

Learning through replication in climate research

R.E. Benestad, H.O. Hygen¹, R. van Dorland, J. Cook, D. Nuccitelli, S. Lewandowsky, K. Hayhoe⁶

Supporting Material - Appendix

Table of Contents

Method - further details.....	4
Replication demonstrations.....	6
Questionable analytical set-up.....	6
Selective use of data.....	7
Curve-fitting.....	10
Biased analytical design.....	13
Ignoring negative tests.....	16
Questionable presumptions.....	20
Questionable representation in data.....	22
Looking at irrelevant aspects.....	23
False dichotomies.....	24
Circular reasoning.....	29
Statistical errors.....	33
Misinterpretation of statistical intervals.....	33
Failure to account for the actual degrees of freedom.....	36
Misconceived mathematics.....	38
Incorrect interpretation of mathematics.....	39
Implausible physics.....	41
Lack of plausible physics.....	41
Incomplete account of the physics.....	42
Irrelevant analogies.....	43
Misrepresentation of publications.....	44
References.....	46
Figures.....	56
Tables.....	60
Manual for replicationDemos.....	64

Method - further details

A mere focus on general points rather than detailed replication may not lead to enhanced understanding for why we don't know what we don't know (Proctor and Schiebinger, 2008). The emphasis must be on specific and detailed scientific/technical aspects in order to understand why different efforts lead to different results. We show how knowledge may progress through replication of 38 papers and how ignorance may be reduced for some controversies. In addition to the replication itself, the assessment of the papers should also involve an analysis of the logical reasoning. Wrong conclusions may result from incorrect logic for several reasons, here categorised from A-D:

- A. One may start from a correct logical premise and execute an erroneous analysis.
- B. One may apply a correct analysis but start from the wrong logical premise.
- C. One may start from the right premise, and correctly apply the analysis, but overstate the significance of the conclusions (the analysis does not actually address the question).
- D. One may start from wrong logical premise and apply erroneous analysis.

In some cases, the relevant category may be subject to different interpretation. For instance, a method may be applied correctly to the data in technical terms (e.g. wavelet), but the results may be misinterpreted because the limitations of the method are not accounted for, or an inappropriate method may be implemented in a way that would be correct in other contexts. The concept of over-fit too may belong to different categories, depending on the context.

The platform for this review was a software called 'R' (R Development Core Team, 2004; version 2.13.1). The R-environment is freely available from <http://cran.r-project.org>, and its user threshold should be low for people who

have some experience with computer analysis scripts such as S+, Matlab, or IDL. Furthermore, there is a number of good introductory manuals for R freely available on-line suitable for the level of this paper, however, some analytical expertise and computer literacy is needed to appreciate the source code. All the data sets contained in replicationDemos are provided with the attribute 'url' which identifies their data source on the Internet. The R-package contains examples and syntax descriptions in addition to the source code for the R-scripts and data. The numbers from various tables discussed here have been copied electronically from the PDF-version of the respective papers. The replicationDemos-package is available from the supporting material, where the zip-version is built for Windows systems and the tar-gz-version is made for Linux and Mac platforms. The package comes with a manual provided at the end of this SM. The installation of the R-package can be done through the menu in Windows (local zip file) or through a few command lines in R for Linux (here the R-prompt '>' is shown) - start R and then write these lines in the R window¹:

```
> install.packages("replicationDemos_1.11.tar.gc",repos=NULL)
> library(replicationDemos)
```

Some of the functions read data directly over the Internet, and hence require Internet access. Table S1 gives an overview of the cases which can be replicated through the replicationDemos package. Some of these may take some time if a large set of Monte-Carlo simulations is carried out. The replication can be implemented through typing the command line in R, after having loaded the R-package ('library(replicationDemos)'): e.g.

```
> Humlum.et.al.2011()
```

¹ For Linux and Mac users; for Windows, replace 'replicationDemos_1.11.tar.gz' with 'replicationDemos_1.11.zip'

The source-code is produced with the following line (the name of the functions without the parentheses `()`):

```
> Humlum.et.al.2011
```

The manual page for the R-function replicating the predictions by Humlum et al. (2011) are provided through:

```
> ?Humlum.et.al.2011
```

In the cases below, similar lines will be used to refer to the functions in replicationDemos used in our replication exercises.

Replication demonstrations

Here an extended list of examples is provided, both which are included in the replicationDemos package and cases which don't need replication for our understanding. All of these have been used to support arguments presented in the media and on the Internet. Many of these have been compiled in reports such as the NIPCC (Idso and Singer, 2009; Idso et al, 2013). Table S2-S3 give a various summaries of the contrarian papers examined here, where Table S2 lists the dominant cycles/frequencies found in the various papers. Table S3 lists all the 38 papers together with the type of shortcomings identified for each, and thus provides a basis for the synthesis provided in the main paper's results section.

Questionable analytical set-up

The way an experiment, a test, or an analysis is set up is important, and it is crucial that the outcome is not pre-determined by their design. They should

however, be objective in the sense that the information embedded in the data decide the outcome. There are various ways that lead to invalid results, including selective use of data, biased design, curve-fitting, and over-fit.

Selective use of data

Humlum et al. (2011a; HSS11a) suggested that natural cycles, e.g. the moon and solar variability, play a role a role in climate change on Earth, and that their influence is more important than changes in the greenhouse gases (GHGs). A replication of their analysis can provide a means for turning these contrarian claims into an educational exercise.

The core of the analysis carried out by HSS11a involved wavelet-based curve-fitting, with a vague idea that the moon and solar cycles somehow can affect the Earth's climate. The most severe problem with the paper, however, was that it had discarded a large fraction of data for the Holocene which did not fit their claims. Their reason for not showing the part of the data before 4000 BP was that they "*chose to focus on the most recent 4000 years of the GISP2 series, as the main thrust of [their] investigation is on climatic variations in the recent past and their potential for forecasting the near future*" (square brackets here denotes replacing 'our' with 'their'). Humlum had also been a co-author on an older article in a popular technical magazine where this absent part of the data had been presented (Bye *et al.*, 2011), and the data stretches almost 50,000 years back in time and is downloaded in one single file².

HSS11a examined the last 4000 years of the GISP2 (Greenland Ice Sheet Project Two) record, and constructed a mathematical model based on a sum of Fourier components and only three periods: 2804, 1186, and 556 years (for Svalbard annual mean temperature since 1912, they found Fourier components of 68.4, 25.7, and 16.8 years). Fourier series are often discussed in science textbooks and it is a mathematical fact that any finite series can be

²ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/greenland/summit/gisp2/isotopes/gisp2_temp_accum_alley2000.txt

represented in terms of a sum of sinusoids, which easily can result in mere 'curve-fitting' (Fourier expansion in this case). According to Stephenson (1973, p. 255), the sum of Fourier series is not necessarily equal to the function $f(x)$ from which is derived, since the function given by the Fourier expansion is mathematically bound to extend periodic regularity. Moreover, most functions $f(x)$, defined for a finite interval, are not periodic, although it is possible to find a Fourier series that represents this function in the given interval (Williams, 1960, p. 74). Pain (1983, p. 252) also clearly states that the Fourier series represent the function $f(x)$ only within the chosen interval, and one can fit a series of observations to arbitrary accuracy without having any predictability at all. This is a form of 'over-fit' (Wilks, 1995), and therefore it's important to verify model to data outside the fitted region.

The underlying data used in (HSS11a) analysis is replicated here through replicationDemos, but the function $f(x)$ for which they found a fit over a 4000-year period was extended to the data outside the training interval (Figure S1). 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 years 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.

Figure S1 shows that the curve-fit for the selected 4000 years does not provide a good description for the rest of the Holocene. The full red line shows their model results and the dashed red lines show two different attempts to extend their model to older data. One initial attempt was made, keeping their trend; however this obviously caused a divergence. So in the second attempt, we removed the trend to give their model a better chance of making a good hindcast. The model failed to represent the earlier data as claimed for the most recent 4000 years in their paper. Clearly, their hypothesis of 3 dominant

periodicities no longer works when extending the data period, and this is not surprising as this is explained in text books on Fourier methods.

In other words, the analysis made by HSS11a was limited to a subset of the data, failing to use the remaining part to evaluate their model. They argued that a comparison between predictions extended from 1850 until present day with trends in the global mean temperature constituted an out-of-sample test, and that their model had "a realistic forecasting time range of about 10–25% of the total record length". Hence, they ignored a large section of data which did not agree with their conclusions, and the fact that the model failed to be universally valid suggests that the chosen method was not objective. Mismatch outside the calibration period similar to that shown in Figure S1 is also a typical behaviour for models which are over-fit, and a limited forecasting horizon in combination with retaining only long time scales (greater than 556 years) is another factor pointing towards a curve-fitting exercise. Any curve-fitting is unsuitable for attributing physical causes. Furthermore, the physics assumed in HSS11a is implausible, as the results would imply that the sun and the moon had no effect on the Greenland temperature in the early part of the Holocene, but suddenly started to play a role for the last 4000 years.

It is also well-known that wavelet fits near the end-points are sensitive to the (subjective) choice of boundary conditions, and to use these results for attribution is therefore difficult. In other words, HSS11a failed to acknowledge well-known shortcomings associated with curve-fitting, but rather based their analysis on unjustified fit to a set of harmonics. In addition, their results lacked a well-formulated physical basis, and they failed to discuss past relevant literature concerning the physics as well as mathematics.

This paper has been used to suggest a false dichotomy between natural variations and trends forced by GHGs through the idea that the moon and the sun can account for most of the past variations in the temperature. It is difficult to sort HSS11a according to the categories of logical fallacies A-D, as one may argue that the paper belongs to either B or D. The initial premise is

incorrect, however whereas the method may be correct in some circumstances, it was inappropriate in terms of the present scientific question. One weakness was the lack of proper evaluation of the method, not using most of the data for Holocene for out-of-sample verification and contemplating how forced cycles are expected to be sustained in the long run. In this sense, the analysis ignored relevant contextual information. The paper also failed to recognise well-known facts about spectral methods and widely acknowledged explanations for the trend in the global mean temperature since 1850, which they purported to be a validating match for their thesis.

