page 472
BHDCN:The main problem with the analysis presented by SSH2011 was the lack of a convincing physical basis, inappropriate hypothesis testing, the inflation of significance, and a small data sample insufficient to support the conclusions.
JES: Our aim with the paper was to produce a forecast, not to give a physical explanation. The model is clearly defined. If it works will be proven by observations the next 5-10 years. That the effect (of SCL) clearly is strongest in and near the North Atlantic, invites to speculate on a physical reason related to the ocean currents, and heat transported from warmer locations where the Sun warms the deep ocean. Unfortunately the length of data series are limited. Still we tested the method by deleting the last observation in each series, and compared it with predictions based on the remaining. All observed (unused values) were in the 95% bracket.
Removing one data point in analyses of 10-11 year means doesn't provide a strong evaluation.

Page 478

Case 9: looking at wrong scales

BHDCN: The analysis on which Humlum et al. (2013) based their conclusions removed the long-term signal through a correlation between the annual time differences in CO2 and temperature, This procedure removes the long time scales, and emphasises the short-term variations. Hence, Humlum et al. (2013) found the well-known link between 5 El Nin˜o Southern Oscillation and CO2. They then incorrectly assumed that this link excludes the effect of anthropogenic emissions.

OH: We set out to investigate the short-term variations (DIFF12; using two 12-month windows), as the long term variation is the integral of these short-time variations. On the time scale investigated in our study, changes of atmospheric CO2 follows after corresponding changes in temperature. For that reason the changes in global air temperature cannot be the result of changes in atmospheric CO2, while the opposite possibility must be considered viable. If one expands the time window considered from 1 yr to 3, 5 or 8 years, our conclusion still holds true. In short, Benestad et al. attempts to give the reader the false impression that there are mathematical errors in our analysis, which is not the case.

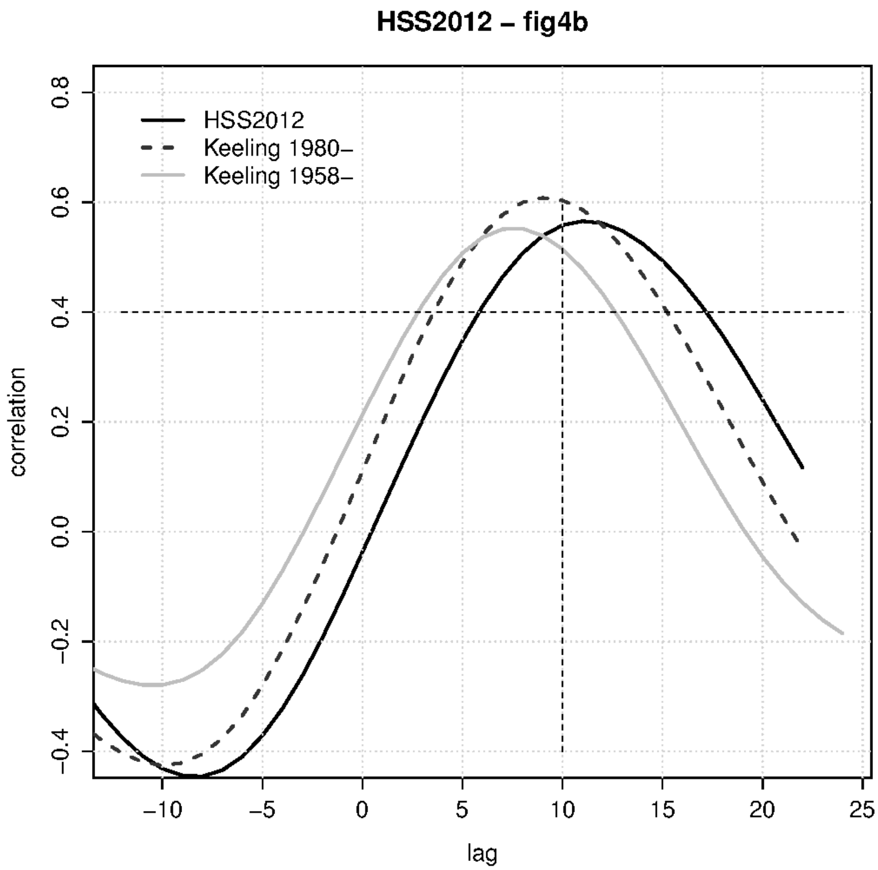
It is well-known that short-term variations such as ENSO drives short-term CO2-changes, but Humlum et al's logic fails when taking that to the long-term trends. This is now explained in Richardson (2013) and Masters and Benestad (2013), and these citations are now included in the revised paper. The revised manuscript elaborates on the flaws, which are mainly of the logical sort. Humlum et al failed to contact the providers of the CO2 data that they used, even though it specifically asks people using their data to check with the providers if they think the use is appropriate. My understanding 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. A follow-up paper on this matter would be beneficial if based on shared source-code for the analysis. The computer code for our analysis is already available in replicationDemos.

BHDCN: Humlum et al. chose to analyse a short series from 1980 describing the global analysis of the CO2 concentrations rather than the almost identical series from Mauna Loa going back to 1958.

OH: We decided to analyse the shorter global data series for atmospheric CO2, rather than the somewhat longer Mauna Loa series, because we wanted to use a global average rather than observations from one single point (Mauna Loa). If one instead uses the Mauna Loa series for analysis, a result entirely similar to that published by us is achieved. Presumably BHDCN are well aware of this, but nevertheless leaves the casual reader with the wrong impression that our analysis is wrong.

This quote from our paper is factually accurate. Our replication was done for both the

Mauna Loa and the global estimate, and there are some differences in the lagged correlation functions - partly due to the different record lengths (figure below). However, the main issue is the use of DIFF12 operator, followed by lagged correlation. The revised paper explains this more carefully.



IN SUMMARY The part of the BHDCN paper which addresses papers with Humlum et al. as author is constructed around a number of misunderstandings or false allegations. In some cases alleged issues discussed by BHDCN by themselves, are not even mentioned in our paper. This is indeed very strange.

There are relevant issues which were ignored in the Humlum et al (2013) paper, such as the the ratio of oxygen to nitrogen (which is reduced through the combustion of carbon and fossil fuels), the carbon cycle, the structure of the oceans, and trace gases used to infer the movement of water masses.

Page 483

Case 13 contamination by other factors

Beck (2010) has repeated his analysis in more detail, and most of the objections mentioned by BHDCN are resolved. He has selected only places where weather conditions are known (wind from the sea, or from isolated islands, ocean crossings etc.). From an estimated total sum of more than 200 000 single samples collected since 1800 in the Northern and Southern hemispheres, he selected 97 404 samples from 901 stations compiled in 87 data files. A thorough literature review revealed that these data are only partly known to climate science and had been rejected and ignored without a thorough validation of the quality, sampling and analyzing methods used. The new dataset contains high quality data including vertical profiles, which can be easily compared to today´s standards. Two periods approximately 70 years apart are identified with slowly rising atmospheric CO2 levels up to values close to 400 ppm. Instead of rejecting Beck´s work with hand waving arguments, BHDCN should look carefully at the series Beck has selected as significant, and compare them with modern measurements.

Local CO2 measurements may be contaminated by local sources, and may not accurately represent 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 their response to case 9 that they chose not to use local CO2 measurements made at Mauna Loa, but argue here that it's acceptable to use measurements since 1800 in Europe that 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. E&E is a social science journal, not a credible climate science journal. At this stage, the E&E paper by Beck is not convincing, especially since there is no information about the different measurements (the link to supporting material is broken).

Case 14: incomplete account of the physics

Miskolzi calculates by a high precision line-by-line code the IR optical depth as function of time based on observed radiosonde data for pressure, temperature, humidity and CO2. The result was that the optical depth is approximately constant the last 50 yrs. And since IR radiation does not care what the absorbers are (CO2 or H2O) – it at least tells us that more CO2 has not changed the radiative transfer. What the physics is to explain this is another question, but data from The International Satellite Cloud Climatology Project (ISCCP) shows that the column of water vapor at 680 -310 mb elevation has decreased from 0.6 to 0.4 cm since 1958 (down 30%). This may have compensated for the extra CO2 in the atmosphere.

A recent paper by Q.-B. Lu (Inter. Journ. Mod. Phys. B, 27, 1350073) concludes: "Remarkably, a statistical analysis gives a nearly zero correlation coefficient (R=-0.05) between corrected global surface temperature data by removing the solar effect and CO2 concentration during 1850-1970. In striking contrast, a nearly perfect linear correlation with coefficients as high as 0.96-0.97 is found between corrected or uncorrected global surface temperature and total amount of stratospheric halogenated gases during 1970-2012."

Many others have calculated 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. Some of the authors on our paper are also submitting a comment on the Lu paper, which is another case of curve-fitting. Nevertheless, a paper by Lu does not imply that the paper by Miskolzi is correct – it is faulty logic to imply so. We have expanded the discussion in our paper to include these points.

Case 15: differences in preprocessing of data
BHDCN: Although a connection between solar activity and Earth's climate is plausible, there is no trend in the recent solar indices that can explain the current global warming.
Use of smoothed values of SCL is discussed also in SSH2012 and we conclude that they are of little value. A test of unsmoothed observed values values give good correlations as discussed in case 4 when a delay of one cycle is introduced. It may be that the lack of trend in the recent solar indices as written above – combined with the reduction of human-made halogenated gases has resulted in the lack of trend in global temperature observed the last 15 years, as concluded by Lu (2013).
As noted above, the Lu (2013) paper is being addressed in a separate paper.

Page 488
Case 17 misinterpretation of spectrographic methods
BHDCN: " claims that the giant planets exert influence on earths´ climate"
HY: No. The paper title is "...influence on Arctic climate". Earth climate is a reference to global climate. Arctic climate is a reference to a regional climate, which is something else.
The text has been revised to refer to the Arctic climate.

BHDCN " Yndestad claims to have identified the lunar nodal cycle".
HY: The lunar nodal cycle has been known since James Bradley in the 17th century, and H G Darwin in the 19th century, and later identified in climate indicators by a number scientists.
BHDCN: "This study is too based on harmonic analysis".
HY: No. The study is based on a wavelet analysis of the longest oceanographic data series in the world. Wavelet analysis is not the same as harmonic analysis. It looks like BHDCN has done a misinterpretation of methods.
OK, wavelet rather than harmonic analysis, although a harmonic analysis would also do for any permanent periodic response.

BHDCN: "Sea-ice and the local Arctic climate is strongly affected by winds and ocean currents"
HY: The study compares data series of the Earth axis position, Arctic ice extent in the Barents Sea, Arctic ice extent in Greenland Sea, Kola section water temperature in the Barents Sea, wind at the Røst island, NAO winter index. The estimated periods in the Arctic oceans, are

confirmed by other scientists. The same periods are identified in sea level, temperature and salinity in inflow of North Atlantic water to the Norwegian sea, and published in: Yndestad H. 2008 William R Turrell, Vladimir Ozhigin. Lunar nodal tide effects on variability of sea level, temperature, and salinity in the Faroe- Shetland Channel and the Barents Sea. Deep-Sea Research I. 55 (2008) 1201-1217.
The periodicities are probably correct, but we question the interpretation of the causes and the method used for attribution.

BHDCN: “Arctic climate involves dynamics with a pronounced non-linear chaotic character”
HY: The wavelet analysis in this study shows that the identified periods have phase-reversals, connected to longer harmonic periods. Phase reversals indicates a time-variant and non-linear system.
OK.

BHDCN: “Case 17 misinterpretation of spectrographic methods”
HY: There is no reference to the misinterpretation in this text by BHDCN.
The title is revised to ’Misinterpretation of periodicities’, which what the section is about.

McKitrick
I should disclose that I was a reviewer of this paper when it was submitted previously to another journal. Since the authors declined to correct most of the problems that led to its earlier rejection, my comments here simply repeat much of what I have already said.
It is true that our paper was first submitted to two other journals, 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 work. 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. There is a conflict-of-interest issue here. We repeat our response to his criticism here - the reviewer is not always right, and there is a real dispute over several issues, many minor.

The BHDCN paper is a bait-and-switch, in which the authors propose a scholarly essay on the methodology of science, then proceed to deliver something quite different: a scattershot of shallow commentary on a list of climatology papers with which they disagree. It is perfectly valid to publish critiques of papers, but they should be written as such, not offered parenthetically in an essay supposedly on another topic. We are asked to take it as “proven” that the authors of this paper have so decisively rebutted all the papers in their Appendix that we can now turn to a philosophical debriefing on the question of why they ever got published in the first place. Yet as proof, all the Appendix offers is a rehash of old, and mostly unpublished, blog posts. The whole

paper is thus a waste of readers' time.
McKitrick thinks that our paper is a "bait-and-switch", but this is his own subjective interpretation. 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 usage of the concept of "agnotology" is confused and contradictory. They introduce the term in paragraph 1 on page 452 as the counterpart of epistemology, that is, as a branch of philosophy.