The same flawed approach was applied by HSS11a to the Svalbard temperature and the results in Humlum et al., (2011b). The HSS11a paper also provides a nice demonstration for why similar types of curve-fitting employed in Scafetta (2012a, 2012b, 2012c) and Loehle and Scafetta (2011; henceforth 'L&H11') fail to provide reliable answers.

The implementation of the replication of HSS11a is:

> Humlum.et.al.2011()

There is a paper by Yndestad (2006) that claims to identify a lunar 'nodal' cycle (18.6-year) in a selection of Arctic measurements and that the Arctic is a forced oscillating system controlled by the pull of gravity from the moon. This study too was based on a wavelet analysis, as those described above, and hence the same type of criticism applies. Furthermore, sea-ice and the local Arctic climate are strongly affected by winds and ocean currents, in addition to being closely coupled (Benestad et al., 2002). The Arctic climate involves dynamics with a pronounced non-linear chaotic character (Benestad et al., 2010), and the tides tend to propagate as coastal Kelvin waves rather than in the ocean interior. Hence, it would be problematic from a pure statistical analysis of a few measurements alone to attribute celestial causes to Arctic secular variations. The paper failed to provide an account on relevant topics

such as the dynamics of the Arctic climate and relevant processes and literature. Its fallacy can be described as category B: correct analysis but start from the wrong logical premise.

Curve-fitting

Loehle and Scafetta (2011; L&S11) purported that 20- and 60-year natural cycles in the global mean temperature estimates were due to natural cycles which they explained in terms of solar and astronomical influences. Furthermore, they claimed there was only a weak linear trend in the global mean temperature, and explained this in terms of a slight negative feedback in the climate system to CO₂.

It is easy to reproduce the analysis and demonstrate why the conclusions drawn by L&S11 are at variance with most of the climate research community. The problem with the L&S11 includes both a lack of clear physical basis and the analytical setup. L&S11 assumed some kind of selective and potent resonance to solar and astronomical forcing while a negative feedback was acting for CO₂. Resonance is inherent to the system, and it is difficult to conceive what it would entail that differed for the different types of forcings. A weak forcing and a pronounced response would imply a positive feedback, or at least an optimal balance between forcing periodicity and damping rate (a 60-year periodicity would suggest very weak damping, which seems unlikely, and hence the most convincing argument for resonance would involve a delayed positive feedback), and a preferred frequency would be an inherent characteristic of the earth climate system. Noisy forcings and transient functions, however, embed a range of frequencies and can therefore feed a resonance. Through our replication toolbox, we can demonstrate such cases, where a simulation of a forced damped oscillator picks up resonant variations if given a noisy forcing, even if the forcing itself has another dominant frequency. Furthermore, a resonant system will respond to a trend in GHG forcings (in mathematical terms, the forcing is proportional to $\ln|\text{CO}_2|$), and if such a resonance implies positive feedbacks, these could also be present in a

situation of GHG forcings. Hence it is extremely hard to attribute a cause for resonant response just from analysing cycles when several forcings are present.

The analytical problem involved a curve-fit, and 60-year cycles were estimated from a mere 160 years of data. In a complex and non-linear system such as Earth's climate, such an exercise is prone to be non-robust and non-stationary as seen in the case of HSS11a above. Furthermore, the analytical setup was not validated against independent data, and the skill of the model was not properly assessed. Furthermore, the methods used in L&S11 suffered from many of the similar flaws as those in HSS11a, even though L&S11 employed a different strategy for spectral analysis. Again, it is important to keep in mind that all curves (finite time series) can be represented as a sum of sinusoids describing cycles with different frequencies. Furthermore, Fourier transforms are closely related to spectral analysis, but these concepts are not exactly the same. Spectral analysis also tries to account for mathematical artifacts, such as 'spectral leakage', attribute probabilities that some frequencies are spurious, and estimate the significance of the results (Press et al., 1989). There is a number of different spectral analysis techniques, and some are more suitable for certain types of data. Sometimes, one can also use regression to find the best-fit combination of sinusoids for a time series, as in L&S11's "empirical decomposition". It's typical, however, that geophysical time series, such as the global mean temperature, are not characterised by one or two frequencies. In fact, if we try to fit other sinusoids to the same data as L&S11, we get many other frequencies which fit equally well, and we see that the frequencies of 20 and 60 years are not the most dominant ones. A trial with a range of periodicities for harmonic fitting in a similar regression analysis suggested that periodicities of 65.75 and 21.5 years gave a better fit in terms of value explained (R^2) than 60 and 20 years respectively.

Fitting sinusoids with long time scales compared to the time series is careless, which can be demonstrated through constructing a synthetic time series that is much longer than the one we just looked at. In our

demonstration it was important that the synthetic series was constructed from a combination of sinusoids for the entire period but with random amplitude and phase. We divided this synthetic time series into sequences with the same length as that L&S11 used to fit their model, and compared the fits for each segment. The 20 and 60 year amplitude estimates varied substantially from sequence to sequence when we adopted the same strategy as in L&S11, and the amplitude for the fits to the shorter sequences were ~ 4 times greater than a similar fit gave for the original 10,000 years long series. This is because there is a band of frequencies present in random, noisy and chaotic data, which brings us back to our initial point: Any number or curve can be split into a multitude of different components, most of which will not have any physical meaning.

The analysis presented in L&S11 can be described as a curve-fitting exercise based on two periods and that assumed that cycles with constant frequency in a non-linear and chaotic system, and similar to case 2, it involved an over-fit of a sets of sinusoids and lacked proper evaluation of the methods. The paper also failed to provide a persuasive account of the physics behind the purported links, and implies a false dichotomy, failing to falsify the thesis that increased GHGs influence Earth's surface temperature. L&S11 furthermore failed to account for relevant work and provide a proper context, and the paper can be classified under Category B: wrong premises.

The analysis used to assess L&S11 can be implemented through the following R-call:

```
> LoehleScafetta2011()
```

Biased analytical design

Scafetta (2012a) argued that celestial forcing in the form of gravitational forces from the giant gas planets explains most of the past climatic changes on Earth, and especially fluctuations of ~ 20 and ~ 60 years, proposing that solar

and heliospheric planetary oscillations would result in synchronous oscillations in Earth's climate, maintained by some resonance mechanism. He also evaluated how well global climate models reproduce the amplitude and phase of ~ 20 and ~ 60 year periodicity, which he attributed to the influence of gravity from celestial objects. Matching frequencies may suggest a connection, but does not rule out other explanations or confounding factors. A chaotic system may have a range of intrinsic frequencies, and there are several forcings with similar time scales (e.g. the Moon's orbit around the Earth and the solar Carrington rotation are both ~ 27 days). In addition, he carried out an evaluation of trends based on an arbitrary curve fitting, using different trend models for different parts of the data, which apparently gave a good match.

Scafetta (2012a) can be reviewed in terms of the physics and the statistical analysis. The account of the physics was vague, as Scafetta argued that resonant response could amplify the weak effect from the giant planets in the solar system, just like L&S11. He failed to acknowledge that resonance is an inherent property of a system, and will pick up any forcing with matching frequency. Synchronization would imply either rapid amplification or a resonance frequency matching that of the forcing. Scafetta (2012a) assumed similar resonance as L&S11, with the same weaknesses. Furthermore, some statistics presented in the paper were miscalculated. By repeating the work done by Scafetta, we can understand why his claims diverge from the mainstream climate science. Our replications demonstrated that the paper presented an inappropriate analytical set-up which favoured one outcome due to its design.

A statistical weakness in the analysis presented in Scafetta (2012a) is the handling of trends, as a quadratic trend that conveniently fitted the data was used for the period 1850-2000, and then a linear fit with a warming rate of $0.009^{\circ}\text{C}/\text{year}$ was used after 2000. The quadratic equation for 1850-2000 $p(t) = 4.9 \cdot 10^{-5} x^2 - 3.5 \cdot 10^{-3} x - 0.30$ (eq. 4, where $x = t - 1850$) gave a warming rate $dp/dt = 2 \cdot 4.9 \cdot 10^{-5} x - 3.5 \cdot 10^{-3} = 0.011^{\circ}\text{C}/\text{year}$ for year 2000. Hence, the method used by Scafetta implicitly assumed that the rate of

warming was abruptly reduced from $0.011^{\circ}\text{C}/\text{year}$ in year 2000 to $0.009^{\circ}\text{C}/\text{year}$ for the future. It also implied that the future warming rate was smaller than the range reported in Solomon et al. (2007), and much of the recent warming was mis-attributed to natural variations based on curve-fitting similar to that of HSS11a.

There seems to be a number of other results in Scafetta (2012a) which are difficult to reproduce, as a replication of figure 5b in the paper suggests that it displays a lower projected trend than produced by the equations cited in the paper (Figure S2). Our replication suggests that there is an inconsistency between the figure in his paper and the information provided in the main text. Scafetta (2012a) also limited the confidence interval to one standard deviation (which implies a 68.6% confidence interval) in the evaluation to see whether the model results overlapped the observations (the more commonly used 95th confidence is roughly spanned by 2 times the error estimate).

The gravest issue with the Scafetta (2012a) analysis involved a series of tests which in effect were 'rigged' to give negatives which is a logical flaw on which his conclusions hinge. They involved a regression analysis to estimate amplitude and phase of 20 and 60 year oscillations in the global mean temperatures, assuming that these were due to the gravitational influence from celestial bodies. The phase and amplitudes found for the observations then were used as a yard stick for the GCM results, and a regression analysis was used where the covariates were the same as for the observations, with exactly the same phase and amplitude specified for the 20- and 60-year oscillations. We know *a priori* that the planets are not accounted for in the CMIP3 climate simulations (Meehl et al., 2007), and hence Scafetta's strategy is not suitable to provide an objective answer. A more appropriate null hypothesis would be that the amplitudes seen for the 20 and 60 year variations would be due to noise, e.g. internal variability. Hence, it is important to allow the phase to be unconstrained in the analysis, as we have done (Figure S3). When we repeat the analysis using a suitable set-up, we don't see a falsification of the null-hypothesis, especially if we account for the fact that

the analysis involves multiple tests and take the field significance into account (Wilks, 2006).

Scafetta (2012a) assumed that his method was validated if it was calibrated on one cycle of 60 years and then was able to reproduce the next 60-year cycle in the data that was not part of the calibration. However, this argument is not justified, as this type of approach fails for the El Niño Southern Oscillation (ENSO), for which it is well known that there were two El Niños during the 1980s, which taken together, resemble two periods of a periodic cycle (Figure S4). While each event provides a good fit for the other cycle, if calibration is performed on half of the decade and evaluated against the other, the predictions by this regression model fails to capture the variations outside this interval. It is therefore important to capture many cycles in a time series before one can establish a periodic signal, as only two cycles will likely not be representative of the entire system.

The implications from Scafetta (2012a) is a false dichotomy between the effect of changing atmospheric greenhouse gas concentrations and the influence from astronomical factors, as the paper did not provide a falsification of the former. The paper provided references of the presence of ~ 20 and ~ 60 -year cycles in other observational record, however, here we are not claiming that there are no such variations. Rather, we claim that variations with comparable amplitudes are found in the GCMs, and hence it is not necessary to include astronomical forcing to account for these. We also show that the method employed by Scafetta (2012a) incorporated logical mistakes, making the conclusions invalid. For instance, the method involved over-fitting a set of sinusoids, which by construction, is bound to match the training data, but may radically fail to predict out of sample observations. Scafetta (2012a) furthermore implied 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. His analysis claimed to test climate models over the instrumental record, ignoring the fact that no GCM is ever expected to match the natural fluctuations in the observed temperature, and in doing so

setting up a strawman. 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. These points are well-known knowledge within the climate research community, however, Scafetta did not acknowledge this knowledge in addition to the fact that he ignored literature on GCMs and relevant climate research. Furthermore, the paper presented no proper evaluation of the proposed model for out-of-sample tests. For these reasons, Scafetta (2012a) belongs to Category D: starting from wrong logical premise (the phase information is relevant), applying erroneous analysis (confidence intervals, missing evaluation, and trend models), due to an inappropriate strategy for which the one answer was favoured, in addition to using wrong statistics. The same type of shortcomings were also present in the curve-fitting (over-fit) presented in Scafetta (2012b), in addition to a clear physical basis was lacking.

The implementation in R:

```
> Scafetta2012()
```

Ignoring negative tests

Solheim et al., (2011; SSH11) argued that 60% of the annual and winter temperature variations at Svalbard are related to the solar cycle length (SCL). The basis for their conclusions was a high correlation estimated between SCL and the temperature estimates, and results from a Durbin-Watson test. The highest correlation reported were -0.82 for the winter mean of the decade lagging one solar cycle. Repeating their analysis with our open-source methods gave different answers. Their connection also 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. Furthermore, the analysis was not objective, inflating the importance 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.

When we reconstructed their table 1 we got nearly the same results, albeit not identical. SSH11 based their analysis of SCL on data from National Geophysical Data Center, which uses a method based on a publication from 1939 (Waldmeier, 1961), however, more recent work on the estimation of SCL account 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, SSH11's estimate of the SCL for cycle 23 (12.2 years) was substantially longer than the estimate of 10.5 years reported by the Danish Meteorological Institute (based on Friis-Christensen and Lassen (1991) and follow-up studies) and 10.8 years estimated by Benestad (2005). Such a long cycle is the basis for their projected cooling (a decrease from -11.2°C to -17.2°C with a 95% confidence interval of -20.5°C to -14°C) at Svalbard over solar cycle 24 (starting 2009). The observed mean over 2009-2012 suggests a continued warming that reached -8.88°C as an average for the 4 years, which means that the mean winter temperature of 2013-2019 (the next 7 winter seasons) must be -21.95°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. This analysis can be replicated in replicationDemos through the following line:

```
> data(ssh2011.tab1)
> DJF(ssh2011.tab1=ssh2011.tab1)
```


SSH11 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-year segments. 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. When we computed the correlation (using 'R's 'cor.test') between SCL and the winter temperature listed in their table 1, we obtained a correlation for the winter of 0.37 with 95% confidence interval between -0.39 and 0.83, and when using the previous cycle SCL, we got -0.84 with a confidence interval between -0.39 and -0.96 (as opposed to -0.52 and -0.97 reported by SSH11). The 'cor.test' test statistic is based on Pearson's product moment correlation coefficient 'cor(x, y)' and follows a t-distribution with 'length(x)-2' degrees of freedom, and an asymptotic confidence interval is given based on Fisher's Z transform.

The estimate of the errors in the correlation in SSH11 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 (η_S) and mean temperature estimates (η_T). Here the symbol (subscripts) 'S' refers to SCL and 'T' refers to the local winter mean temperature. We estimated the error in SCL from the standard deviation of the difference between the SCL estimates from SSH11 and Benestad (2005): $\eta_S = \sigma_S$, where σ_S is the standard deviation of the SCL difference: $SCL_{SSH11} - SCL_{B05}$. Then we re-calculated the 95% confidence interval of the correlation estimates by adding white noise to temperature and SCL with standard deviations of σ_S for SCL, and for temperature we took the error of the mean estimate to be $\sigma_T/n^{1/2}$ where n is 10 for each 10-year long segment. The Monte-Carlo simulation of the correlations between temperature and SCL were then estimated as: $cor(T + \eta_T, S + \eta_S)$, and was repeated 30,000 times with different random realisations of the error terms η_T and η_S . The Monte-Carlo

simulations gave a 95% confidence interval for the correlation between -0.85 to 0.08, substantially wider than both 'cor.test' and SSH11. However, the latter two did not account for the uncertainties in the SCL estimates, which amplify the real uncertainties. Due to substantial uncertainties in the SCL, the Monte-Carlo simulations that we propose represent the most appropriate approach assessing the confidence intervals.

The Monte-Carlo simulation also revealed that the SSH11 correlation estimate was not centred in the simulated correlation error distribution, but was biased towards higher absolute values. The correlation estimate based on the Benestad (2005) SCL, on the other hand, gave a better match with the mean correlation from the Monte-Carlo simulation, although this too had a greater absolute value than the mean error estimate. Furthermore, the bootstrapping approach adopted by SSH11 seemed to give a biased error distribution, and we did not get the same 95% confident limits as they did (we made 30,000 iterations). 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 SSH11 of SCL was indeed 'fortunate' within the bounds of error estimates by getting correlations in the high end of the spectrum.

Since SSH11 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: $p_w = 1 - (1 - \alpha_{global})^{1/K}$ (Wilks, 2006). The reason is that from 100 random tests, about 5% are expected to achieve scores that are at the 5% significance level. Another question is how many other temperature series that have been examined, as the appropriate number of tests to use in the Walker test should include all (also any unreported) tests in order to avoid a biased selection or lucky draw. When we estimate the p-value of their correlation from the null-hypothesis derived from the Monte-Carlo simulations, we find that all the p-values exceed p_w , and hence their results are not statistically significant at the 5%-level.

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 (“The publication bias”; Wheelan, 2013), and apply e.g. the Walker test. The failure to do so will give misleading results.

The main problem with the analysis presented by SSH11 was the lack of a convincing physical basis, inappropriate hypothesis testing, the inflation of importance, and a small data sample insufficient to support the conclusions. Furthermore, there was missing contextual information in the paper concerning the dynamical character of the Arctic climate, the effect of an Arctic amplification on trends, and established inter-dependencies. The paper can be filed under Category A in terms of logic: starting from a correct logical premise and execute an erroneous analysis, both in terms of the boot strapping and the failure to account for multiple tests.

The replication of SSH11 is implemented with the following command lines in R:

```
> Solheim.et.al.2011()
```

Questionable presumptions

Scafetta and West (2007, 2006a, 2006b, 2005) argued that the recent increases in the global mean temperature were influenced by solar activity rather than increased GHG concentrations. The analysis, on which Scafetta and West based their conclusions, assumed that the global mean temperature was not influenced by factors other than solar variability on decadal to multi-

decadal time scales. Furthermore, they dismissed the role of increased concentrations of GHGs, based on the model fit to the solar trend, assuming that a solar influence excludes the effect from increased CO₂-levels.

Scafetta and West assumed that all the climate variability over wide frequency bands spanning 11 and 22 years were due to changes in the Sun. They developed a model which was not evaluated against independent data, and hence they had no information about its skill. Benestad and Schmidt (2009) demonstrated that the strategies employed in Scafetta and West (2005, 2006a, 2006b, 2007) were unsuitable for analysing solar-terrestrial relationships, and the source code for replicating these studies is included in replicationDemos. Scafetta and West's strategy failed to account for 'spectral leakage', common trends, and the presence of a range of frequencies in chaotic signals. They applied a transfer function based on the ratio of the standard deviation for respective temperature and total solar irradiance (TSI) after having applied a broad band-pass filter (7.3-14.7 and 14.7-29.3 years) to both. Moreover, their analysis a priori assumed that no other factor was affecting Earth's climate over these wide ranges of time-scales, and hence it is not surprising that they arrive at a misguided answer that seemed to suggest a strong solar influence. Furthermore, their conclusions hinged on a set of assumptions which were not justified and the methods were not subject to proper evaluation. These papers can be associated with Category B in terms of logical failure: starting with wrong premises, where the main logical flaw is that the relation between the sun and climate is given by the design of the model. These papers also implied a false dichotomy between the effects from solar variations and GHGs, without addressing the latter issue. Furthermore, the papers neglected relevant context such as literature on GCMs and critical views on solar-terrestrial links.

The replication of the Scafetta and West papers, as done by Benestad and Schmidt (2009), is implemented with the following command lines in R:

```
> Scafetta2006()
```

Questionable representation in data

Beck (2008) described a curve for atmospheric CO₂-concentrations which is at variance with corresponding results presented in Solomon et al. (2007). He compiled measurements from different locations at different times, often in Europe near CO₂-sources. The implication is that the upward trend in the current CO₂-measurements (Keeling curve) is not extraordinary.

Modern satellite-based measurements (NASA/AIRS) show that the concentrations in these regions may be substantially higher than the background level because of their proximity to the emission sources. Beck presented dramatic changes in CO₂-concentrations, which cannot be explained in terms of the carbon cycle (exchange between air, sea, and surface, involving photosynthesis and ocean acidification). Hence, the ignorance and neglect of relevant context makes such analyses prone to misguided interpretations, as in Humlum et al. (2013).