Then they use it in paragraph 2 as a method ("An agnotological study of the climate sciences can shed light..."). Then it is used synonymously with replication analysis and as a rhetorical technique ("the communication of misleading claims is a case of agnotology." para 2 p. 463). Its use in the heading of Section A4 implies it is a form of misrepresentation. Finally the title itself implies agnotology means learning from mistakes, but the paper does not allege "mistakes", instead it alleges widespread research malfeasance, such as ignoring data that doesn't fit a hypothesis or ignoring known physical theories.
Presumably, "agnotology" means an absence of information and a lack of basis for knowing. Yet all the authors' examples allege the opposite situation, namely cases in which (BHDCN assert) there is so much information, and matters are so decisively settled, that we can now assume the debates are all over and BHDCN won them all. There is no special philosophical issue behind their analysis, it is just garden-variety argumentation, most of it at a very trivial level. The agnotology angle appears to be a contrivance to try and make a weak paper sound erudite.
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 intend our paper to go into a special philosophical issue. The manuscript has been re-structured, and less emphasis is placed on agnotology, although it is still mentioned towards the end of the paper.

The authors repeatedly discuss replication as an essential part of science, insinuating that the papers they critique are at fault regarding disclosure of data and methods. Yet they provide no evidence that non-disclosure was an issue for the papers they study. The authors of the papers they critique appear to have made their methods and data freely available, and BHDCN do not claim that their work was thwarted by non-disclosure. While they point out that replication work is rare, they don't present any case studies in which replication was actively impeded by failure to release data and/or code. Examples of such studies would be Dewald et al. (1986) and Anderson et al. (1994).
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 New Scientist (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.

Worse, BHDCN insinuate that their analysis of MM04 and MM07 was hampered by secrecy, by ending their discussion on page 490 with the statement: "Another problem was the lack of openness and transparency, which prevented finding out why the conclusions in some of these cases differed to attempts to replicate (Le Page, 2009)." This is completely misleading. The Le Page article refers to an unrelated incident involving different authors, whereas the data and code for MM04 and MM07 have always been available, and nobody has ever claimed to be unable to replicate the findings. I have pointed this out to BHDCN in response to their previous drafts and it is very objectionable to see them repeat their falsehood here once again.
McKitrick misinterpreted our reference to Le Page, 2009 – which concerned Scafetta and not MM04 or MM07. It is hard to see how he makes the connection to his own work. There has not been an intention to suggest that their analysis of MM04 and MM07 was hampered by secrecy, and the paper has not claimed that MM04 or MM07 were secretive. The reference to Le Page (2009) was never made in any connection to MM04 nor MM07. The fact that MM04 are replicated in 'replicationDemos' suggests that any claim of not being able to replicate is not an issue ("The replication of the MM04, as done by Benestad (2004), is implemented with the following command lines in R:"). In the revised manuscript we have made this distinction clear by stating that the problem was concerning a specific case of a series of Scafetta paper, reported in New Scientist, where the journalist was witness to the discussion and the irrational refusal to get to the bottom of the issue (LaPage, 2009).

Case 7 (A2.5) refers to McKitrick and Michaels (2004) and insinuates that the results were not tested using a withholding/prediction test, an accusation made even more explicitly on page 490. But Section 5 of MM04 presented just such a test, and Section 4.1 additionally tested the results against the influence of atypical outliers, and in neither case were the conclusions affected.
This test did not account for mutual dependencies due to both spatial correlation and common values for national indices within each country. The paper has been revised and expanded to explain more carefully the important points.

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, but the failure to pass this test had no general implications, as

explained in McKitrick and Michaels reply to Benestad, which BHDCN do not mention.
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 inaccuate, and appears to be an effort disregard the results. In fact, 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. If McKitrick had read the Benestad (2004) paper properly, he'd know that the assertion about different sample sizes 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)". We have revised the paper to explain our points more carefully.

The McKitrick and Michaels (2007) paper (Section 4.5, Figure 2) presented 500 split sample withholding/prediction tests in which 30% of the data were randomly withheld each time and predicted by a model fit to the remaining 70%. MM07 Section 4.2 tested against the influence of outliers. The skill of the model is amply demonstrated by the reported findings, which BHDCN do not mention, even while falsely claiming the MM07 paper was flawed for not doing such tests.
We argue that: "A follow-up paper, McKitrick and Michaels (2007; MM07) involved similar flaws, revealed in Schmidt (2009) who concluded that the basis of their results was a set of 20 correlations for a small selection of locations mainly from western Europe, Japan, and the USA." Again, there is spatial correlation and common national indices, resulting in dependencies. The issue is not resolved in a random split-sample-test, but it is important to ensure that there are no dependencies between the training and evaluation samples. We have nevertheless cut this paragraph from the paper, as it does not add much new information.

Their discussion of spatial autocorrelation (SAC) in the MM07 results omits all the relevant aspects of that debate. Benestad (2004) conjectured, without providing any evidence, that SAC would reduce the effective degrees of freedom in MM04 sufficiently to undermine the significance of the conclusions.
This assertion is false; The evidence was provided through the split-sample test, which is provided in replicationDemos. Furthermore, the failure of this test did have general implications, contrary to McKitrick's view. Even randomly picking data in this case 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. The influence of correlation on the effective sample size should be well-known facts and elementary statistics, and MM04 should in the outset have recognized this issue. We have now explained this more carefully in our paper.

Schmidt (2009), cited by BHDCN, repeated this claim but once again did not test it, and he

confused SAC in the dependent variable with that in the residuals. BHDCN make no mention of the extensive treatment of the SAC issue in McKitrick and Nierenberg (2010), who presented a suite of robust LM tests for SAC on both dependent variables and residuals, and showed that MM07-type model residuals were not affected by this issue, and even if the models were re-estimated with a correction for SAC the conclusions were upheld. They showed, moreover, that Schmidt's regression on GCM-generated data was affected by SAC for which he neither tested nor corrected, and had he done so his own results would be insignificant.
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. The paragraph where this is discussed has been dropped - it's not our major point.

Schmidt's results did not show, contrary to BHDCN's claim, that the MM07 coefficient estimates were inside the model-generated distribution. As is clearly shown in McKitrick and Nierenberg Section 2.2 (emphasis added), they were unambiguously outside the distribution:
With regard to the [claim by Schmidt], the distributions of the coefficients estimated on GCM data do not encompass the coefficients from either the MM07 data set or any other observational grouping in Table 2. In the next section this will be shown after reestimating the model using a correction for spatial autocorrelation. Anticipating the findings, for none of the socioeconomic coefficients does the 95% Confidence Interval estimated on model-generated data encompass the coefficients estimated on observed data. Consequently the null hypothesis as stated by Schmidt (that there is no contamination) is rejected. I am at a loss to think of wording that could be any clearer, yet BHDCN repeatedly make the opposite statement about the findings in question. Finally, Schmidt's remarks about "Japan, Western Europe and the USA" appear only in regards to his Figure 3a which is not part of his discussion of MM07, so its repetition in Case 7 is misleading. It is unacceptable that BHDCN repeat their untrue statements on these matters since they have been presented with the above information in response to both of their two previous drafts.
It suffices here to focus on the McKitrick and Michaels (2004; MM2004) and Benestad (2004) papers, while MM07 and the others are mere peripheral references. We are citing the results from Schmidt (2009), which we find credible and in agreement with Benestad (2004). Dr McKitrick disagrees, but we do not find his arguments persuasive. Nevertheless, this is very peripheral to the discussion, and we have removed the paragraph citing MM07.

To the extent they want to claim that "agnotology" arises from authors deliberately ignoring contrary information, they are themselves serving as striking examples.
McKitrick misunderstands the term "agnotology"; there is also a component of

evaluation and assessment involved, e.g. what information is relevant and what is not. Or which is more reliable? This requires suitable training within the scientific discipline, but it also may require a good understanding of the underlying data, the physics, statistics, and computation skills.

Their discussion of the Douglass et al. paper (Case 6) focuses on the idea that the confidence interval around the mean is not the appropriate measure of the distribution of model results, instead people should examine the range of the data. But the purpose of the literature is to say something about the distribution of GCM outputs on the assumption that they are taken to be varied implementations of the same underlying physics, i.e. that they represent samples of a single data generating process. To characterize the central tendency of a data generating process one uses the first and second moments. That is why Douglass et al. and Santer et al. (and others) have argued about the correct definition of the standard deviation around the mean trend. The range, by contrast, can be made arbitrarily wide simply by running the models often enough, and it is not the appropriate measure for the question being posed.
The issue at stake here is the common statistical way of testing: can one distinguish one number from a sample and say that it belongs to different populations? This is elementary statistics, and in the revised paper attempts have been made to explain this even more carefully. By comparing the observed values to the first moment (the mean), then one does not test whether the value is distinguishable from the numbers from which the mean is estimated. The inconsistencies Douglass et al (2007) are revealed even with the simplest tests. The mix-up between the sample spread 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 incorrect, as described in our paper.

If BHDCN want to argue that the debate is totally misplaced they need to develop their argument in proper depth and address the literature in a scholarly way, not through brief, peremptory commentary. Foster and Rahmstorf (2011) is cited as support for Santer et al. (2008), yet it does not test the model-data mismatch so the usage is misleading. Also they ignore McKitrick et al. (2010) who used longer data sets than Douglass et al. or Santer et al. and applied panel and HAC estimators robust to non-zero covariance and higher-order AR processes. The McKitrick et al. findings were closer to those of Douglas et al. than Santer et al. regarding the significance of the model-observational mismatch, especially in the Correction (2011) that fixed an error in the GISS data. BHDCN make no mention of this, despite the obvious importance they attach to the question of a model- observational mismatch.
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 if we were to expand our replication exercises with further cases. At the present, including this paper will add much substance, other than citing another paper of McKitrick's. Longer trends do not rectify the problems of mixing up the error bar of the estimate of the mean and the sample spread, and the confusion about error bars associated with the mean estimate and the interval for a data sample. The objective of our paper is to show that there are a number of flawed papers that have influenced the public opinion, and provide an interesting case of agnotology.

Their discussion of Long Term Persistence (Case 10) is lacking in technical depth, yet the authors dismiss the work of Cohn and Lins (and others) without even attempting to present a statistical rebuttal. It is difficult to see the purpose of this section. The IPCC and others often use an AR1 error model on which to base claims of trend significance. The Cohn and Lins findings are consistent with a wide range of papers on the subject (e.g. Rybski et al. 2006, Lennartz and Bunde 2009, Mills 2010, McKitrick et al. 2010, plus the many others discussed at http://www.climatedialogue.org/long-term-persistence-and-trend-significance/) that find more complex long-memory and higher-order AR processes in climatic time series, thus showing quite categorically that AR1 models exaggerate trend significance. BHDCN seem to disagree with all this, but do not present their own statistical model, much less defend it. Their statement "All processes involving a trend also exhibit some LTP" is vague and nonsensical. After all, the IPCC uses a trend+AR1 model for all its standard error calculations on the assumption that the data do not exhibit LTP.
For case 10, the text has been expanded and the man issue is more clearly spelled out: to test the hypothesis that unforced trends may arise and resemble the recent global warming, one needs to rule out the role of the forcing. We can show that the presence of forcing influences the autocorrelation function (ACF), and there is one function in replicationDemos that shows this. For the total forcing, we see a Hurst coefficient of ~0.9, and one needs to account for the fact that the past temperature may have been influenced by forcings of different kinds and characteristics. 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. The discussion is also expanded to the connection between (low-dimensional) non-linear chaos and LTP: how to distinguish to similar looking behaviour?

The climatedialogue.org exchange features Benestad and van Dorland trying to argue that forcing trends can induce LTP, hence its detection might be interpreted as evidence for forcing rather than random natural variability. Koutsoyiannis commented: "I agree that (changing) forcing can introduce LTP and that it is omnipresent. But LTP can also emerge from the internal dynamics alone as the above examples show. Actually, I believe it is the internal dynamics that determine whether or not LTP would emerge." And I particularly refer Benestad to Bunde's comment: "When testing to what extent GHG is responsible for LTP, we found it is not, please have a look at our 2004 GRL, where we also specified the methods. It is very unfortunate that

Rasmus does not seem to be able to read this and our other articles on LTP." These are sound views from knowledgeable experts, and Bunde in particularly directs the BHDCN lead author to papers explaining the methods available to them to make their arguments. Yet rather than doing so, BHDCN resort to handwaving and insinuations that the many LTP papers in the literature are forms of misinformation.

Additional discussion has been added in the manuscript regarding this case. There has recently been a discussion on this issue on ClimateDialogue.org, as McKitrick too references. 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 on the climate dialogue blog, McKitrick presents only one side of the dispute. My criticism to the LTP work is that they rarely place the analysis in the wider context, such as known physical inter-relationships (processes) and other geophysical data. This point is now brought into the paper. Furthermore, a missing aspect has been the distinction between non-linear chaos and LTP, which seem to produce similar-looking behaviour. It is well-known that the climate system involves non-linear chaotic dynamics, and we know that the memory of the initial conditions disappears after some time. So it's unclear whether such behaviour may be equivalent to LTP behaviour, or if it affects the LPT analysis. These points are now included in the discussion, and constitute too to agnotology - uncertainty due to different types of interpretations.