The analysis carried out by Beck (2008) did not reflect the global background levels, but the results were affected by the contamination from local sources and suffered from a lack of homogeneity. His results were not corroborated by independent studies of related aspects, such as the carbon cycle and carbon budgets. Since Beck (2008) was based on contaminated measurements that did not represent the global CO₂ background levels, it may be classified as B (start from the wrong logical premise) or C (not addressing the actual question). The analysis neglected relevant information and the contextual information such as the carbon budget, sinks and sources. Furthermore, the paper provided insufficient information about the data.

Looking at irrelevant aspects

Humlum et al. (2013; HSS13) argued that changes in CO₂ follow changes in the temperature, and that this implies that the increases seen in the Keeling curve are not man-made. Their claims implicitly support the CO₂-curve

presented by Beck (2008), and the thesis that the increase in the CO₂ concentrations seen in the Keeling curve is not due to the burning of fossil fuels, has long been an aspect of agnotology surrounding the global warming issue. The analysis on which HSS13 based their conclusions filtered out 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, HSS13 found the well-known link between El Niño Southern Oscillation and CO₂. They then incorrectly assumed that this link excludes the effect of anthropogenic emissions.

HSS13 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. They also applied a differencing operator (DIFF12) to the data followed by a lagged correlation, and in effect removed all trends and long time scales. A comparison between the shorter global and longer Mauna Loa series had some effect on the lagged correlation, however, the main problem was the use of DIFF12 followed by the correlation, as this strategy is designed to neglect trends. It is easy to demonstrate that the method Humlum et al. used is unable to pick up the longer time scales, as shown in replicationDemos. In other words, the analysis emphasised the short time scales, and the analytical set-up was pre-disposed to ignore the anthropogenic component to the CO₂ concentrations. Hence, the analysis contained a logical flaw since conclusions based on short-term fluctuations were drawn for the long-term time scales.

Another problem was that their study did not account for the carbon-budget, such as sources and sinks. It is not clear whether the increased CO₂ was assumed to originate from the ocean surface or the deep ocean, and their discussion ignored the literature concerning diffusion of trace gases in the oceans. They also neglected the work documented in the fourth assessment report of the IPCC (Solomon et al., 2007) regarding changes in the O₂/N₂ ratios, the acidification of the world oceans, and isotope ratios (Kern and

Leuenberger, 2013). Further criticism of HSS13 have been published in comments to the article (Masters and Benestad, 2013; Richardson, 2013).

The way HSS13 fails logically suggests it can be attributed to category C: addressing a different question. Another point was missing relevant contextual information, such as facts about the carbon cycle and ocean dynamics. The replication of the HSS13 is implemented with the following command lines in R:

```
> Humlum.et.al.2013()  
> diff12demo()
```

False dichotomies

The papers (Friis-Christensen and Lassen (1991; FL1991), Lassen and Friis-Christensen (1995; LF2000), Svensmark (1998; S1998), and Svensmark and Friis-Christensen (1997; SF1997) claimed that changes in the sun can explain a large part of the recent global warming. These papers have been used by Scafetta (see the earlier examples) and others as a support for their purports. Furthermore, they have contributed to the notion that galactic cosmic rays (GCR) play an important role for Earth's climate, that has been popularised through the media. They have also implied that GHGs, such as CO₂, play a relatively small role for Earth's climate, and dispute the view presented by the mainstream climate research community (National Research Council, 2001; Oreskes, 2004; Solomon et al., 2007). The conclusions from these papers rest on a curve-fitting exercise and are based on little physics. The data handling has also been questioned (Laut, 2003), and recent up-to-date replication has suggested that the predictions diverge from the observations. Stauning (2011) took advantage of two additional solar cycles to recalculate the relationship between sunspot and temperature data. The trends in temperature and solar cycle length showed a strong divergence after 1976. These analyses are similar to the classical studies on the relationship between

sunspots and climate performed over the centuries and that eventually have failed to stand up to new data (Benestad, 2002). Another point is that there is no trend in the solar proxies over the last 50 years (Benestad, 2005, 2013b; Lockwood and Frölich, 2008).

Damon and Laut (2004; DL2004) pointed out several flaws in the FL1991, LF2000, S1998, and SF1997, and argued that the apparent good match in FL1991 were obtained by “adding to a heavily smoothed (“filtered”) curve, four additional points covering the period of global warming, which were only partially filtered or not filtered at all”. Mixing data subject to different pre-processing and filtering is deemed to result in unreliable answers and is prone to introduce spurious artifacts. Another question is whether filtering solar cycle lengths could be justified, as each epoch lasted approximately 11 years, and hence implied that very slow changes in the sun would correlate directly with short term variations on Earth in a warped fashion. It is hard to conceive how the mean temperature in the period from 5 years ago to the next 5 years will be influenced by the solar activity from 25 years in the past to 25 years in the future. DL2004 also found trivial arithmetic errors in LF2000, being responsible for an incorrect curve. It is also difficult to explain why SCL should affect the climate, although the notion is based on the idea that short SCL is associated with more intense solar activity, however, it is a conundrum why there is a correlation between SCL that is stronger than corresponding correlations with the number of sunspots or the total solar irradiance.

For the analysis by S1998, DL2004 argued that the use of data from the U.S. Defence Meteorological Satellite Program in SL1997 and S1998 was inappropriate as they did not represent total global cloud cover. More appropriate data from the International Satellite Cloud Climatology Program (ISCCP) were inconsistent with their hypothesis (Laken et al., 2012), and DL2004 observed that the more recent and conflicting part of the ISCCP data were shown in the SF1997 article but were omitted from the S1998. Independent investigation of the solar cycle lengths is in line with DL2004 (Benestad, 2005), and open-source replication method is available:

> Benestad2005()

These papers have also had an influence on Svensmark (2007), Shaviv (2002), Courtillot et al. (2007) and Veizer (2005). The papers S1998, Svensmark (2007), Shaviv (2002), Veizer (2005), and Courtillot et al. (2007) argued that GCR affect Earth's global cloud cover which subsequently modulates the planetary albedo. The latter group of papers also assumed that a strong connection between GCR and climate implies a weak role for GHG such as CO₂. The GCR are known to be modulated by solar activity through its influence on the inter-planetary magnetic field (IMF), and are responsible for the creation of cosmogenic isotopes such as C-14 and Be-10 (Benestad, 2002). The analysis presented in these contrarian papers was based on data from various sources, ignoring part of the data, and lacked proper verification. The GCR have been introduced to the general society through popular science books (e.g. Svensmark, 2007) and videos (e.g. 'The Cloud Mystery', 'The Great Climate Swindle'), and have represented an important feature of agnotology in northern Europe. The influence of GHG has often been dismissed on the grounds of the speculated correlation between GCR and climate, assuming that the GCR-connection excludes the effect of changes in the GHG concentrations.

Veizer (2005) failed to present any resemblance between the GCR-proxies discussed in the paper and a proxy for temperature, and he provided no quantitative statistical analysis on the correspondence between these quantities. Furthermore, the purported dependency involved a neglect of the fact that many other factors may be more important in terms of generating cloud condensation nuclei. In replicationDemos, estimates based on Be-10 and temperature from the Vostoc ice cores can be shown together, and any correlation between the two seems to be due to long-term trend over 40000 years rather than more 'ephemeral' fluctuations on thousand year time scales. The correlation between Be-10 and the temperature proxy over the last 40000 years was -0.78 but this number reflected the long time scales (greater than

5000 years). The high frequency component was estimated by subtracting a low-pass filtered record, using a Gaussian window with a width of 5000 years. The correlation between the high-frequency components were only -0.23 with a 95% confidence interval of -0.65 to +0.30.

In other words, it is difficult to discern any credible evidence linking GCR and recent climate change, due to weak correlation and the number of other factors present. Veizer (2005) did not exclude other possibilities, but assumed that the other factors would be weak if there were a strong connection between GCR and climate. The paper can be described as category C: not addressing the right question. Furthermore, the paper suffers from missing contextual information, e.g. literature critical to the GCR-link, and there are major uncertainties concerning relevant physics.

The replication of the Veizer (2005) is implemented with the following command lines in R:

```
> paleaproxy()
```

Svensmark (2007) did not answer the serious criticism forwarded by Damon & Laut (2004) and Laut (2003). The original analysis presented by Svensmark was based on total cloud cover, which later turned out to provide a poor fit, and he then replaced these with data describing low-level cloudiness. He then used a different version of the cloud data to others, claiming that the original data were incorrect due to calibration problems and that the recent global warming was caused by GCR (Laken, et al., 2012).

On a similar note, Courtillot et al. (2007) presented an analysis between solar irradiance and geomagnetic field, but ignored part of the data record for which the data diverged (Bard and Delaygue, 2008). Other errors included a confusion between the interpretation of solar irradiance changes and net forcing. They also argued that periods with high GCR-flux, found in cosmogenic isotope records, coincided with periods with high ice raft debris in the North

Atlantic and assumed that high iceberg drift activity implies cold global conditions, which has not yet been established; Icebergs tend to originate from calving of ice sheets and glaciers, and should not be confused with sea-ice. As opposed to sea-ice, calving activity may not necessarily increase with lower temperatures.

Shaviv (2002) considered extreme time-scales of millions of years. He argued that our solar system takes about 250 million years to circle the Milky Way galaxy and that our solar system crosses one of the spiral arms about every ~ 150 million years. This number was arrived at by measuring the rotational velocity of stars in the Milky Way disk or other spiral galaxies. The pattern speed of the spiral arm in the Milky Way has not been firmly established, and a number of values are listed in table 3 of Shaviv (2002) for the pattern speed of the spiral arms, taken from other publications ranging from 1969 to 2001. However, he disregarded most of these results and derived "period for spiral arm crossing" of $p=134 \pm 25$ Myr for four spiral arms in the upper extreme of the published range. Nevertheless, such astronomical considerations are a far shot from present state-of-the-art measurements and understanding of cloud physics here on Earth. The distant past of the solar system and our galaxy is known to a far lesser extent than modern climate science.

These papers also neglected cloud condensation nuclei (CCN) from other sources, and their implications concerning GCR and the recent global warming cannot explain why the warming has been greatest during night (Solomon et al., 2007): The albedo mechanisms would be more important for the day side of the earth. It has also been established that there is no significant trend in GCR and other solar activity proxies in the last ~ 50 years (e.g. Richardson et al., 2002, Benestad, 2002, 2005, 2013b) and that in the most recent decades, there even has been a small trend in opposite direction to what is expected for solar forcing to cause a warming (Lockwood and Frölich, 2008). A review of the studies of GCR-climate links have not been supported through subsequent investigation (Laken et al., 2012).