The issue here is whether the external forcing influences the ACF in a way that subsequently influences the FARIMA models used in Cohn and Lihns. Additional analysis has been added to replicationDemos, demonstrating that with climate models, the forcings have a clear influence on the ACF. I did not see that Bunde responded to the figure showing that. We also expect that models similar to FARIMA are calibrated to mimic the behaviour of the time series - including the trends. Hence, for proper hypothesis testing, the FARIMA should be calibrated with data equivalent to results from the unforced climate model simulation.

Case 12 refers to McKitrick and McIntyre 2005, which focused on the bias arising from using decentered data in a PCA algorithm that is only valid when the data are centered. BHDCN dismiss the bias as irrelevant, ignoring the fact that Mann et al 1999 placed explicit emphasis on the shape of the PC1 in their analysis, and that many subsequent authors used the biased PC1 in their own reconstructions, and that the PC1 error biased the computation of critical values, a topic which was central to the MM2005 article as well as the later exchange with Huybers. Many salient details of these points were discussed at length in, among other places, the 2006 NRC report (North et al.). In fact, these issues have received so much airing elsewhere that it is hard to see the point of Case 12 at all, especially when the authors resort to a Wikipedia entry as one of their main sources. The authors do not seem to have taken the trouble to properly research the issue, and as such their brief commentary lacks credibility. Nor does it illustrate their elusive concept of "agnotology", it just seems yet another axe to grind at the end of a long, tendentious paper.

References
Anderson, R.G., and W.G. Dewald (1994). Replication and scientific standards in applied economics a decade after the Journal of Money, Credit and Banking project. Federal Reserve Bank of St. Louis Review (Nov): 79-83.
Dewald, William G., Jerry G. Thursby, and Richard G. Anderson (1986). Replication in empirical economics: the Journal of Money, Credit and Banking project. American Economic Review 76(4): 587-603.
McKitrick, Ross R. and Nicolas Nierenberg (2010) "Socioeconomic Patterns in Climate Data." Journal of Economic and Social Measurement, Vol 35 No. 3-4 pp. 149-175.
The discussion of this case has been expanded and it now includes more explanations, starting from the elementary aspects of linear algebra and PCA. It seems 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 and how much of the variance do they explain. This is just a mathematical fact, which we now spell out. 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 McKitrick's claim, The reference to Wikipedia is not a "main source". It was 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. This reference is now dropped in the revised manuscript.

Bothe
1. I welcome every published exchange highlighting failed replications of results in the literature but such comments should be identified as such and published using the appropriate paths. If one manuscript can highlight problems with multiple papers, it's an asset. However, the provided commentary in Benestad et al. appears to be rather superficial and using a too broad brush.
The cases and the R-package are included to make the discussion specific, and their

function is to provide concrete evidence. They are provided in the appendix and are intended as examples of the type of replication we have in mind. In this sense, it is importance of being to the point and being specific about this issue, and hence we can learn from mistakes. These facts are in contrast to Bothe's interpretation: 'shallow' and 'broad brush'. We would like to learn about other papers which provide a discussion along three levels, where one actually discloses the source code and the data necessary to do the replication. Papers as far as we know rarely reach this depth. While each case is specific, we can understand the impression that they provide a wide-range sample of cases, but we have shown that there is a common denominator: their effect on the public perception of climate change, and many of the papers are not independent, but involve the same authors. We also argue that we can learn from all these examples, as they exhibit clear-cut cases of fallacy. Hence, we disagree with the view of being "appears to be rather superficial using a too broad brush", however, we agree that these problems probably are common beyond the 'sceptical' part of the literature. We have made this note in the paper too.

I agree, some shortcomings are common to a number of the mentioned publications, but these problems are unfortunately also common beyond the 'sceptical' part of the literature. Thus, I think, it would be more useful to publish methodological notes, e.g. comparable to the manuscript by Wilks (2006) that Benestad et al. reference. Similarly, the literature indeed lacks a thorough and open review of planetary or lunar influences on climate but the manuscript does not provide one either.

There is some merit in writing a review of several controversial papers - although there are not many such papers in the literature. We hope that our paper will spark a discussion about these topics, which is much needed. In the revised manuscript, we do address the question of bias in terms of the selection of cases here. The list of examples that we chose was a selection based on simple clear-cut cases. For instance, there is no doubt that the results in some papers were based on analysis where part of the data had been removed. Mixing up error estimate for the mean with sample range is too an obvious flaw. The same goes with 'curve-fitting' or the analytical design in one of Scafetta's paper.

One does not always have to write long articles to expose such shortcomings. Short accounts are not the same as being superficial, but it requires that the reader has a good understanding of statistics and analytical design. We aim our paper to fellow climate scientists and experts with hands-on experience with such matters as discussed here. It is understandable if people without this training do not appreciate our work - this is often the case in general. Questions regarding the hockey stick shape of a principal component requires a familiarity by the reader in terms of how the principal components are used and of the linear algebra involved in order to grasp the concepts.

If replication, open data and open methods are the topics at heart of the manuscript, a relevant

commentary should extend beyond a rather narrow group of papers. One could accompany the negative examples by positive ones. Furthermore, one could search for publications whose methods have been criticised but with which one personally agrees. Replication is as or even more important for high profile publications. Possibly there are surprises? That would not be a case of false but rather of correct balance. It is rather unfortunate irony that some people may ask some of the co-authors to lead by example with respect to open methods and open data.
We agree completely that replication is as or even more important for high profile publications, and we hope that our paper will spark a discussion that leads to greater awareness about this issue. It is not easy, and this paper has itself met a great deal of resistance, on a number of different grounds. This is also how science plays out in the real world, and we want to provide some documentation of that. Our paper does not just look at the science, but is also addressing a real concern regarding the different views on climate change between publishing climate scientists and the general public. We want to stick to that, and therefore we think that our list of cases are sufficient for this purpose. However, we hope this inspires similar papers with a more general view on replication.

2. I doubt whether a rather superficial paper helps in closing the gap in understanding. This holds especially if the two “sides” of the gap are classified in such a binary and close to condescending manner. This aspect of the manuscript in its current form should take place in the media, in the blogosphere and possibly in everyday exchanges (e.g. in a real-world or a virtual pub). I guess, it may be possible to rewrite this aspect in a more suitable manner. However, from my point of view, this aspect is best served by providing thorough commentary and discussions on the supposedly controversial topics within the scientific literature (see above).
Again, we disagree that the paper is superficial. We may disagree, and we would like to point out that the reviewer may not always have the correct answers, and here we are somewhat in line with Hanekamp’s concern regarding non-epistemic consensus and ‘anti-pal review’. It is important to discuss the disputed points, and sometimes the disagreement will prevail. We think the paper is in a suitable form, and we believe it is likely to boost discussions about these issues. Our paper serves a useful purpose as long as there are no serious fallacies on which is is based.

3. I would welcome a study or a commentary on agnotology in climate (change) science (e.g. in Climatic Change), a commentary or essay on replication (e.g. in Climatic Change or Nature Climate Change), and also philosophical essays on either topic (in a suitable philosophical journal). But the present manuscript doesn’t fit any of these categories.
It may be difficult to introduce new ideas and bring in new ways into the science community, and we think this paper is a point in case. The view concerning categories is based on expectations about what is proper in terms of a scholarly paper. Science is not so much about conformity, however, the important aspect is whether the science is correct: the empirical evidence, theories, and the logic. One flaw can topple a theory or

falsify a hypothesis, and this is what we show with our cases. The list cases makes our paper specific, rather than superficial. Which of the cases do you not find convincing?

Nevertheless, I agree with the authors that it is not necessary for a manuscript to be strictly scientific or philosophical. However, if authors mix both aspects, their writing has to be strong on both accounts. The present manuscript is weak on either. In addition, Since the paper is apparently conceived as a review article it requires a much clearer structure.
The structure has been revised.

In the end I just second the final comment of anonymous reviewer 2: I recommend that the authors restructure their paper around common themes (e.g. logical fallacies or common methodological mistakes), write it clearly and concisely, avoid snarky comments against denialists (irritating though these characters might be!) and work on a coherent presentation, instead of publishing a laundry list of replication studies and wrap it in dubiously written philosophical verbiage. This will make for a much stronger contribution to the scientific literature.
The manuscript has been revised, and we have introduced a common-theme structure, and the philosophical aspect has been toned down. All the cases are moved into the appendix for specific documentation, providing a more coherent presentation. There has been no specific examples of "snarky comments", and it's not clear what is being referred to - we do not want to be seen as being snarky.

Hanekamp

Some of the comments from Hanekamp are somewhat difficult to respond to where they are not to-the-point but more vague and focussing on peripheral matters. He accuses our paper for failing logic, words which we think describes some of his own comments. In some cases, his points are irrelevant.

Our paper tries to make sense out of the controversial issues that have plagued climate sciences in the recent years, and presents a number of cases which suggest that we can learn from mistakes. The set of cases was selected because they have received some attention in the public debate about climate change and because they are clear-cut examples where it's easy to demonstrate mistakes.

Hanekamp accuses our paper for having "rudimentary approximate of the philosophy of science" which "is only reserved for those authors that do not share 'the view presented by the mainstream climate research community". We think this statement is false, and we will challenge him or anybody else to find other similar (flawed) cases which are in line with the view of the climate science community. There have been some, such as a bug in the climateprediction.net simulations, a corrigendum in Nature (doi:10.1038/nature02478), error in the GISTEMP data, and glaciers in the Himalayas (2nd work group report, 2007), but these have been acknowledged and dealt with. We have learned.

The revised paper has tone down the philosophical aspect, and hence some of the comments are no longer relevant. We acknowledge that our revised paper will still be controversial for many of the commentators; however, we think that the fact that some issues are disputed does not mean that a paper on these issues is unpublishable. We also note the point raised by Hanekamp that reviews are not always perfect and the reviewer is not always right - sometimes for non-epistemic reasons. Sometimes people do not like a paper due to irrational reasons, and we have notes a number of subjective arguments here, especially as none have dived to the depths of the associated R-package and tested the data and the methods provided.

Comments

The central question that is raised immediately by the paper is whether agnotology exclusively is the playing field of those who criticise the ostensible consensus position on human-induced climate change or whether it is found in the latter as well, e.g. as to manufacture some kind of consensus.

Hanekamp asks the rhetorical question: "The central question that is raised immediately by the paper is whether agnotology exclusively is the playing field of those who criticize the ostensible consensus position on human-induced climate change or whether it is found in the latter as well, e.g. as to manufacture some kind of consensus." This is not a logical conclusion from our discussion, and our answer is "no" - agnotology is not exclusively the playing field of one side. This is an insinuation and an attempt of Hanekamp to read between the lines, based on failing logic. The second suggestion about manufacturing some kind of consensus too is a red herring. Science is about discussions, discourse, and replication, and a scientific consensus emerges when one explanation is far more persuasive than alternative ones. We think this paper is an example of that, and we are disappointed that Hanekamp himself wants to stifle this debate by recommending the rejection,

Thus, on what grounds do the authors think that human-induced climate change is scientifically well established and thus beyond any scrutiny? (Indeed, which position in any scientific discourse is beyond scrutiny?)

The consensus view is certainly not beyond scrutiny, but we will expect that such scrutiny should be solid and persuasive – not papers where data have been removed, based on mere 'curve fits', or involve logical flaws. I think we agree that no scientific discourse is beyond scrutiny, but the critical thinking should not end at the first stage of scrutiny. Indeed, we argue in our paper that "A continuous replication of published results and dissemination through scientific fora can nevertheless contribute towards a convergence towards the most convincing explanations". The revised paper makes an even stronger point that science is never settled and that thorough scrutiny is welcome.

Consensus seems the only reference ('evidence') in the paper whereby the whole exercise

consequently becomes question begging.
This is an inaccurate representation of our paper.

The basic aspect of science, however, is proof (either empirically or logically), not assumption.
Here we are in complete agreement, and I'd like to point out that we also have provided proof in terms of our replication of the cases we refer to (the R-package), and we provide the data and source code for our analysis. Alternatively, the results are conditional upon a set of stated assumptions. In his comment, Hanekamp himself makes a set of assumptions, which are not clearly stated (and not logical). It would help the discussion to make these explicit (e.g. about symmetry).

Thus a simple referral to the consensus of some view on human-induced climate change simply cannot do.
We agree completely, and we think Hanekamp must have misread our article. Our position is the question why are there differences between the consensus and a number of scholars with different views, and whether we can understand the reasons behind them.

As a result, the argument as a whole collapses into circularity. This is the critical and unacceptable asymmetry the paper introduces devoid of any clarification or evidential basis. Incidentally, this criticism is emphatically not an implicit reference to any purported agnotological 'misbehaviour' on the part of representatives of the consensus view such as the IPCC or the US NRC.
This does not make sense. Please explain more carefully the reasoning. Can the criticism be more specific?