The way that these papers have handled and selected the data have been questioned, but due to lack of open-source methods and sharing of data, it has difficult to pin-point the exact reason for the differences. These papers failed to account for relevant information, such as additional data, known processes, and literature on atmospheric physics. These papers can be described as category A: start from a correct logical premise and execute an erroneous analysis (data processing). They have been used to imply a false dichotomy, however, they have not excluded the possibility that increased GHGs may be the reason for trends in the global temperature. Moreover, they tend to ignore contextual information such as known physics, and past attempts to draw a connection between sunspots and climate (Benestad, 2002).

Circular reasoning

The question about long-term-persistence (LTP) is especially interesting, although the issue here was not whether LTP is present or not. It is associated with genuine scientific questions that appear to be unresolved: On the one hand, LTP has been used to model hydrological processes such as the Nile river level and may be the most successful strategy to model processes with unknown dynamics., while the climate system, on the other hand, involves a set of known processes in terms of a set of equations describing well-known physical laws. Our general concern here is that LTP often is assessed in isolation where other relevant geophysical data or general scientific understanding of the underlying physics are not considered. Hence, not making the maximum use of information. Furthermore, we note the risk of over-fitting models accounting for LTP. For the two papers addressed here in particular, we criticise the way the null-models (which account for LTP) was applied, and argue that they were inappropriate and 'contaminated' by the signal.

Cohn and Lins (2005) observed that tests for trends are sensitive to the expectations (the choice of the null-hypothesis), and argued that LTP makes standard hypothesis testing difficult. The implications of their conclusions was

that the observed recent global warming is not extraordinary, but a normal consequence due to LTP, however, their analysis of LTP did not take into account other information, such as physical mechanisms known to be involved. It is widely acknowledged that the atmosphere and oceans are driven by non-linear dynamics, giving rise to chaos (Lorenz, 1963). In such chaotic systems, the memory of the initial conditions is lost after some time, and the dynamics can be simulated based on the state of the system and the laws of physics alone: $dX/dt = f(X)$ (Palmer, 1996). In LTP processes, on the other hand, x_i depends on all previous points ($x_i = f(x_1, x_2, \dots, x_j), j < i$; Armin Bunde³), which implies a long-term memory that is absent in chaotic data. Lyapunov exponents can be used to assess the presence of chaos (Pesin, 1977), whereas the Hurst coefficient can provide an indicator of the LTP character (Kantelhardt et al., 2001). Non-linear chaos may, however, resemble LTP through the presence of multiple regimes in a system's strange attractor, and similar models used to describe LTP may also provide a description of chaotic systems, based on the understanding that both have a stochastic character. A fundamental difference between the two concepts involves the source of information: whereas LTP assumes uncertainty but derives a systematic pattern based on the past record, the simulation of chaotic systems introduces information based on an understanding of the physics and dynamical processes.

Processes involving a forced trend also exhibit some LTP if the total forcing has a LTP character, and the test by Cohn and Lins involved some degree of circular logic: Forcings with LTP characteristics increase LTP, and so an LTP derived from data that have been influenced by *external* forcings is not representative for the *intrinsic* LTP of the system. It can easily be shown that the total forcing data used for input to climate model simulations have this kind of behaviour (Hurst coefficient for the total forcing ~ 0.98).

> demo(ltpforc)

³ <http://www.climatedialogue.org/long-term-persistence-and-trend-significance/>

For the real climate system, the total forcing is a combination of the influence from GHGs, other anthropogenic effects (e.g. landscape changes), and natural variations (e.g. solar, volcanic). Thus, in order to be physically consistent, arguing for the presence of LTP also implies an acknowledgement of past radiative forcing in the favour for an enhanced greenhouse effect. This point can be demonstrated by applying the function 'testLTP' in replicationDemos to compare the auto-correlation function (ACF) estimated from a set of results produced by the global climate model ECHAM5 (Demuzere et al., 2009; Keenlyside and Latif, 2002) with constant boundary conditions and with historic forcing respectively. An objection to comparing ACFs is that adding a trend to a stationary, noisy, non-LTP signal will produce a LTP-like tail, which is a major reason why the ACF should not be used to demonstrate LTP in short time records (Rypdal, 2013), but this is also the reason why the calibration of LTP models should not be done with data subject to past external forcing. Nevertheless, the presence or absence of LTP is not the issue here, which instead is to show that methods such as auto-regressive moving-average (ARMA), auto-regressive integrated moving-average (ARIMA), and auto-regressive fractionally integrated moving-average (FARIMA) are designed to mimic the data, trends, and autocorrelations. These aspects are influenced by the assumptions made in the hypothesis testing, and if the same types of models were trained on the results from a long simulation with a climate model with no forcing, then they would be expected to mimic the internal variations. It would be correct to then apply such models to generate a null-distribution for the simulations that account for a forced response, however, time series models trained on externally forced climate model simulations would not provide a true description of the internal variability null-distribution. Hence, while it is true that statistical tests do depend on the underlying assumptions, it is not given that statistical models such as ARMA, ARIMA, FARIMA provide an adequate representation of the null-hypothesis when trained on past observational records which have been subject to forcing. It is important to avoid interpreting part of the signal as 'noise', as all these statistical models do

represent a type of structure in time, be it as simple as a serial correlation, persistence, or more complex recurring patterns. Thus, the choice of model determines what kind of temporal pattern one expects to be present in the process analysed. The flexibility of adapting these stochastic time series to different time series has a downside, since the flexibility may lead to over-fitting the models, e.g. so that they mimic the long-term trends. The statistical LTP models employed by Cohn and Lins were just convenient models which to some degree mimic the empirical data (tuned for several parameters), and are arguably far inferior compared to the physics-based general circulation models (GCMs) for providing appropriate null-distributions (long control simulations). No GCM reproduces the observed global warming unless an enhanced greenhouse effect is taken into account (Solomon et al., 2007; Marcott et al., 2013), and there is a well-known physical reasoning for why it has to be so (Weart, 2004).

Another difficulty with the notion that the global mean temperature varies randomly with substantial long-term departures from its mean, is that it then would imply a more unstable system with similar warming as we now observe throughout our history. However, the indications are that the historical climate has been fairly stable during the Holocene (Solomon et al., 2007). Cohn and Lins ignored all physical considerations in their analysis, and a serious problem with the idea that departures (such as the recent global warming) is random and natural, is that such changes in the global surface temperature would have physical implications in terms of energy conservation and the climate sensitivity. Similarly, phase scrambling (Franzke, 2012) are sensitive to embedded long-term trends which may not be part of the noise. Hence, such models are not suitable for testing trend hypotheses when it is not known a priori what fraction is part of long-memory noise and what is really the signal.

In summary, the difficulty with the analysis presented by Cohn and Lins was distinguishing between noise and signal, and treating both as noise resulted in misguided conclusions. We have not taken a position on the

question of the presence of LTP, or whether LTP is compatible with non-linear chaos. Another problem with the idea that the climate is highly sensitive to variations in its own state where small differences in the state lead to pronounced spontaneous variations, is that this implies a high climate sensitivity and strong positive feedback mechanisms. We know that there is a forcing present associated with increases in CO₂ concentrations, and Cohn and Lins could not show that the climate sensitivity discriminate against some types of forcings and not others. Furthermore, the papers ignored contextual information, such as the literature on atmospheric and climate modelling, chaos, and other relevant geophysical data. Hence, it is 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. A more appropriate indicator for climate change would be the oceanic heat content or the sea level, which are less characterised by multi-annual variations. The case examined here can be placed under category B: correct analysis but start from the wrong logical premise. The paper implies a false dichotomy between natural and anthropogenic climate changes. We also note that the idea that most of the variations are LTP-noise is incompatible with a number of the other papers here, e.g. arguing that the cycles in geophysical data records are due to astronomical forcings, however, we do not imply that those studies are right - at least one of these, most likely both, are wrong. The methods employed are prone to overfit, and were not evaluated for such caveats.

The replication of the Cohn and Lins (2005) is implemented with the following command lines in R:

```
> testLTP()
```

Statistical errors

Misinterpretation of statistical intervals

Douglass et al. (2007; henceforth DCPS07) claimed that upper air trends in the Tropics predicted by global climate models were inconsistent with the trends measured by radiosondes and satellites. This purported discrepancy has been echoed on various Internet sites, been promoted by the Norwegian organisation "*klimarealistene*", and included in the NIPCC report. The flaw in the analysis presented in this paper can easily be exposed through replication, and in order to appreciate the technical aspects of this analysis, it is necessary to look to elementary statistics. The core of the analysis presented in the paper involved a comparison between model results and observed data, based on a statistical hypothesis testing, which in this case comprised of a standard one-sample test, where a sample (trend in observations) was compared to a population representing the hypothesised behaviour.

Here, *population* means the statistical distribution of the sample of trends taken from an ensemble (a set) of model simulations, and is a standard statistics concept. The appropriate null-hypothesis would be that the sample (observed trends) belonged to the population describing the model results. In the following discussion, we will use standard notations and definitions from statistics, where the *range* describes the difference between the largest and smallest values of a data sample. The *interval*, on the other hand, is taken to be a set of two numbers describing the lower and upper limits between which the true value of a parameter is located with a given probability. In statistics, the evaluation of a parameter, such as the mean value of a population (the first moment), is imprecise because the data sample is a random set of numbers drawn from a larger universe, and may vary from one sample to the next. The confidence interval accounts for the effect of random sampling on the parameter estimate, providing lower and upper limits within which the parameter is most likely to be located. An interval can also refer to a population, describing the lower and upper limits within which a random sample is likely to be located with a given probability. Hence, an interval may

refer to different things depending on the context, which can result in confusion. In general, the interval describing the effect of random sampling on the estimation of parameters, such as the mean value, tends to diminish with the sample size. The population interval, on the other hand, is expected to be invariant with sample size, being defined by the population's probability distribution function (pdf).