Logically, the approach the authors chose must by default capture the symmetry of the debate they themselves initiate. More to the point, they need to prove as carefully as possible that their approach can only apply to those that criticise the alleged consensus view on human-induced climate change rather than simply assume it with reference only to some kind of consensus. Consequently, no amount of examples they might bring forward in support of their claims can be regarded as proof of their argument whatsoever.
We think Hanekamp has misunderstood our objective. Agnotology, learning from mistakes (now, the title has changed) tries to make more sense out of some of the central controversies that have emerged around climate sciences in the past years. We want to show that there are a number of papers promoted into the public sphere which have influenced the public perception about climate change, and that many of these papers fail the test of scrutiny.

Worse, the corollary of the authors' argument seems to be that one requires only a single basic agnotology example in the discourse the authors assume to be free thereof in order to refute the paper in its entirety. (It seems reasonable to assume that someone already did produce such an example or examples.)

This argument does not make sense. Please explain more carefully the reasoning. Can the criticism be more specific?

No attempt thereto, in line with some form of falsification, has been made. So, the very reproducibility they require from others does not apply to their own work, which introduces another kind of inacceptable asymmetry.

This argument does not make sense. Is this a different form for asymmetry to above? Indeed, our paper discusses falsification (the cases provided) and we provide data and open source code so that others can test our analysis. Hence, we expect that others will be able to test our methods – it is important to require the same from ourselves as from others. The revised paper now also discusses the issue of 'asymmetry' in the sense of sampling controversial papers.

Hanekamp presumes that there must be some kind of "symmetry" in the peer-reviewed publications, but he has not provided any evidence for such a symmetry: "This introduces asymmetry in the paper which will be the focus in the subsequent comments." Hence his objection to our paper rests heavily on this asymmetry, and we could ask if he expects symmetry (of similar kind to that he expects in climate sciences) regarding the question of tobacco and cancer, continental drift, quantum physics, or general relativity.

On a more general note, one can always invoke an 'asymmetry' argument, i.e. that the comments posted on our paper are all negative because it provokes a response - positive comments tend to come as e-mails. In the revised paper, we include a discussion about symmetry (basically equivalent to the sampling strategy in terms of cases - it's relevance depends on the questions one wants to address).

Furthermore, the very a priori (ex cathedra might be a more accurate descriptor) exclusion of the consensus view on human-induced climate change from the analysis the authors bring forward can result in nothing more than fallacious appeals to authority and popularity (of which the term 'consensus' is but a variation).

Hanekamp should provide some counter-examples. So far, he only assumes that a symmetry exists, and assuming is not a valid proposition (as he too observes). We do not see ourselves are '*ex cathedra*', but that it's a useful distinction (consensus vs outliers) to make in order to try to understand different views. This is a bit off-topic subjective interpretation that fails logically - we have never presented the idea that any consensus should be "exonerated from the very enquiry". This does not make sense, because Hanekamp makes an illogical leap to this position. We have revised to manuscript by spelling out (on the risk of being seen as snarky) that this is not the issue. Of course, science is never settled and of course, consensus should be subject to enquiry - the issue is whether opposing arguments are more convincing or whether they rests on faulty logic or errors.

Again, why would the consensus view as defined by the authors be exonerated from the very enquiry the authors initiate other than by an appeal to authority and popularity?

This interpretation is strange. In our paper, we argue "The merit of replication, by re-examining old publications in order to asses their veracity, is obvious. Published results in particular should be replicable, and access to open source codes and data should be regarded as a scientific virtue that facilitates more reliable knowledge. Results are far more persuasive if one can reproduce them oneself, although replication of published results requires scientific training, numerical skill, and mastering of statistics. One concern is that modern research is veering away from the scientific virtue of replication and transparency." We do not want to limit this to one side, but it must be a continuous process for any view. We will make this point clearer in the revised version of our paper.

Besides, why would scientific peer-review within the sphere of the purported consensus-view on human- induced climate change be any guarantee for scientific quality, which the authors think is lacking in critical literature? Scientific majority views could afford quite proficient mechanisms to keep other views at bay precisely through peer-review, also known in this case as pal-review.

We have not taken any position about the quality in the consensus-view, and not made a general statement about the opposing view (to call the cases we chose 'critical' may be inappropriate, as they contain severe flaws as exposed in our demonstrations). The point about keeping other views at bay is relevant to our agnotology paper, and a case emerges through the discussion comments provided by Dr McKitrick. We agree that this is a concern, and we need to rely on the editor's digression to judge the various referees' reports. We include some discussion on this issue in the revised paper, noting that openness/transparency, replication, and the reviewing of past results is a good measure to reduce the risk of non-epistemic biases (along similar lines as proposed in this paper). Having said this, Hanekamp would not like this paper to be published (according to his own comment) because it opposes his own view about how science should be conducted.

The dominant epistemic community within climate science could well hinder freedom of research and publication, which could result in impeding certain research themes that are not regarded as in line with the dominating paradigm and thereby ignored for less than charitable reasons.

This would be unfortunate, but there is difference between valid conclusions and stifling of opposing views. We don't think that Hanekamp sees the differences, and there is no logical connection between the two. In the physics community, there is a preprint-sever (arXiv), with no review. There is also a wide range of journals, and often one needs to move onto another journal because of one difficult referee. This happened in this case (McKitrick). Sometimes the road to publication is long-winded due

to some particular reviewer's personal preference. On one point, we do agree, though: that the reviewer is not always necessarily correct (we take this as implied by Hanekamp's concern here).

(1) Think of for instance the history of the theory of continental drift proposed by the meteorologist Alfred Wegener in 1912 and the ad hominum opposition levelled against him.
(2) The Russian economist Nikolai Dimitrievich Kondratiev is best remembered for his theory of 'long waves' or 70-year cycles in which economies reflect the rise and fall of dominant technologies. Less well known is his prediction of the inevitable superiority of capitalism over Marxist planning. Correct perhaps, but a dangerous conclusion to reach and publish in Stalin's Russia and one which saw Kondratieff executed. His misfortune serves as a stark reminder of the power of dominant paradigms (ideologies) to resist change.
We hope Hanekamp would welcome our paper - as a contribution to the discourse around the science concerning climate change. He does not need to agree with us, but then our contribution may spark a scientific/philosophical debate. It would be a shame if he wants to stifle the debate by rejecting our paper. We think a better avenue is rather that he subsequently submits his objections to our ideas - and explain why he thinks our logic is failing. This would harmonise better with his own arguments based on Wegner and Kondratiev. Who knows - maybe our description is the right one after all? There is no logical connection between our paper and the cases presented by Hanekamp, as we are not in the position of power that can censor or stifle others. We would not want that either, and in our paper we do the exact opposite - we shed light on controversial papers in an attempt to learn from the controversies. As we explain in our paper, the authors of those papers are free to respond through subsequent publications.

The suggestion here is not that the alleged consensus view on human-induced climate change is wrong and other views are correct. It simply poses the problem of peer-review and its quality safeguard.
This is a fair point, and we do discuss this in our paper. One issue is obstacles to new and controversial views (assuming persuasive evidence, logic, and valid conclusions) and the other issue is sloppy reviews. We hope our paper can trigger more discussion about the peer-review.

The authors, again, make no attempt to support their asymmetric assumption: what seems to be wrong with those who criticise the professed consensus view on human-induced climate change is assumed to be right in the consensus discourse.
We are making no such assumptions, but we provide some examples for which we argue contain flaws - which is demonstrated through replication described in the appendix. This is how science should work. Hanekamp seems to make some assumptions about asymmetry and about veracity. Again, the revised manuscript discusses the issue of symmetry.

A final note here is related to the purported well-established scientific suppositions on climate change. This is an odd position to take (to say the least) from the perspective of the history and philosophy of science. One can be entirely indifferent to this subject to nevertheless be completely astonished by such a view. Chemistry is one of the oldest scientific fields in the history of science, and chemists have seen many theories come and go during centuries of research. How can a young field such as climate science be so sure of itself, as the authors in their paper seem to be, other than to be totally ignorant of the history and philosophy of science?
Hanekamp presumes that the lessons made in chemistry must be true for other disciplines, and he interprets our paper in a surprising way. Sure, we are aware about the history and philosophy of science (e.g. Thomas Kuhn, Karl Popper, and Emmanuel Kant) and that history has overturned previous paradigms. However, we are not saying that the consensus is right (however, we can assume that one explanation has swayed most of the scholars in the field of climate science because it seems to be the most persuasive view – a fair definition of consensus). We are discussing whether we can learn from controversial papers by replicating the work and check the details. We are trying to show how science can be utilized to reduce uncertainty – this even works for a "young" field as climatology (comparing their age is a bit of a red herring – climatology does not start from scratch, as chemistry did).

Discussion and conclusion
With the presence of multiple kinds of asymmetry as highlighted above, this paper essentially is unpublishable. Why would any reader be persuaded by the claims made in the paper other than being forced to accept fallacious arguments from popularity (consensus), authority and some examples of supposed failures of those who criticise the alleged consensus view on human-induced climate change, whereas the consensus view is ex cathedra exempt form any kind of agnotological scrutiny?
Hanekamp's assumption of symmetry needs to be established before he can say that his arguments are valid. Demonstrations of how science work (through openness, transparency, replication, and testing, as we have done here) normally is the most persuasive strategy. We disagree that the consensus view is '*ex cathedra*' exempt from any kind of agnotological scrutiny, and we are not taking this position in our paper. In fact, the consensus has been under intense scrutiny since the first IPCC reports, and the ClimateGate episode revealed a surprising resourceful opposition to the notion about AGW. So far, the 'critics' have not managed to find much substantial. This is a strawman argument from Hanekamp, devoid of logic and evidence.

Inductively, such an argument simply fails and is an affront to scientific logic. Of course, objections can be raised here. It might be argued, for instance, that not every detail can be discussed in one manuscript on human-induced climate change. Put differently, the consensus view as defined by the authors cannot be discussed and defended in depth in just one paper. That is obviously the case, but if that is so (and I do not doubt it) then the whole exercise is vacuous. In terms of education I will definitely use this manuscript in my philosophy of science

classes, whether published or not, but not in the way the authors might have envisioned it. It can only be used as an example of how some scientists maltreat very elementary aspects of understanding and doing science. It shows that universities should require far more of their students with respect to arguments, reasoning (logic), and evidence.
This paper fails utterly on all three counts.
Here we have different views. We say the same about Hanekamp's comment, which is a subjective opinion, lacking evidence and logic. Many of his points are imagined. He presumes a symmetry and complains about asymmetric presentation of cases. We will challenge him to find this symmetry. Without it, his whole line of reasoning collapses. In any case, he has mistaken the objective of the paper when he gets hung up on the question of asymmetry.

Ellestad

Discussion comment SC C291 was submitted through Dr Jan-Erik Solheim, however, we will presume it reflects Ellestad's views as his name is on the letterhead of the comment.

Their analysis do not reflect the methodology required to draw the given conclusions. In addition, there are numerous misquotations and partly irrelevant elements on which they base their critique.
It's surprising to read that Ellestad thinks our analysis does "not reflect the methodology required to draw the given conclusions". Already in our abstract, we state that "we show that a number of papers in the scientific literature contain severe methodological flaws", and on p.455 we state "We attempt to provide a comprehensive review by examining the methods used in an effort to replicate the results of a range of different studies, highlighting the value of replication". Indeed, the cases presented in the appendix and the R-package replicationDemos provide our methods and analysis. So, exactly what methodology does he think would be required to draw conclusions like ours? He needs to be more specific. This is also the case regarding the alleged misquotations (some provided by Solheim have been corrected).

In Norway we have experienced reference to the BHDCN-paper in the public debate where it more easily can be accepted as a sound scientific paper which is not the case. This probably reflects the main intention of the paper and why also popular scientific articles, names and organizations that are not actively taking part in the scientific work, are included in the paper.
Ellestad should elaborate on why he does not think that our paper is scientifically sound. Has he found errors or mistakes in our analysis? IWed be happy to discuss such, as this would be in the spirit of agnotology: learning from mistakes. We'd like to refer him back the my response to the comment from Solheim et al., and we maintain that our results are solid, based on the cases that we have analysed. Ellestad thinks we intend to leave an impression of being concerned and knowledgeable about the principles of

agnotology in climatology. Indeed, we work as climate scientists at national meteorology services, and we are both concerned and knowledgeable. We speak to journalists and lay people, we make measurements, we analyse and program computer code, we follow the news, we read scientific papers, we attend scientific conferences, we pursue climate science, and we publish our results in the peer reviewed journals.

BHDCN start, as we have experienced several times with Benestad, with a description of the general principles of agnotology which intend to leave the impression that they are very concerned and knowledgeable about the principles and therefore have the responsibility to take on their writings. This is tactics not science. The paradox becomes clearer when BHDCN defend the "Hockey stick" (in section A3.2 case 12) which has been thoroughly analyzed and criticized by top experts and now is supported only among a gradually smaller group of scientists.
Ellestad makes the curious remark about tactic not science and a paradox associated with the “Hockey stick”. He also purports that the work by Mann et al., (1998) is “supported only among a gradually smaller group of scientists” [sic!]. No proper documentation nor evidence provided.

The organization Klimarealistene is attacked In the following, I will concentrate on the activities of the organization Klimarealistene which is a non-political organization, including members with various political views. The object of the organization is to participate in the public debate and distribute relevant information about climate as a supplement to the official information.
Ellestad proceeds to declare that “klimarealistene” is non-political, without defining what he means by that. This is followed by a description of how the organization distributes material on climate research (e.g. from “ClimateGate”), a booklet on “alternative” views about climate change to schools, and points to “severe weaknesses of the AGW theory”.

However, he does not mention the fact that the late politician Svenn Korseth sr. from the far right “progressive party” used to be the leader of this organization, and that associates of “klimarealistene” have taught politicians from this party “climate skeptical” soundbites (by the way, we never claimed in our paper that the organisation was political, but perhaps we should?).

We have produced booklets which describe the well accepted natural variations like ENSO, PDO, AMO, sunspot activity and their influence on cloud formation as well as the Urban Heat Island (UH)-effect. In addition, we have pointed to the severe weaknesses of the AGW theory and the climate models like the missing hotspot finger print, flattening of the global mean temperature over the last 16 years and the severe discrepancy with the model calculations as well as the low level of understanding of cloud formation and the questionable positive forcing of water vapor. We consider this to be highly relevant and generally accepted as important climate phenomena.
Much of this is covered by the IPCC, however, the leaflets misrepresents many of these topics and takes a view that is closer to the Heartland Institute’s publication ‘NIPCC’.

None of the authors of the leaflet has a strong publication record on these natural phenomena.

What is controversial, but not necessarily wrong, is our statement that natural variations seem to be the dominating mechanism. Among the given critical statements by BHDCN is that the relation between sunspots and climate has failed to stand up to new data, and information presented on sunspots and cosmic rays, which include the work by CERN, has been characterized as nonscientific.
We agree that what is controversial is not necessarily wrong, but we also think that controversial issues should be subject to thorough investigation. This is the basic message from our paper. Our paper has not characterized any work by CERN as 'nonscientific', nor relationships between solar variability, cosmic rays or climate. This is a red herring.

Our distribution of the booklet to the schools reflects the low standard of information provided by the school authorities. The information the primary school children in Norway get on natural climate variations is only two sentences about the Milancovic relations. The rest of the climate variations are stated to be AGW. This is a far stronger statement than the conclusion in the IPCC 2007-report that "most of the observed increase in temperatures since the mid-20th century is very likely to have been due to the increase in anthropogenic greenhouse gas concentrations." We interpret ´most´ as more than 50 %, which means that up to 49,9% of the warming can have other causes, which should be explained to the pupils.
This is a strange interpretation. There is a tacit knowledge about natural variations, and their presence is often taken for granted. There is a great deal of knowledge concerning the phenomena giving rise to such oscillations, but there are also some unknowns. Nevertheless, there is little evidence suggesting that the present global warming is due to such natural causes - hence a scientific consensus. The leaflet distributed to the schools diverges from this consensus.

For Norway being close to and partly within the Arctic area and with a Gulf stream (AMO) along our coast and into the Arctic, we find it highly relevant to inform about the observed systematic variations over years which certainly influence our climate. This is done in the official publication "Klima i Norge 2100" (pages 44-46; Fig. 3.3.2, and 3.3.3), which shows temperature flattening the next 20 years, but not communicated to the general public.
The need to improve the dissemination to the general public is acknowledged, but much of the information is propagated through the media which has its own agenda.

Other remarks
Klimarealistene do not give input to research activities as stated under Case 9. We publish popular science about climate and participate in the public debate. Our meager income, based on contributions from members, is mostly used for printing, postage and public meetings. Our written material is produced without salaries.

Ellestad argues that "klimarealistene" do not give input to research activities (as stated in Case 9; Humlum et al., 2013). However, in the acknowledgement of Humlum et al., (2013), the authors state that they "are finally grateful for informal discussions in Oslo with Drs. O.H. Ellestad [himself], O. Engvold, and P. Brekke. Also Drs. T.V. Segalstad and H. Yndestad has for long time been important sources of information and inspiration." - all associated with "klimarealistene". The irony cannot be ignored here: the very comment of the leader of "klimarealistene", Ole Henrik Ellestad, was submitted through Jan-Erik Solheim. In fact, there is evidence suggesting that Ellestad's assertion is false: an e-mail that went astray and ended up in Benestad's mailbox because it mentioned Benestad, where Ellestad provides a remark on the Humlum et al. (2011) and Solheim et al. (2012) papers discussed in the appendix of our paper. It reveals how the papers were viewed as a piece in the strategy of "klimarealistene" (I only disclose some small excerpts here, translated to English):

From: Ole Henrik Ellestad [mailto:ole.henrik.ellestad@hotmail.com] Sent: 30 May 2011 17:50 To: j.e.solheim@astro.uio.no; Kjell Stordahl; Oddbjoørn Engvold; Ole Humlum; Subject: Articles
"[I] think this with truncation and using the last part for evaluation, then take the forecast
is exemplary in a field that is so full of subjective opinions... With these two [papers] and the two in FFV, klimarealistene are significantly better equipped in the battle by having documented good agreement between observations and mechanism. It takes a lot to dismiss this and I see the argument that they are struggling. [I] reckon that

Benestad and ... are partly on a level no one would even dare to approach if they had not been backed by the IPCC and the press".

We will maintain that Ellestad and "klimarealistene" are engaged in a propaganda campaign. On February 3rd 2012, Ellestad wrote a letter to the director at MET Norway in an attempt to gag Benestad, after he had commented on the Humlum et al (2011) paper on the website of a Norwegian newspaper. (A scanned copy is available on-line: https://drive.google.com/?usp=chrome_app#folders/0B5ZHm1tjzEtDWjhWZmxlQzVVSWc)
references:

Humlum, O., Stordahl, K., Solheim, J.-E., 2013. The phase relation between atmospheric carbon dioxide and global temperature. Glob. Planet. Change 100, 51–69.
Humlum, O., Solheim, J.-E., Stordahl, K., 2011. Identifying natural contributions to late Holocene climate change. Glob. Planet. Change 79, 145–156.
Mann, M.E., Bradley, R.S., Hughes, M.K., 1998. Global-scale temperature patterns and climate forcing over the past six centuries. Nature 392, 779–787.
Solheim, J.-E., Stordahl, K., Humlum, O., 2012. The long sunspot cycle 23 predicts a significant temperature decrease in cycle 24. J. Atmospheric Sol.-Terr. Phys. 80,

267–284.

In summary we find the approach of the critique by BHDCN highly impropriate.
We would not expect that Ellestad and "klimarealistene" would appreciate our paper, but attribute their view to a dislike of criticism rather than incorrect statements.

Rypdal
At the surface of things, the conceptually simplest approach to detection of anthropogenic global warming should be the estimation of trends in global surface temperature throughout the instrumental observation era starting in the mid-nineteenth century.
We think some of the comments here are a bit off-topic - more details below. However, the discussion on LTP and trends has also been expanded to accommodate for some of the points made by Drs Rypdal as well as McKitrick.

These kinds of estimates, however, are subject to deep controversy and confusion. The authors of Agnotology contribute to this confusion by their shallow statements about "circular reasoning." The only evidence offered is an R-script which shows that the autocorrelation function (ACF) develops an LTP-like tail if one adds a trend to a stationary, noisy, non-LTP signal.
We agree on the point about the controversy about trends, and to us, this brings in the agnotological nature of this question and confusion reflecting the associated uncertainties. We also agree that in general that there may be several equally valid definition of trends. Here, the question really is the whether the observed long-term change in the global mean temperature (let's call it a 'trend') can be explained in terms of natural/internal variations or if it requires an external cause. Our analysis examines *two specific papers* which look into this question (Cohn and Lins, 2005; Franzke, 2012), both of which base their analysis specifically on FARIMA type models or "phase scrambling" to represent the null-model against which the trends are assessed.

We disagree that our only evidence is an R-script showing ACFs, but our main point is the logical flaw in calibrating a model on data that contain forced response to use as null-models to test the presence of a forced trend. Moreover, the approaches mentioned by Rypdal may be suitable for simulating noise with memory, but we argue that they do not represent appropriate null-models because they are trained on data records which have been subject to both natural and anthropogenic forcings in the past. In fact, the proper approach would be to demonstrate that the *total* forcing (GHGs, solar, volcanic, etc) is devoid of similar long-term behaviour, and that the time series models are not 'fooled' by their temporal structure. This was not done in any of the two papers that we discuss.

This is known to anyone who works seriously with LTP, and is a major reason why the ACF is never used to demonstrate LTP in short time records. It is my view that there are several, equally valid, definitions of the notion of a trend.

We also appreciate Rypdal's point about the use of the autocorrelation function (ACF) in studying long-term persistence (LTP), yet the autocorrelation is a central aspect of long-term persistence and power spectra. It is not in our interest to elaborate the exact models for the LTP here (we assume that the model development itself was thorough and state-of-the-art), as it suffices to note that the failure to account for the forced response invalidates their use in this context. If the null-models such as FARIMA and "phase scrambling" are to mimic the past data, they will be sensitive to auto-correlations as they need to match the ACF. The estimation of their parameters are furthermore subject to sampling errors, and we argue that a substantial change in the ACF indicates that the null-models must account for the LPT-character in the forcing fields in order to provide reliable answers.

Which one that will prove most useful depends on the purpose of the analysis and the availability and quality of observation data. At the core of the global change debate is how to distinguish anthropogenically forced warming from natural variability.
For calibrating the models, one needs the unforced variability - subtracting the anthropogenic still leaves a forced component. We do have a good idea of what the current forcing is, and hence can account for these, so the missing part is the internal variability. We can easily show that the *total forcing* data used for input to climate model simulations have this kind of behaviour (H~0.98), and there may be a risk that the natural component has a LTP character that is imposed on the global mean temperature. An R-script to show this has been included in replicationDemos. Furthermore, we already know that the climate system involves a non-linear chaotic dynamic, which is the most common explanation for the internal variability.

As the authors of Agnotology point out, a complicating factor here is that natural variability has forced as well as internal components. Power spectra of climatic time series also suggest to separate internal dynamics into quasi-coherent oscillatory modes and a continuous and essentially scale-free spectral background. Over a vast range of time scales this background takes the form of a long-term persistent, fractional noise or motion [1].
We agree on the view that there are natural and internal variations, but we disagree on how the suitability of applying these models depends on the context. We also agree that short-memory may not be most appropriate for assessing trend, however, the null-models must not include forced response.

Hence, the issue is threefold: (i) to distinguish the climate response to anthropogenic forcing from the response to natural forcing, (ii) to distinguish internal dynamics from forced responses, and (iii) to distinguish quasi-coherent, oscillatory modes from the persistent-noise background.
We will propose the use of LTP models that takes care not to mix up the signal and noise (distinguish between forced response and internal 'noise'). One strategy could be to use paeloclimatic records and data from the time before the industrial revolution (ii).

This, however, will contained forced response from natural drivers. Another approach could be to analyse time series where we do not expect a trend to be present, and yet represent the type of natural fluctuations that are expected to take place in the real world. One candidate would be data based on the barometric pressure, as we expect the atmospheric mass to be constant (e.g. AMO, ENSO or the AO).

It is furthermore not sufficient to look at one single index or series, but one needs to explain the comprehensive picture: we know that the global temperature is just one manifestation of a more general situation which involves ocean heat content, sea level, the cryosphere, and the hydrosphere.

This conceptual structure is illustrated by the Venn diagram in panel (a) in the figure. Panel (b) in this figure illustrates three possible trend notions based on this picture. Fundamental for all is the separation of the observed climate record into a trend component (also termed the signal) and a climate noise component. The essential difference between these notions is how to make this separation. The widest definition of the trend is to associate it with all forced variability and oscillatory modes as illustrated by the upper row in panel (b). With this notion the methodological challenge will be to develop a systematic approach to extract the persistent noise component from the observed record, and then to subtract this component to establish the trend. The physical relevance of this separation will depend on to what extent we can justify to interpret the extracted trend as a forced response with internally generated oscillatory modes superposed. If detailed information on the time evolution of the climate forcing is not used or is unavailable such a justification is quite difficult. In this case one could construct a parametrized model for the trend based on the appearance of the climate record at hand and physical insight about the forcing and the nature of the dynamics.
We have expanded the discussion on trends.

The next step could be to estimate the parameters of the trend model by conventional regression analysis utilizing the observed climate record. The justification of interpreting this trend as something forced and/or coherent different from background noise will be done through a test of the null hypothesis which states that the climate record can be modeled as a stochastic process with certain memory properties. Examples of such processes are persistent fractional Gaussian noises (fGns) or fractional Brownian motions (fBms). For comparison one can also test the null hypothesis against a conventional short-memory notion of climate noise, e.g., the first-order autoregressiveprocess (AR(1)).
This was not done in the two papers that we discuss in our paper. such an extension is out of scope for our paper.