DCPS07's conclusions were based on an inappropriate analytical set-up and relied on an analysis which confused the confidence interval for the *mean estimate* with the confidence interval of a *population* (see the Central Limit theorem; Wheelan, 2013), leading to a conclusion which was inconsistent with the results presented in the paper itself. Here, we use the same notation as in the original paper, where the subscript 'SE' denotes the error estimate, and $\sigma_{SE} = \sigma / (N-1)^{1/2}$, where N is the sample (model ensemble) size. DCPS07 mistook $\pm 2\sigma_{SE}$ to be the confidence interval of the data sample, whereas the correct interpretation should be the confidence interval for the estimated ensemble *mean value*. In other words, the analysis involved an incorrect description of the population characteristics, and the statistic that DCPS07 really wanted was the interval describing the population of the trends predicted by climate models, the 95% confidence interval for the population.

There are two additional points which emphasise the miss-characterisation of the population in DCPS07. For one, a population interval does not decrease with increased sample size N , which DCPS07 implied when they used $\sigma / (N-1)^{1/2}$ to describe confidence interval. The population interval is determined by the pdf and is invariant to the sample size (the range, on the other hand, tends to increase with the population size). To illustrate this point, we set up an experiment where we define a random normally distributed variable with mean value (first moment) of zero and a unit standard deviation (second moment) involving a random number generator to make a sample of 1000 values. This experiment is presented in the 2 lines of R-code below:

```
> testDCPS07 <- function(N) {x = rnorm(N,0,1); sSE = 1/sqrt(N-1);  
100*sum(abs(x) > 2*sSE)/N}  
  
> testDCPS07(1000)
```

If we used the criterion from DCPS07, $\sigma_{SE}=1/(N-1)^{1/2}$, then for $N=1000$ we see that about 94% of the values fall outside $\pm 2\sigma_{SE}$. We can repeat this with $N=10$, and the answer would be 60%, and for $N=10000$, we get 99.5%.

The second point is that their choice of confidence interval was not self-consistent, and hence is logically flawed. A simple evaluation of the self-consistency for the confidence interval used in DCPS07 involves testing the individual model results against the population statistics. This is done in replicationDemos based on values in their tables I and II (obtained from copying the digital values from a PDF-version of the paper), and for some vertical levels, the confidence limit in DCPS07 excluded up to 59% of the models from which it was derived. In other word, the method used in DCPS07 implied that the population was inconsistent with the samples from which the population was derived.

DCPS07 used the confidence interval for the mean value rather than for the sample and their test constituted an evaluation of how many models were consistent with the mean of the ensemble. A more appropriate confidence interval would be taking the first and second moments without dividing by the sample size: $\mu \pm 2\sigma$. The internal inconsistency and logical flaw of the analysis indicate that the conclusions of DCPS07 could have been dismissed, even before Santer et al. (2008) highlighted additional flaws in the paper. A critical evaluation of the method should provide sufficient indication that it was unsuitable for the type of testing intended in the paper, although literacy within statistics too should prevent a mix-up between different concepts (Bektas, 2013). The paper belongs to category A: starting from a correct logical premise but executing an erroneous analysis.

The replication of the DCPS07 is implemented with the following command lines in R:

```
> DCPS2007()
```

Failure to account for the actual degrees of freedom

A paper by McKittrick and Michaels (2004a; MM04) claimed that much of the historical temperature trends could be explained from local economic activity, level of literacy, and the heat island effect. The analysis was based on a multiple regression analysis between local temperature trends and a set of economic co-variates, where the regression was used to test hypothesised links between economic indicators and local temperature trends. Hypothesis testing and regression analyses generally assume that each data value in a sample is independent in terms of the other values, and that there is no dependency between the data points. In other words, each number represents a measurement which does not depend on any other measurement making up the sample. Potential dependencies between the data, such as correlations, must be accounted for by e.g. the estimation of the effective sample size (Wilks, 1995). However, multiple regression is prone to over-fit when a large number of variables are included (Wilks, 1995; Benestad and Schmidt, 2009), and a proper evaluation of models based on such regression must involve out-of-sample tests against independent data. For such tests, it's important to ensure that there are no dependencies between the training data and the evaluation data.

The analysis of MM04 involved national data and historical temperature trends from a network of thermometers, and for countries with more than one temperature record, the national economic data would be the same and hence not independent. Furthermore, long-term temperature anomalies are smooth functions in space, and the temperature trends are correlated across the borders of nations. MM04 did not take these aspects into account and hence

the effective degrees of freedom, as pointed out by Benestad (2004). Proper testing needs to account for the fact that the economic co-variates would contain the same data within the border of each country, and temperature trends are smooth functions in space. Benestad (2004) replicated the MM04 analysis and applied an out-of-sample which involved splitting the data according to latitude, using one part (50%) for calibration and the other (50%) for independent evaluation. The results from this test demonstrated that the analysis of MM04 was flawed. Out-of-sample tests must ensure that there were no dependencies between the samples used for calibration and evaluation, and hence these samples would involve data from different regions. The latitude suffices for splitting the data into two equal-sized sub-samples for which the dependency due to spatial correlation is minimized (though not quite eliminated), and the analysis was repeated for one sub-sample, making predictions for the other. The evaluation presented in Benestad (2004) showed that multiple regression was unable to predict the trends corresponding to those in the independent sample, given the set of economic indicators suggested in MM04, showing that the multiple regression analysis in MM04 was over-fit.

The criticism presented in Benestad (2004) was not heeded; McKittrick and Michaels, (2007) repeated the claim of made in MM04 without acknowledging the criticism presented in Benestad (2004). While McKittrick and Michaels (2004b) responded to this, they defended their original positions by dismissing the criticism stating they were "*unaware of any paper in the refereed applied climatology literature that has performed the test [splitting the sample, using one for model calibration and the other for validation] suggested by Dr. Benestad... if he has ever seen such a test applied anywhere in a published atmospheric science paper he should have provided an example, which he did not*". While it may be true that they were ignorant about such studies, there are plenty of literature on such tests, typically used to evaluate statistical forecast models and various forms for cross-validation. Split sample (also referred to as out-of-sample) tests are often the norm for testing statistical models (Benestad et al., 2008, 2007; Wilks, 1995). Furthermore, a

valid model is expected to provide good predictions regardless of how the splitting is performed. Hence, MM04 drew their conclusion based on inappropriate statistics, not recognising that the temperature trends vary slowly over space, and their regression analysis misapplied weights to the different covariates resulting in poor predictions of independent data. Furthermore, recent studies indicated that the analysis for the ground temperatures is in accordance with the satellite-based analyses (Foster and Rahmstorf, 2011). MM04 can be listed under category C: overstating the importance. The paper ignores relevant contextual information, such as literature on model evaluation, physical inter-relationships and the warming of the oceans, melting glaciers, and increasing sea level rise, which cannot be explained by the economic variables considered in MM04.

The replication of the MM04, as done by Benestad (2004), is implemented with the following command lines in R:

```
> MM2004()
```

Misconceived mathematics

Although statistics and mathematics often have much in common, they may be distinguished in how they derive conclusions. Here, the term statistics has been used to refer to ways we can draw conclusions from a random sample, describe probabilities, and the way a finite sample is representative of a larger universe. The category 'mathematics', on the other hand, is used here to describe logical abstracts and the set of well-defined absolute rules applying to those, such as algebra and geometry. Using this distinction, the following 'mathematics' section puts more emphasis on equations and valid ways to transform a term between equally valid forms of representation.

Incorrect interpretation of mathematics

McIntyre and McKittrick (2005; MM05) claimed that the reconstruction carried out by Mann et al. (1999, 1998) resulted from inappropriate data processing before a principal component analysis (PCA). They attributed the shape of the curve describing the reconstruction (“hockey stick shape”) to the leading principal component (PC), and argued that since it had a ‘hockey stick shape’ the results were likely an artifact. They argued that red noise processes tend to produce such shapes if the data were not ‘centred’ before computing anomalies.

PCA is a common way of transforming a data matrix (X) into a new set of basis functions in data space, while keeping its information intact. It is a type of eigenvalue analysis in data space, and is also known as empirical orthogonal functions (Lorenz, 1956). The purpose is often to reorganise the data in a way that makes use of the redundancy in the data and makes subsequent analysis faster and less prone to incorrect weighting. For instance, some patterns may recur and be responsible for more of the variance than others, and these are represented in the first modes (principal components). The modes in PCA describe eigenvectors in data space, and can be implemented through singular vector decomposition (SVD): $X = UWV^T$ (Press et al., 1989; Strang, 1988), where U and V contain sets of orthogonal vectors. Here W is a diagonal matrix holding the singular values in descending order, and these values describe how much variance each of the singular vectors account for in the dataset. When there is a degree of coherence, the last values in W represent noise, and hence PCA can provide a means of reducing noise taking n to be the number of principal components containing the signal:

$$x' = \sum_i u_i w_i v_i, \quad i=1, \dots, n$$

There are ways to determine the most appropriate number for n (Wilks, 1995), however, for regression analyses it is important that the subset of principal components describe most of the variance. The fact that the principal

components are orthogonal makes them suitable for regression analysis (Benestad and Schmidt, 2009). The shape of the leading PC is not really relevant as the original information is recovered through the matrix multiplication of the different components according to the above equation. Hence, the question that matters is how many components are included in the subsequent weighting of these components, and how much of the variance is embedded in these components.

MM05 neglected the calibration involved in the process of reconstructing the past temperatures, and failed to address the important question of how many PCs were included in the calibration and how much of the variance they could describe. This failure suggest that they did not understand the process, as the shape of each individual PC, which they stressed, is less relevant as regression analyses weight the different PCs according to how well they match the calibration data. Another point is that the actual 'blade' of the hockey stick graph were not a result of the PCA, but consisted of instrumental data which had been added to the reconstructions (Mann, 2012).

The arguments presented in MM05 were irrelevant for the question they wanted to address, i.e. whether the PCA used in Mann et al. (1999, 1998) would lead to spurious results. However, the general features of the Mann et al. (1998,1999) reconstruction have also been found in other independent analyses (Solomon et al., 2007), and the work has been further evaluated by the Committee on Surface Temperature Reconstructions for the Last 2,000 Years, National Research Council (2006). Furthermore, the MM05-paper were criticised by Wahl and Ammann (2007), Huybers (2005), and Von Storch and Zorita (2005). These criticisms, however, did not convince McIntyre and McKittrick, and further exchange followed in the literature (Mann et al., 2009; McIntyre, 2005a, 2005b; McIntyre and McKittrick, 2009). The source code for the Mann et al. (1998) analysis has been available on-line since 2005⁴, although there have been accusations of not sharing the data and the code. In summary, we would classify MM05 as category C: the analysis does not

⁴ http://www.meteo.psu.edu/holocene/public_html/shared/research/MANNETAL98/

address the actual question. The paper ignored the context of the analysis, i.e. that PCA is only a pre-processing step before a regression-type analysis.