Rejection of this null hypothesis will be taken as an acceptance of the hypothesis that the estimated trend is significant, and will strengthen our confidence that these trends represent identifiable dynamical features of the climate system.
OK

The authors argue that the value of this kind of analysis of statistical significance is of little interest since the result depends on the choice of null model for the climate noise. One can, however, test the null models against the observation data, and here analysis seems to favor the fGn/fBm models over short-memory models. Bayesian iteration can be used in this process. This is not circular reasoning, but a systematic approach to hypothesis testing and establishment of knowledge.
We do not believe we have argued that. We say that the null-hypothesis must represent the noise and not be contaminated by the signal. Otherwise, the logical aspect breaks down.

If forcing data are available over the time span of the observed temperature record we can utilize this information in a parametrized, linear, dynamic-stochastic model for the climate response. The trend then corresponds to the deterministic solution to this model, i.e., the solution with the known (deterministic) component of the forcing. In this model the persistent-noise component of the temperature record is the response to a white-noise stochastic forcing. The method is described in a recent paper submitted to J. Climate where only exponential and scale-free long-range persistent responses are modeled, without allowing for quasi-coherent oscillations [2]. The approach in that paper adopts the trend definition described in the second row of panel (b) in the figure.
Our paper is not discussing Rypdal's own recent work submitted to Journal of Climate (which probably does not deserve to be in the list of cases for which we argue are flawed and lead to misguided conclusions), and we merely focus here on two published papers (Cohn and Lins, 2005; Franzke, 2012), for which we argue that the null-models do not represent the unforced variability because they were calibrated on past data which contain a response from external forcing. Hence, their use in hypothesis testing involve circular logic. This does not mean that we dismiss the presence of ~60-year variations, and we certainly do not dismiss in advance Rypdal's work as "cultural production of ignorance".

Here the trend is the forced variability, while all unforced variability is relegated to the realm of climate noise. It is possible, however, to incorporate forced and natural oscillatory dynamics into such a response model. The simplest way could be to add the response of a forced, damped harmonic oscillator to the scale-free response. These extra degrees of freedom would add an oscillatory response to the deterministic forcing (this would be a forced, oscillatory response), but also an oscillatory response to the stochastic forcing which would be interpreted as an internal oscillatory mode.
This is out of scope

According to the approach described in that submitted paper we have to classify all deterministic forced responses as trends, implying that a trend defined this way is not necessarily slow. For instance, the irregular sequence of volcanic eruptions provides a shot-noise like forcing signal. After having estimated the parameters of the forced response model using the full forcing data

and the observed temperature record, the residual can be analyzed to assess the validity of different noise models. The responses to fast components in the forcing (like volcanic spikes) will be shifted to the forced response, rather than being incorrectly represented as parts of the internal noise. The test of different noise models via analysis of the residual will therefore give more correct results in the forced-response model than the trend-fit approach.
This is out of scope

The lower row in panel (b) depicts the trend notion of foremost societal relevance; the forced response to anthropogenic forcing. Once one has estimated the parameters of the forced-response model, one can also compute the deterministic response to the anthropogenic forcing separately. One of the greatest advantages of the forced-response methodology is that it allows estimation of this anthropogenic trend/response and prediction of future trends under given forcing scenarios, subject to rigorous estimates of uncertainty.
This is out of scope

A systematic approach to estimating the significance of the linearly rising trend in global land and ocean temperatures, and the of the 60 yr oscillation observed in these records, throughout the instrumental can be found in a manuscript that we are about to submit [3].
This is out of scope

I feel provoked by to have this work dismissed in advance by the authors of Agnotology as ”cultural production of ignorance.”
It would be understandable if this was true, but we have never proposed that Dr Rypdal’s work has contributed to a ”cultural production of ignorance”. We do not even mention his work in our paper.

Rypdal II
This discussion is dominated by a clash between two factions of climate science which both adhere to pre-world war scientific paradigms. Both ignore well-established in-sights about the dynamics of complex natural systems originating from Kolmogorov’s discovery of the importance of scaling laws in physics and brought to maturity by Mandelbrot.
We see this differently, and disagree with Dr. Rypdal’s subjective interpretation. This is also off-topic, and not really relevant to our paper, and the purpose of this statement is not immediately clear.

Nicola Scafetta advocates the view that oscillations in the climate system (including their phase) have to be predictable. But there is no physical justification for such an assumption. A toy-model counterexample is a damped harmonic oscillator subject to a stochastic force. The oscillator could be a natural mode in the climate system and the weather stochastic force could be the forcing of global temperature from atmospheric systems. This stochastically forced oscillator has a preferred frequency, but the phase will drift chaotically. If the physics of a climate oscillation is of this nature any theory that predicts the phase must be wrong, as will any theory

that contends it can predict what is unpredictable. My personal impression is that many of Scafetta's theories belong here.
There are many if's here, and this is not relevant for our paper since we are not discussing these aspects. Moreover, the phase in chaotic systems, such as internal climate variability, too is unpredictable after some time.

Scafetta contends that any theory (e.g., the GCM-type climate models) that cannot predict the phase of climate oscillations "are flawed in the sense they are missing something". The basis for his conclusion, however, is an unspoken assumption that the oscillation is not internal with a stochastic drift of phase, but driven by, and synchronized with, some (mystical) external forcing. This is one of a myriad of conclusions in Scafetta's work that are implicitly based on untold biases and assumptions.
Our angle on this is different, but it is not inconsistent with Rypdal's view here. However, this is not relevant to the discussion of our paper.

Rasmus Benestad, on the other hand, recognizes the existence of chaotic unpredictability, but believes that this precludes the existence of long-term persistence (LTP).
No, this is not the position made in our paper - we criticise the strategy adopted by two specific papers where the null models may be contaminated by the signal.

Such persistence may give rise to large internal variability on long time scales that can be incorrectly interpreted as forced trends.
This is understood, but such persistence also has a physical cause, be it non-linear chaos or slow response.

This, and other strange ideas about the dynamics of complex systems can be found in dr. Benestad's blog contributions in the debate about LTP on climatedialogue.com. Benestad misses the well-known fact that even though microstates can be unpredictable due to chaos, complex systems ubiquitously exhibit self-organization and order on macroscale. What about the red spot on Jupiter? According to Benestad it does not exist because Jupiter's weather systems are chaotic.
This is off-topic and not relevant for the paper. Furthermore, this statement is false, and a figment of Dr Rypdal's imagination. Of course, chaotic systems can lead to self-organised states, such as the red spot on Jupiter. We also see that here on Earth, such as cyclones. Rypdal's interpretation is based on faulty logic and misconceived assumptions. Rather, we wonder how one can distinguish chaos from LTP, and are they equivalent? The one assumes that all states depend on all predious states (Armin Bunde) whereas the memory is lost after some time in chaotic systems. This is a topic connected to agnotology.

He claims that the omnipresence of self-similar scaling in many geophysical systems is an illusion because the various systems are governed by different physics. By this statement he

dismisses completely statistical physics where deep universal laws emerge from statistical principles and largely independent of first physical principles.
This is a false statement - where does this come from? There is no dismissal of statistical physics, especially not in the paper. This is completely off the mark.

The existence of emergent laws, arising on increasingly higher level of description of complex systems, and which cannot be deduced in a straightforward way from the microscopic laws of physics, is perhaps the most fundamental new insight in physics since quantum theory.
OK, but again, off-topic.

I recommend reading of the introductory chapter in the recent book by S. Lovejoy and D. Schertzer: Weather and climate: emergent laws and multifractal cascades.
OK, but again, off-topic.

In different ways, Scafetta and Benestad are both "denialists" in the sense that they keep on refuting to accept modern scientific developments.
This statement is based on false allegations and misinterpretations. Besides, this is not directly relevant for the paper under discussion.

Ironicallly, by perpetuating ignoring these insights they both contribute to the cultural production of ignorance. Benestad seems to have learnt very little about long-term persistence during the seven years since he wrote his first blog on RealClimate the Cohn and Lins paper, so in that context he does not seem to be the right person to write about "learning from mistakes." There is need for debate around the paradigms underlying climate science. This paper, and this discussion, has not served the purpose very well.
There are differences in opinion here, and the case is not as clear-cut as Rypdal presents it. Several of the LTP papers have not been all that persuasive, however, Rypdal and his fellow LTP-supporters almost seem a bit religious about LTP. There is a reason why LTP is not used in weather forecasting, and it is important not to ignore the wisdom derived from meteorology ("chaos") in ones attempts to understand the intrinsic noise ("weather"). Dr Rypdal is sure that he is right, but the issue of LTP is not the whole story here. We are not dismissing LTP as such, and the paper focusses specifically on two papers which happen to involve LTP - the take on these is that the null-models used in these studies were inappropriate because they were trained on data which included the forced response (both natural and anthropogenic). There is a big leap to details about LTP processes, and Dr Rypdal's comments are not making logical sense. We have revised the discussion to avoid similar misunderstandings in the future.

Scafetta I
Benestad response to the C. Loehle objections is quite vague. In my extended response I uncover numerous math and physical errors present in Benestad et al. (2013). So I do not

repeat. Herein I just highlight a couple of issues.
1) About the quasi 60-year cycle observed in the global surface temperature and used in Loehle and Scafetta (2011) and in other Scafetta's papers, Benestad states "For noisy geophysical data, it is hazardous trying to identify cycles when you only have a small number of them (2), and from such a curve you cannot really attribute much physical significance"
It is evident that Benestad's statement does not demonstrate anything.
This cited statement explains the position that we take in our paper, which is also demonstrated through the accompanying R-package 'replicationDemos'. Scafetta, the co-author of the paper that we criticise, is not making a convincing case here.

Benestad can not disprove that the global surface temperature since 1850 is characterized by a major pattern with a quasi 60-year oscillation, which is evident at naked eyes to any unbiased person and extensively demonstrated in Scafetta's papers many times.
This is not the issue here - the question is whether this is a part of a sustained pattern on the long term, and whether this is caused by variations of similar frequencies in the sun.

Benestad's statement is severely misleading because it gives the impression that Loehle and Scafetta did not consider longer sequences than the global surface temperature records which cover only 160 years. On the contrary, Loehle and Scafetta referenced numerous papers using climate proxies covering several centuries that demonstrate that a quasi 60-year oscillation is one of the common patterns that characterize climate records. This is also clearly shown in Figure 4 of our paper Loehle C. and N. Scafetta, 2011.
Figure 4 in that paper is based on individual proxies, and the 60-year cycles aren't even good fits. It's simply not convincing.

G. Bulloides abundance variation record found in the Cariaco Basis sediments in the Caribbean sea since 1650 [Black et al., 1999]. B) tree-ring chronologiesfrom Pinus Flexilis [MacDonald and Case, 2005] as an index to the PDO. Both records show five large quasi 60-year cycles since 1650.
We are not saying that there are no variations with decadal time scales, but our point is that the method applied to the temperature record is inappropriate because that particular record is too short to provide reliable answers. The other proxy indices are not equivalent to the instrumental global mean temperature record, and if they were, part of the work would be to see if the phase was consistent between all these.

Other more advanced figures are present in more recent papers published by Scafetta which also include the solution of the secular trending problem.
This doesn't provide much support to this case - Dr Scafetta needs to be more specific.

It is unclear whether Benestad's failure to properly understand Loehle and Scafetta (2011) about the quasi 60-year oscillation is due to the fact that he did not read our manuscript or he is explicitly trying to misrepresent our paper for the purpose of misleading the readers of the journal.

We believe that we understood the method correctly, and we also think it's appropriate to criticise misapplied analysis. To forward allegations about attempts to misrepresent for the purpose om misleading is not a very convincing defence of one's position - if we misread the paper, it could also be that the paper was not sufficiently clear. As we said earlier, the science is never settled, and science is made by trial and error - and discussions. We hope to learn from mistakes. It's also ironic that Dr Scafetta brings up the topic of misrepresentation, seeing his own recent publication in PRP (http://www.pattern-recogn-phys.net/1/37/2013/prp-1-37-2013.html). A comment on this paper has now been accepted (with minor changes - submitted version on arxiv: http://arxiv.org/abs/1306.2011). One of the reviewers responded: "I have to agree with the author of the comment that the way Scafetta (2013) discusses the Benestad & Schmidt (2009) paper is dishonest and misleading. It is somewhat surprising that this was not recognized during the review process of that paper".

2) About the Soon and Baliunas case and ClimateGate emails. Contrary to what BHDCN2013 lets a reader to believe, cases such as Soon and Baliunas (2003) are very complex, as documented for example here:
http://en.wikipedia.org/wiki/Soon_and_Baliunas_controversy
http://newzealandclimatechange.wordpress.com/2011/11/27/climategate-2-and-corruption-ofpeer-review/
For example, the Wikipedia article says that "Jones replied Mann that "I think the sceptics will use this paper to their own ends and it will set paleo back a number of years if it goes unchallenged. I will be emailing the journal to tell them I'm having nothing more to do with it until they rid themselves of this troublesome editor", referring to de Freitas." And "By May the journal's editors Hans von Storch and Clare Goodess were receiving numerous complaints and critiques of the paper from other scientists, to such an extent tht they raised the issues with de Freitas and the journal's publisher Otto Kinne. In reply, de Freitas said they were "a mix of a witch-hunt and the Spanish Inquisition". Note that the accusations against de Freitas (the editor handling Soon and Baliunas (2003)) were unjustified, as demonstrated by Otto Kinne (the director of the journal) here http://wattsupwiththat.com/2011/11/28/a-response-from-chris-de-freitas/

We do not find this point credible.

Scafetta II

These comments by Scafetta are bit of a 'filibuster' (despite the instruction to be concise), and many points are off-topic. We will address the points that are directly

relevant to our paper, mainly respond to Scafetta's "in brief" comments and conclusions rather than the lengthy report which often veer off the topic.

This comment contains a response to Benestad et al. (2013) where the authors critique a set of papers that they dislike and make a number of unjustified accusations toward their authors. I am going to demonstrate that, despite their "good intentions", Benestad et al. (2013) is filled with misconceptions and/or falsehoods, and with severe mathematical and physical errors.
Scafetta accuses us for making a number of "unjustified accusations" toward the authors of the papers we have selected in our study. This is a misrepresentation of our work – in fact, we provide a discussion for why we think the papers draw misguided conclusions based on replication. We do provide the source code of our work (replicationDemos). Scafetta is the author of several of these papers and he is a contributing author to the NIPCC report discussed in our paper. We are not surprised that he dislikes our paper.

In brief, this paper is a very superficial and unprofessional internet-blog style study.
This is a subjective opinion, and we disagree.

The arguments advanced by the authors simply originate from poor reading and understanding of the critiqued works and general lack of mathematical, statistical and physical knowledge.
We think it's rather the opposite, which we now can document. The recent comment Benestad (2013) exposes Scafetta's own poor reading at best, or in worst case a deliberate and dishonest misrepresentation.

They even promote their critiques by claiming that the IPCC climate models are NOT supposed to reconstruct the correct phases of the temperature patterns!
The internal variability is due to non-linear chaos, where memory of the initial conditions is lost after some time. Hence, the phase of the variations therefore is stochastic. This is a well-known fact within climate sciences.

Simultaneously, the authors do not use their same logic to critique the alternative theories that they favor and advocate such as the IPCC AGW. In fact, one important aspect of the scientific method is comparing alternative theories to determine which one better agrees with the data. Benestad et al. (2003) carefully avoided this direct comparison.
We are looking at the methods of a selection of papers here. This comment is a bit off-topic in this context.

Often, the authors simply highlight secondary apparent discrepancies of the critiqued theories in reconstructing the data with an "absolute" precision claiming that such minor discrepancies invalidate the proposed theories. However, the authors do not provide any alternative theory capable to better interpret the major patterns of the same data.
We will argue that we do not simply "highlight secondary apparent discrepancies of

the critiqued theories in reconstructing the data with an “absolute” precision claiming that such minor discrepancies invalidate the proposed theories” - instead, we look at aspects on which the conclusions hinge.

Moreover, despite their “good intentions”, Benestad et al. (2013) do not really demonstrate anything because often they do not even make explicit the functions they use or their data analysis results or figures. They simply provide some R-routines that in their opinion the reader is supposed to run by himself to find out what happens.
This argument is false - the fact is that replicationDemos contains the open source code for the functions, which make anyone who is capable of programming in scripting languages able to examine the algorithm line-by-line (R, S+, Matlab, or IDL). Non-experts, however, will of course struggle, but that is general for all disciplines.

Therefore, a real direct comparison with the critiqued works cannot even be made in most cases. In history of science the same flawed superficial logic has been often used by those who have opposed the emerging physical theories proposed since the 16th centuries from Galileo to Einstein and beyond. This work does not serve a scientific purpose, but a political one.
Scafetta associates his work and situation with that of Galileo to Einstein, and argues that our paper “does not serve a scientific purpose, but a political one”. This association is misplaced and the accusation that our paper has a political purpose is unwarranted. Besides, a direct comparison can be made through our replicationDemos package, and if e.g. Scafetta himself had made his own code available [1], we could even compare his code with ours.

Introduction

Benestad et al. (2013) critique some papers including a few authored by me (e.g. Scafetta, 2010, 2012a; Loehle and Scafetta, 2011; and other studies) that claim that the anthropogenic global warming theory as advocated by the IPCC (which says that humans contributed about 100% of the warming observed since 1900, as also explained below) is somehow “erroneous”. However, data analysis and an increasing number of papers are establishing that the climate system is very likely characterized by large natural oscillations. Moreover, contrary to Benestad et al. (2013) claims, the criticized papers used not only global surface temperature records since 1850, where these oscillations (e.g., the quasi 60-year oscillation) are quite evident, but also much longer climatic records covering millennia and centuries.
This is a misrepresentation of the analysis discussed here, which was based on the instrumental global mean temperature. There are some proxy indices which show multidecadal variations, but that is not the the issue here. We look at the analysis carried out by Scafetta and others and test their methods.

Simultaneously, data analysis establishes that the IPCC models do not reproduce these oscillations also during the period 1850-2013 that these models are supposed to reconstruct accurately. For example, the temperature standstill after 2000 is missed by the models, but also the strong warming between 1850 and 1880 and between 1910 and 1940 and the cooling from 1880 to 1910 and 1940 to 1970 remain unexplained by the models.

This is a misrepresentation - the GCMs are expected to describe the response to external forcings and the statistics of the internal variability (amplitude, geographical and temporal structure). Scafetta makes assumption about the cause of the global mean temperature hiatus, but if the mechanisms causing it are not represented in the models, there is no reason to expect the model to reproduce it. The decadal variations are most likely due to internal variations, for which the phase is stochastic, but there are also other explanations (deep-ocean heat uptake, aerosols, stratospheric vapor). We have added some discussion on this topic, as it is relevant to agnotology.

These models are, therefore, very likely flawed in the sense that they are missing something (e.g. natural oscillation mechanisms) while overstating something else (e.g. the effects of anthropogenic emission climate forcing).

This claim is wrong due to ignorance about the nature of models and the climate systems, and false assumptions. Furthermore, we are not really discussing “IPCC models” in our paper (the IPCC has no models, but he means the climate model simulations presented in the IPCC reports), but rather the method used to analyse these model results presented in one of his papers. Nor are we discussing the climate sensitivity in our paper. The testing of the climate models have been done elsewhere.

Once these oscillations are taken into account, Scafetta demonstrated that about 50% of the observed 20th century warming can be interpreted as due to the detected natural variability, and his proposed model also well forecasts the standstill after 2000. This result implies that the correct climate sensitivity to CO2 doubling is about half of the average 3 oC value currently implicit in the IPCC models. The real climate sensitivity at the observed scales should therefore be about 1.3-1.4 oC and likely between 0.9 oC and 2 oC. Similar results are being more recently confirmed by researchers simulating the 60-year temperature oscillation with the quasi-60 year Atlantic Multidecadal Oscillation such as Tung and Zhou (2013) on PNAS that Benestad et al. (2013) do not cite nor critique with the same zeal. As a consequence of the lower climate sensitivity to GHG emission, Scafetta concluded that the same 21st century IPCC emission scenarios would produce far less alarming 21st century warming projections than what currently proposed by the IPCC and its advocates such as Benestad et colleagues: see also Scafetta (2010, 2012a, 2013c).

This is a bit off-topic. Our paper examines the methods and how the conclusions drawn in the selected studies hinge on the choices made in the design of the analysis.

In brief, I do not believe that BHDCN2013 can be accepted for publication for two major reasons: 1) BHDCN2013 naively critique a large number of papers authored by numerous people published in the peer reviewed scientific literature without adding anything to science. If BHDCN2013 believe that their arguments are scientifically correct they should submit proper critical scientific comments to the original journals and let the criticized authors to write proper responses so that the readers may properly evaluate the arguments by considering both sides. Sadly, both anonymous referees failed to properly emphasize this elementary point, despite the fact that BHDCN2013 is so poorly written and argued that both referees disliked the paper for other reasons and also suggested its rejection in the present form. This point is, however, partially mitigated by the fact that Earth System Dynamics allows open comments.

Scafetta has his own idea about how science is progressing and what is 'proper' procedure concerning criticising past work. We argue in our paper that it is also necessary to look at a number of papers as a group to see if there are similarities. We do not argue against comments (e.g. hence Benestad, 2013), and we believe there is a need for both a broad view of controversial papers as well as case-specific comments. Furthermore, this discussion paper allows the authors of the original papers to comment on our paper, just as what Scafetta has done. Hence, his major point number 1 does not apply. We also disagree about our paper being poorly written – this is Scafetta's own subjective view.

2) BHDCN2013 contains numerous misconceptions and/or falsehoods in addition to philosophical, mathematical and physical errors that cannot be fixed without making their paper completely useless.

We disagree with Scafetta's point #2 about our paper containing "numerous misconceptions and/or falsehoods in addition to philosophical, mathematical and physical errors that cannot be fixed without making their paper completely useless". Se further details below.

in his comment McKitrick has revealed that this same paper was already rejected by another journal and that BHDCN2013 have not truly addressed the issues raised by those referees that yield the rejection of their paper. And in its reply Benestad revealed that the paper was apparently submitted and rejected not once but twice!

For the reference to McKitrick, please see our response to his comment. We see that Scafetta has a strong interest in our paper being rejected because it exposes some fundamental weaknesses in his own work and papers on which the NIPCC is based and both Scafetta and McKitrick have associations with the Heartland Institute, also mentioned in our paper (http://blog.heartland.org/2011/09/heartland-institute-counters-al-gores-24-hours-of-reality-climate-panic/). In both cases, there was one reviewer (McKitrick was one of them) who strongly opposed the acceptance of our paper - we think it was for non-epistemic reasons.

Therefore, I do not see how the editor of the Earth System Dynamics can accept BHDCN2013 by ignoring: (1) the overall negative review of his two anonymous reviewers, although these reviewers were quite cynical toward the criticized authors; (2) the negative reviews this paper received elsewhere; (3) the rebuttals of the accused authors without also publishing (free of charge) the responses from the criticized authors such as the present one.

Scafetta too is taking an advocating role here. We are not surprised that he dislikes our paper, but we do not think he has very strong arguments.

2. My general "agnotological" philosophical impression The general impression that I had is that Benestad, Hygen, van Dorland, Cook, and Nuccitelli have not understood at all the criticized works and engaged in "straw-men" and "red-herring" tactics to mislead the scientific community and society about ongoing frontier research that they personally dislike up to the point that they try to defame the critiqued authors with a number of undemonstrated "accusations".
Our work is not as Scafetta presents it: 'Simply arguing that there might be some error here and there is not a "demonstration" that the error truly exists'. In fact, we have done his analysis and showed why he has reached incorrect conclusions. We are not looking at "straw men", using "red herring" tactics, or presenting "personal doubts". In his defence, Scafetta picks our differences between 60 or ~66-year oscillation (or ~20-year period), which relates to the question whether the earth's climate is subject to the influence from planets in the solar system. The point made in our paper is that there are some broad spectral peaks not even centered on the frequencies cited in Scafetta's work, and it is naïve to attribute these to planetary forcing the way he does.

It is quite unclear to me what BHDCN2013 want to demonstrate with this figure, and a reader should be very careful because this figure shows something "good" and something "bad": the "good" part is my model; the "bad" part is what BHDCN2013 added. Indeed, contrary to their "accusations" that they do not reproduce my figure, it appears to me that they well reproduce my figure 5b (Scafetta, 2012a) :a reader should note that my figure is made of the red, the blue, the black and the grey lines and the error bars curves after 2010 and everything looks "good". The only difference with my figure is that they added their thin green - model dash line whose meaning is not explained. It appears to me that such green-model may be the one obtained with the so-called "best harmonics" reported in "red"at the bottom - right of the figure
The figure shows that there is a mismatch between the test in Scafetta's paper and the figure - now more carefully explained. We have more carefully described the curves in Fig. 2 in the revised manuscript. Scafetta judges that his fit is superior based solely from a visual inspection, but a proper evaluation needs to take into account the number of fitted parameters. His fit involves more fitted parameters (e.g. a quadratic trend model), which will result in a closer description. That should not be confused with a better model – over-fit is indeed discussed in our paper. Here Scafetta implicitly acknowledges that we actually replicate the work discussed, even though he tries to give the opposite impression elsewhere (see below).

Most of their critique refers to the scientific problem of "replication", which is evidently an important part of science. However, the correct way of proceeding in an objective scientific critique is first to accurately replicate the result and acknowledging the logic of the critiqued

studies within their own “full” hypothesis, which demonstrates that the critics well understood the critiqued study, and then demonstrate whether factual errors are present. Simply arguing that there might be some error here and there is not a “demonstration” that the error truly exists.
This is a misrepresentation of our work.