Implausible physics

It is assumed that all valid explanations of natural phenomena are consistent with the laws of physics, obeying the universal conservation laws in terms of energy, mass, charge, and momentum. Here, the term 'physics' is also used to describe the connection between different phenomena and processes taking place, and the link between the cause and effect.

Lack of plausible physics

Scafetta (2010) assumed that changes in the Earth's rotation rate, which he somehow associated with climate variability, is entirely due to planetary forcing (Jupiter and Saturn and the effects their alignment has on their gravitational forces), neglecting other factors such as changes in the circulation in the earth's interior, which may be more important (Appell, 2012). There is no known mechanism explaining how the climate responds to minute changes in the planet's rotation rate, and Scafetta offered no estimates for the Coriolis force or sensitivity tests with different values for the Coriolis coefficient.

The idea of changes in Earth's rotation affects climate was picked up by Solheim and Humlum in the Norwegian magazine '*Fra Fysikkens verden*' (1/11), however, it can be traced back to a conference proceeding from 1992 (Mörner, 1992). It is hard to trace this idea further back in time than Mörner (1992), as 14 out of the 15 citations in his paper were made to his own work. These ideas have been promoted to Norwegian schools through the organisation „klimarealstene“, who include Solheim and Humlum. The most appropriate category for this paper would be B: applying erroneous analysis (not excluding other explanations), and the paper ignored contextual information such as physical mechanisms and other disciplines of geophysics (dynamics of the core).

Incomplete account of the physics

Miskolczi (2007, 2010) attempted to calculate the significance of greenhouse effect through estimating how much of the upwelling infra-red radiation (IR) is absorbed in the atmosphere. He purported that the atmosphere is saturated with respect to CO₂. The claims made in this paper have been promoted by organizations such as 'Friends of Science', been propagated through the Internet, and contributed to the misguided idea that the increases in the CO₂ concentrations have little effect on the global mean temperature.

The Miskolczi (2010) paper was published in same journal as Beck (2008), *Energy & Environment*, and arrived at wrong conclusions due to neglecting relevant physics, such as convection, latent heat of evaporation and sensible heat. Based on these two examples, the journal gives the impression of lacking thoroughness, as Miskolczi (2010) is also difficult to follow since the manuscript has the character of being an unfinished draft with undefined terms, and making few references to relevant previous work; 6 of the 19 citations were to his own work while only 11 references could be considered as scientific journals.

His calculations for the atmospheric absorption of upwelling IR neglected latent and sensible heat fluxes, e.g. associated with vertical motions due to adjustment by hydrostatic stability. This negligence alone invalidates his results, as the time scale associated with hydrostatic adjustment is shorter than the time scale of reaching local radiative equilibrium. Miskolczi's calculations also assumed that the amount of absorbed upwelling IR from the ground equals the downwelling IR from the atmosphere, and that the height, from which the bulk of the outgoing long-wave radiation (OLR) emissions occur, is insensitive to the atmosphere's optical depth. The former is hard to justify if the re-emission from the atmosphere is isotropic, as a volume of air is expected to emit equal amount of IR radiation upward and downward. The total amount of IR emitted by the air is expected to balance the amount of IR received from the ground if the atmosphere is in equilibrium, transparent to

sunlight, and has no other source of energy. The latter claim would mean that an observer viewing IR from space would see down to the same height level even if the IR optical thickness increases, which logically doesn't make sense. Miskolczi (2010) also argued that atmospheric moisture has decreased, in contrast to independent observations (e.g. see the NOAA climate indicators⁵). The conclusion drawn by Miskolczi is also difficult to consolidate with the situation on Venus, which has a heavy atmosphere that mostly consists of CO₂ and has a potent greenhouse effect (Pierrehumbert, 2011). Miskolczi's analysis failed on multiple accounts and his conclusions are invalid, and can be described as a category A: executing an erroneous analysis, which also neglected relevant information such other energy fluxes and relevant literature on the atmospheric physics, and hence lacks the comprehensive picture.

Irrelevant analogies

Gerlich and Tschuschner (2009) claim to have falsified the existence of an atmospheric greenhouse effect by comparing it to a heat pump driven by an environment that is radiatively interacting with planetary atmosphere but equilibrated to the atmospheric system. Their conclusion was based on a misguided comparison between glass houses and the atmospheric greenhouse effect. A comment by Halpern et al. (2010) showed that their methods, logic, and conclusions were in error, e.g. by their attempt to apply the Clausius statement of the Second Law of Thermodynamics to only part of the process. Like Miskolczi (2010), they ignored most non-radiative heat flows applicable to the Earth's surface and atmosphere. Another similarity was publishing in a journal that did not specialise on atmospheric or planetary physics: International Journal of Modern Physics B, Condensed Matter Physics; Statistical Physics; Applied Physics. As with Miskolczi, the paper is an example of an erroneous analysis: category B.

⁵<http://www.ncdc.noaa.gov/bams-state-of-the-climate/2009-time-series/humidity>

Misrepresentation of publications

The regression analysis discussed in Benestad and Schmidt (2009; BS09) was misrepresented in the papers Scafetta (2013a,b), and a comment on the papers has been published to set the records straight (Benestad, 2013). In the abstract of BS09, it is stated that *"We demonstrate that naive application of linear analytical methods such as regression gives nonrobust results"*. The paper iterates this point further *"The regression analysis ... should in this context be regarded as a naive approach that is prone to yielding biased results, and we caution against using such techniques without a critical interpretation"*, and *"Here we use the regression to demonstrate how spurious results may arise from colinearity and "noise" by examining the variability in the coefficients"*. Scafetta (2013a,b) turned this around and accused the paper for inappropriate use of this method: *"An improper application of the multilinear regression method is found in Benestad and Schmidt (2009), indicated herein as BS09"* and *"The first way BS09 multi-linear regression fails is mathematical. The predictors of a multilinear regression model must be sufficiently linearly independent, i.e. it should not be possible to express any predictor as a linear combination of the others"*. Furthermore, Scafetta incorrectly gave the impression that a regression with 10 covariates was used for the comparison and the conclusion of a 7% solar contribution.

One similarity between Scafetta (2013a,b) and the papers Gerlich and Tschuschner (2009) and Miskolczi (2010) is the use of journals not specialising in atmospheric physics, and hence one explanation for lacking rigour may be related to the reviewing process. One of the reviewers of the Scafetta (2013b) paper was the same Mörner as mentioned above, with a similar mindset as Scafetta in terms of external forcings on climate. There are some similarities between Scafetta's thesis and the purports made by Humlum et al. (2011a), who all joined up on a recent paper defending "the planetary theory of solar variation" (Scafetta et al., 2013c). While Scafetta has been a contributing author to the NIPCC, Humlum et al. have written extensively for popular science and engineering magazines in Norway ('Fra Fysikkens Verden',

1/11, p. 21-26; 'Teknisk ukeblad', 16/11, 10/11), in which they have promoted the purports from a number of the cases presented here. In addition to a misrepresentation which is hard to believe has been made in good faith, Scafetta has shown little interest in getting to the bottom of disputed questions (Le Page, 2009). Furthermore, Scafetta (2013d) maintains his position and claims that these points, pointed out in Benestad (2013a; B13), themselves are misleading. He argued that "B13 did not find any physical nor mathematical error in S13. Thus, Scafetta (2013a)'s scientific results remain fully confirmed.". The point of the comment is presented above, concerning phrasing rather than more profound physics and mathematics, which were discussed in BS09. Scafetta (2013d) further argue that "*A regression model is misleading also if it is based on just two collinear constructors, as B13 claimed to have done. At the end, BS09's "7 %" claim is only supported by the GISS ModelE prediction; the result remained not validated by robust data analysis and, therefore, BS09's argument falls into circular reasoning"*, however, the analysis discussed in BS09 (eq.4 in BS09) is a replication of Scafetta's own estimate T_{sun} , but with different representation of the total solar irradiance (S). Furthermore, the bi-variate regression discussed in BS09 used to compare solar and GHG forcings concerned a period when the solar forcing is not colinear with the GHG forcing (Benestad, 2013b). These facts were not appreciated in Scafetta (2013d)

References

Appell, D., 2012. A pressing matter. *Physics World* 37-41.

Bard, E., Delaygue, G., 2008. Comment on "Are there connections between the Earth's magnetic field and climate?" by V. Courillot, Y. Gallet, J.L. Le Mouél, F. Fluteau, A. Genevey *EPSL* 253, 382, 2007. *EPSL* 265.

Beck, E.-G., 2008. 50 YEARS OF CONTINUOUS MEASUREMENT OF CO₂ ON MAUNA LOA. *Energy & Environment* 19.

Bektas, M., 2013. *Statistical Snacks*, **ASIN:** B00DWJZ9Z2. <http://www.amazon.com/Statistical-Snacks-ebook/dp/B00DWJZ9Z2>

Benestad, R.E., 2013a. Comment on "Discussions on common errors in analyzing sea level accelerations, solar trends and global warming" by Scafetta (2013). *Pattern Recognit. Phys.* 1, 91–92.

Benestad, R.E., 2013b. Are there persistent physical atmospheric responses to galactic cosmic rays? *ERL*, 8 035049 doi:10.1088/1748-9326/8/3/035049

Benestad, R.E., 2002. *Solar Activity and Earth's Climate*. Praxis-Springer, Berlin and Heidelberg.

Benestad, R.E., 2004. Are temperature trends affected by economic activity? Comment on McKittrick & Michaels. *Climate Research* 27, 171–173.

Benestad, R.E., 2005. A review of the solar cycle length estimates. *Geophys. Res. Lett.* 32.

Benestad, R.E., Chen, D., Hanssen-Bauer, I., 2008. *Empirical-Statistical Downscaling*. World Scientific Publishing, Singapore.