However, often Benestad et al. construct “straw-men” arguments based on partial and misleading presentations of the supporting arguments used in the critiqued works. They also do not report equations, figures and tables with their data analysis results that can be point-by-point contrasted versus those reported in the criticized works. For some curious reason, the authors think that such tedious work, which is necessary for providing scientific demonstrations and to make explicit the facts to a reader, should not be their primary responsibility but it should be left to the readers themselves by using some R-codes that Benestad et al. provided!
This is a misrepresentation of our analysis. Scafetta does not think our calculations were clear, but their recipe is provided in the R-packages ’replicationDemos’. His remark “a reader is left to just “trust” their words that they have done this and that” is completely off the mark. Furthermore, our equation is so simple (it’s in the R-code) that we thought it would not be needed. We do presume that the readers will have some skills in computer scripting and numerical analysis. The results from Scafetta’s tables were digitally copied from a PDF-version of his paper.

Moreover, the demonstrated “errors” must be objective and so serious as to invalidate the analysis and the scientific conclusion of the critiqued studies. It is evident that minor irrelevant discrepancies in data analysis as well as inappropriate application of the propose methodologies outside the physical constraints of the original analysis, do not invalidate the interpretative scientific logic proposed in the critiqued studies.
We argue that in the selected cases, there are major flaws, as well as more minor mistakes.

Small discrepancies are typical when slightly different data or slightly different analysis methodologies are used to reproduce some result. On the contrary, BHDCN2013 focus their critique emphasizing secondary irrelevant data analysis details or systematically misapplying the adopted mathematical methodologies outside the physical time-scale of validity proposed by the authors. They do this often using red-herring tactics, while neglecting the major scientific message contained in the works they critique (e.g. the existence of large natural climatic oscillations at multiple scales from the 60-year cycle to the millennial one not captured by the IPCC models) that, evidently, they could not properly disprove.
Some of the minor issues are included to explain incorrect interpretations, however, the major issues are well-described, e.g .presented in Figures 1 & 3. One major conclusion in Scafetta (2012) was about model-observation similarities, which we document was based on flawed logic. This is elaborated further in the revised manuscript.

BHDCN2013 mislead the reader by arguing that only two 60-year oscillations exist since

1950 and that this fact "demonstrates" Scafetta's argument "wrong". They even plot in their Figure 4 a "demonstration" using the ENSO signal from 1980 to 1990 (!), without realizing that such an argument is nothing but a red-herring fallacy because what the ENSO does from 1980 to 1990 is an irrelevant topic presented in order to divert the attention of a reader from the original issue referring to the quasi 60-year oscillation.
The ENSO-analogy demonstrates that the method used by Scafetta in general is not appropriate for studying cycles. However, it is true that ENSO is not equivalent with the global mean temperature, nor are the paleoclimatic indexes of the Pacific Oscillation, Atlantic Oscillation, or Indian summer monsoon (which have different phase relationships). The question is not whether there are natural variations present with timescales in the range 50-80 years, but the methodology that Scafetta used to represent the global mean temperature. Furthermore, it's particularly the association to an astronomical origin which we criticize, in addition to the curve-fitting. We also show that the search for astronomical signal in the climate models is logically flawed, for reasons explained.

Hence, Scafetta's attempt to demonstrate that the climate models fail to reproduce the ~60-year variations is misplaced and based on an invalid approach. Scafetta never explains properly the physics of the astronomical influences, other than some vague resonance to gravitational forcing. We do indeed discuss how this physical explanation fails in our paper.

BHDCN2013 ended up expressing more a litany of "personal opinions" and "personal doubts" (which are not even supported by convincing numbers and figures) misleadingly presenting them as "incontrovertible facts" on a number of issues than presenting accurate and convincing scientific demonstrations disproving the results or the theories proposed in the critiqued works.
This is a misrepresentation of our work.

BHDCN2013 also carefully avoided using their same critical logic to scrutinize the works that support their own advocated AGW theories that, as demonstrated in the peer reviewed literature BHDCN2013 criticize, contain far more serious shortcomings such as the macroscopic failure of the IPCC general circulation models in properly reproducing the observational temperature data at multiple scales such as the temperature standstill after 2000.
This point is repeated and is addressed above

BHDCN2013's attempt to dismiss scientific works with just a "philosophical" approach instead of using very accurate, explicit, detailed and extended physical and mathematical calculations and graphs is naïve, at least.
This is a misrepresentation of our work - we do in fact replicate and publish an accompanying R-package with necessary data and open source code for our method.

They should have scientifically disproved the papers being critiqued before attempting to write a

“philosophical” treatise. On the contrary, BHDCN2013 mislead a reader by giving an impression that the critiqued papers have been already so robustly rebutted in the literature that they can now propose a philosophical “agnotological” summary and interpretation of the case.
The philosophical aspect has been toned down in the revised paper.

Essentially BHDCN2013 did not realize that the quadratic fit from 1850 to 2000 simply captures, at a second order approximation, the secular trending of the warming from 1850 to 2000. This 1850 - 2000 warming trending is due not to anthropogenic forcing alone, but to whatever is causing it. On the contrary, the linear component after 2000 is supposed to capture and simulate only the anthropogenic contribution as deduced by the models but with a reduced climate sensitivity, as estimated by taking into account the existence of the oscillations such as the 20 and 60-year oscillation that would be responsible of about 60% of the warming observed from 1970 to 2000. So, the two components cannot be directly compared as BHDCN2013 misinterpreted.
We also have different views on trend representation and think Scafetta’s definition is unpersuasive and ad hoc. He provides no viable justification for his choice in his paper, and we are especially critical to using two arbitrary trend models for different parts of the series.

More recently, I have published a formal paper in the peer review literature where I demonstrate in detail some of the math and physical errors made in Benestad and Schmidt (2009). This is Scafetta N., 2013a. Discussion on common errors in analyzing sea level accelerations, solar trends and global warming. Pattern Recognition in Physics, 1, 37–57. DOI:10.5194/prp-1-37-2013
This is really a bit off-topic, where Scafetta makes reference to a recent paper of his in Pattern Recognition in Physics, which we will include in the revised version of our paper – this paper clearly demonstrates how he has unacceptably misrepresented the work Benestad and Schmidt (2009; BS09, available from http://pubs.giss.nasa.gov/abs/be02100q.html). A comment on this paper has now been published (Benestad, 2013), which shows the flaws in Scafetta’s reasoning. It also shows that he is making up this criticism and he has an extremely weak case.

If I understand well BHDCN2013 and their “free-phase” climate model, if the global surface temperature presents local “maxima” around 1880, 1940 and 2000 (the 60 - year cycle) while a climate model produces local “minima” around the same periods, e.g. minima in 1880, 1940 and 2000, then the model reconstructs the temperature patterns “well” because the phases do not matter!
Scafetta’s remark ’model reconstructs the temperature patterns “well” because the phases do not matter!’ suggests that he does not understand the situation. We know the model should not reproduce the phase of the internal variations in the non-linear and chaotic dynamical system; because we know that any astronomical forcing is absent.

His comment that "This is pure non-sense" reveals a fundamental lack of understanding of the climate models, the climate system, and the underlying physics.

BHDCN2013's argument is clearly logically flawed. In science only a definitive mathematical/physical demonstration can determine whether a published scientific claim is erroneous. It is evident that the mere resignation of a number of editors who simply disagreed with the results published in a paper does not demonstrate by itself that the incriminated papers (in some case published with the previous approval of the same resigned editor as in the case of Spencer and Braswell) are necessarily fundamentally erroneous: note that small imprecisions may always exist.
The editorial resignation is a matter of fact, proving there to be controversies. We do not say it's a "science demonstration". This is a "straw man" argument. We have noted that Humlum et al. (2011a) do not make a claim about the giant planets anywhere in their paper – see our response to their comment. We still believe that Humlum et al (2011) performed a curve-fit that is not suitable for attributing causes.

On the contrary BHDCN2013 misunderstood it for the Pearson's chi-squared test defined sometime as where in the denominator the theoretical value Ei (without square) is used simply because such a test is valid ONLY when the observational value Oi is Gaussian distributed around the theoretical expected value Ei with a standard deviation given by the root of Ei, which is an assumption that sometimes is made in statistics in specific cases. Thus, in the above equation, the denominator value Ei represents the "square" of the theoretical standard deviation. In the case discussed in my paper, the "square" of the standard deviation is not given by the absolute value of the model prediction itself but by its own measured variance, as correctly used in my equation.
The issue concerning the chi-squared distribution versus test is a minor point here, however, the sentence referring to this has been dropped in the revised version of the paper.

Conclusion
I believe that BHDCN2013 have written a very poor and weak work under any point of view: philosophical, mathematical and physical. Above I have demonstrated a number of different errors, shortcomings and misinterpretations. Other critiqued authors have highlighted other shortcomings. Therefore, I need to suggest the rejection of this work or that this work is published as it is together with the rebuttals of the criticized authors (free of charge).
I need to conclude that BHDCN2013 reminded me the pamphlet "Hundert Autoren gegen Einstein" (A Hundred authors against Einstein) published in 1931, which today, according to Goenner, is considered a mixture of mathematical–physical incompetence, hubris, and the feelings of the critics of being suppressed by modern physicists
(from http://en.wikipedia.org/wiki/Criticism_of_the_theory_of_relativity).
We have tried to argue that Scafetta's conclusion is wrong through the point-by-point

response above. We are happy that that the paper is published together with the comments. We do not see the logical link to the association with Einstein - it's purely irrational.

Some comments on the body of Scafetta comments, not included here follow: We have never stated that 'science on climate change is already perfectly understood and "settled"' - We think this is a misconstrued idea of Scafetta's. In our revised manuscript the fact that the science is not settled is spelled out. Our paper is concerned with falsification of some controversial articles, for which we have provided a source code and data, in addition to an explanation. With our analysis and Scafetta's comment, I hope we can make some progress. However, it would be helpful if Scafetta could disclose his own methods, so that others could examine his analysis in detail [1].

It is interesting to note Scafetta's view "given the fact that the climate is a dynamical system, not just a stochastic system". On this account, he deviates fundamentally with ours and e.g. those who think that long-term persistence may provide an explanation for the recent trends. Furthermore, he dismisses the effect from non-linear chaos, and the rich complexity of the earth system. The world is not as simple as that.

Picking one sample (here one model: GISS ModelE) and from it deduce that the whole sample (here ensemble of climate models - CMIP3) involves faulty logic, and hence is a logical flaw in Scafetta's argument. Furthermore, Scafetta refers to a "free-phase" climate model – which is a very strange way to put it. We argue that one cannot expect to see a signal from the planets in the climate models with the same phase as in the real world, because we know the models do not include this type of forcing. Furthermore, we think that the slow fluctuations arise from non-linear chaos, and the phase essentially is unpredictable. There is no reason to believe that the models should replicate the phase of these oscillations. What we show is that they more or less reproduce their magnitude.

Scafetta refers to blogs when referring to a previous paper by – ironically he accuses our paper for having a blog-like character. Again, the paper he cites in Pattern Recognition in Physics (why that journal?) that provides an unacceptable misrepresentation of the regression analysis in Benestad and Schmidt (2009). This can easily be checked by reading the two papers (both open access), and the recent comment on this paper explains the serious misrepresentation made by Scafetta. It is completely incomprehensible how he managed to misread the paper the way he did. In any case, the R-code for replicating the work is available in 'replicationDemos', and we would urge Scafetta to divulge his own code [1].

Also, see our response to the comments posted by Loehle and to Solheim et al. Scafetta's concern about editorial review do fit in our discussion about agnotology and peer reviewed papers. We too argue that the reviews sometimes are too weak, letting through papers such as those from Scafetta, where the gravest example probably is

Scafetta (2013).

References:

Benestad, R. E. and Schmidt, G. A., 2009. Solar trends and global warming, J. Geophys. Res. 114, D14101, 2009. (Open access from http://pubs.giss.nasa.gov/abs/be02100q.html)

Humlum, O., Solheim, J.-E., Stordahl, K., 2011. Identifying natural contributions to late Holocene climate change. Glob. Planet. Change 79, 145–156.

Scafetta N., 2013. Discussion on common errors in analyzing sea level accelerations, solar trends and global warming. Pattern Recognition in Physics, 1, 37–57.DOI: 10.5194/prp-1-37-2013.

Benestad, R. E., 2013: Comment on "Discussions on common errors in analyzing sea level accelerations, solar trends and global warming" by Scafetta (2013)., Pattern Recogn. Phys., 1, 91-92, doi:10.5194/prp-1-91-2013.
http://www.pattern-recogn-phys.net/1/91/2013/prp-1-91-2013.html

[1] Sceptical Climate Researcher Won t Divulge Key Program - Environment - 18 December 2009 - New Scientist.
Accessed July 5, 2013.
http://www.newscientist.com/article/dn18307-sceptical-climate-researcher-wont-divulge-key-program.html#.UdaMk86UAqU.