Benestad, R.E., Hanssen-Bauer, I., Førland, E.J., 2002. Empirically downscaled temperature scenarios for Svalbard. *Atmospheric Science Letters* 3, Issue 2-4, 71–93.

- Benestad, R.E., Hanssen-Bauer, I., Førland, E.J., 2007. An Evaluation of Statistical Models for Downscaling Precipitation and Their Ability to Capture Long-Term Trends. *International Journal of Climatology* 27, 649–665.
- Benestad, R.E., Schmidt, G.A., 2009. Solar Trends and Global Warming. *JGR* 114, D14101.
- Benestad, R.E., Senan, R., Balmaseda, M., Ferranti, L., Orsolini, Y., Melsom, A., 2010. Sensitivity of summer 2-m temperature to sea ice conditions. *Tellus A*.
- Bye, E., Humlum, O., Stordahl, K., 2011. Fortsatt dominans av naturlige klimaendringer. *Teknisk Ukeblad*.
- Climate of fear, 2010. *Nature* 464, 141–141.
- Cohn, T.A., Lins, H.F., 2005. Nature's style: Naturally trendy. *Geophysical Research Letters* 32
- Committee on Surface Temperature Reconstructions for the Last 2,000 Years, National Research Council, 2006. *Surface Temperature Reconstructions for the Last 2,000 Years*. The National Academies Press, Washington, D.C.
- Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., Way, R., Jacobs, P., Skuce, A., 2013. Quantifying the consensus on anthropogenic global warming in the scientific literature. *ERL*, 8 024024
- Courtillot, V., Gallet, Y., Le Mouél, J.-L., Fluteau, F., Genevey, A., 2007. Are there connections between the Earth's magnetic field and climate? *Earth and Planetary Science Letters* 253, 328–339.
- Damon, P.E., Laut, P., 2004. Pattern of Strange Errors Plagues Solar Activity and Terrestrial Climate Data. *Eos* 85, 370.
- Demuzere, M., Werner, M., van Lipzig, N.P.M., Roeckner, E., 2009. An analysis of present and future ECHAM5 pressure fields using a classification of circulation patterns. *International Journal of Climatology* 29, 1796–1810.

- Douglass, D.H., Christy, J.R., Pearson, B.D., Singer, S.F., 2007. A comparison of tropical temperature trends with model predictions. *International Journal of Climatology*, .
- Foster, G., Rahmstorf, S., 2011. Global temperature evolution 1979–2010. *Environmental Research Letters* 6, 044022.
- Franzke, C., 2012. On the statistical significance of surface air temperature trends in the Eurasian Arctic region. *Geophysical Research Letters* 39, n/a–n/a.
- Friis-Christensen, E., Lassen, K., 1991. Length of the solar cycle: an indicator of solar activity closely associated with climate. *Science* 254, 698–700.
- Gerlich, G., Tscheuschner, R.D., 2009. FALSIFICATION OF THE ATMOSPHERIC CO₂ GREENHOUSE EFFECTS WITHIN THE FRAME OF PHYSICS. *Int. J. Mod. Phys.*
- Halpern, J.B., Colose, C.M., Ho-Stuart, C., Shore, J.D., Smith, A.P., Zimmermann, J., 2010. COMMENT ON "FALSIFICATION OF THE ATMOSPHERIC CO₂ GREENHOUSE EFFECTS WITHIN THE FRAME OF PHYSICS". *Int. J. Mod. Phys. B* 24, 1309–1332. *B* 23, 275–364.
- Humlum, O., Solheim, J.-E., Stordahl, K., 2011a. Identifying natural contributions to late Holocene climate change. *Global and Planetary Change* 79, 145–156.
- Humlum, O., Solheim, J.-E., Stordahl, K., 2011b. Spectral Analysis of the Svalbard Temperature Record 1912–2010. *Advances in Meteorology* 2011, 1–14.
- Humlum, O., Stordahl, K., Solheim, J.-E., 2013. The phase relation between atmospheric carbon dioxide and global temperature. *Global and Planetary Change* 100, 51–69.
- Huybers, P., 2005. Comment on "Hockey sticks, principal components, and spurious significance" by S. McIntyre and R. McKittrick. *Geophysical Research Letters* 32.

Idso, C., Singer, F., 2009. Climate Change Reconsidered: 2009 Report of the Nongovernmental Panel on Climate Change (NIPCC). The Heartland Institute.

Idso, C.D, Carter R.M., and Singer S.F. 2013. (Eds.) Climate Change Reconsidered II: Physical Science. Chicago, IL: The Heartland Institute.

Kantelhardt, J.W., Koscielny-Bunde, E., Rego, H.H., Havlin, S., Bunde, A., 2001. Detecting long-range correlations with detrended fluctuation analysis. *Phys. Stat. Mech. Its Appl.* 295, 441–454.

Keenlyside, N., Latif, M., 2002. Enso Hindcast Performance of The Mpi-om/echam5 Coupled Model. Nice.

Kern, Z., Leuenberger, M., 2013. Comment on “The phase relation between atmospheric carbon dioxide and global temperature” Humlum et al. [*Glob. Planet. Change* 100: 51–69.]: Isotopes ignored. *Glob. Planet. Change* 109, 1–2.

Laken, B.A., Pallé, E., Čalogović, J., Dunne, E.M., 2012. A cosmic ray-climate link and cloud observations. *Journal of Space Weather and Space Climate* 2, A18.

Lassen, K., Friis-Christensen, E., 1995. Variability of the solar cycle length during the past five centuries and the apparent association with terrestrial climate. *J. Atmos. Terr. Phys.* 57, 835–845.

Laut, P., 2003. Solar activity and terrestrial climate: an analysis of some purported correlations. *Journal of Atmospheric and Solar-Terrestrial Physics* 65, 801–812.

Lockwood, M., Frölich, C., 2008. Recent oppositely directed trends in solar climate forcings and the global mean surface air temperature. II. Different reconstructions of the total solar irradiance variation and dependence on response time scale. *Proc. R. Soc. A* 464, 1367–1385.

Loehle, C., Scafetta, N., 2011. Climate Change Attribution Using Empirical Decomposition of Climatic Data. *The Open Atmospheric Science Journal* 5, 74–86.

- Lorenz, E., 1963. Deterministic nonperiodic flow. *J. Atmospheric Sci.* 20, 130–141.
- Lorenz, E.N., 1956. Empirical Orthogonal Functions and Statistical Weather Prediction (Sci. rep. No. 1). Department of Meteorology, MIT, USA.
- 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.
- Mann, M. E., Bradley, R. S., Hughes, M. K., 1999. Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations. *Geophys. Res. Lett.* 26, 759–762.
- Mann, Michael E, 2012. *The hockey stick and the climate wars: dispatches from the front lines.* Columbia University Press, New York.
- Mann, Michael E., Bradley, Raymond S., Hughes, Malcolm K., 2009. Reply to McIntyre and McKittrick: Proxy-based temperature reconstructions are robust. *PNAS* 106, E11–E11.
- Marcott, S.A, J.D. Shakun, P.U. Clark, and A.C. Mix, 2013, A Reconstruction of Regional and Global Temperature for the Past 11,300 Years, *Science* 8 March 2013: 339 (6124), 1198-1201. [DOI:10.1126/science.1228026]
- Masters, T., Benestad, R., 2013. Comment on “The phase relation between atmospheric carbon dioxide and global temperature”. *Glob. Planet. Change* 106, 141–142.
- McIntyre, S., McKittrick, R., 2005. Hockey sticks, principal components, and spurious significance. *Geophys. Res. Lett.* 32, L03710.
- McIntyre, Stephen, 2005a. Reply to comment by von Storch and Zorita on “Hockey sticks, principal components, and spurious significance”. *Geophysical Research Letters* 32.
- McIntyre, Stephen, 2005b. Reply to comment by Huybers on “Hockey sticks, principal components, and spurious significance”. *Geophysical Research Letters* 32.

- McIntyre, Stephen, McKittrick, Ross, 2009. Proxy inconsistency and other problems in millennial paleoclimate reconstructions. *PNAS* 106, E10–E10.
- McKittrick, R., Michaels, P.J., 2004. A test of corrections for extraneous signals in gridded surface temperature data. *Climate Research* 26, 159–173.
- McKittrick, R., Michaels, P.J., 2004b. Are temperature trends affected by economic activity? Reply to Benestad (2004). *Climate Research* 27, 175–176.
- McKittrick, R., Michaels, P.J., 2007. Quantifying the influence of anthropogenic surface processes and inhomogeneities on gridded global climate data. *J. Geophys. Res.*
- Meehl, G.A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J.F.B., Stouffer, R.J., Taylor, K.E., 2007. The WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research. *Bull. Amer. Meteor. Soc.* 88, 1383–1394.
- Miskolczi, F.M., 2007. Greenhouse effect in semi-transparent planetary atmospheres. *Q. J. Hung. Meteorol. Serv.* 111, 1–40.
- Miskolczi, F., 2010. THE STABLE STATIONARY VALUE OF THE EARTH'S GLOBAL AVERAGE ATMOSPHERIC PLANCK-WEIGHTED GREENHOUSE-GAS OPTICAL THICKNESS. *ENERGY & ENVIRONMENT* 21.
- Mörner, N.A. (1992) 'Present El Niño-ENSO events and past super-ENSO events effects of changes in the Earth's rate of rotation, ORSTOM-CONCYTEC, Lima 1992, p. 201-206.
- National Research Council (U.S.), 2001. *Climate change science: an analysis of some key questions*, The compass series. National Academy Press, Washington, D.C.
- Oreskes, N., 2004. The scientific consensus on climate change. *Science* 306.
- Palmer, T.N., 1996. Predictability of the Atmosphere and Oceans: from Days to Decades, in: Anderson, D.L.T., Willebrand, J. (Eds.), *Decadal Variability*, NATO ASI Series. Springer.
- Pain, H.J., 1983. *The Physics of Vibrations and Waves*, 3rd ed. John Wiley & Sons Ltd.

- Pesin, Y.B., 1977. CHARACTERISTIC LYAPUNOV EXPONENTS AND SMOOTH ERGODIC THEORY. *Russ. Math. Surv.* 32, 55–114.
- Pierrehumbert, R.T., 2011. Infrared radiation and planetary temperature. *Physics Today* 33–38.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1989. *Numerical Recipes in Pascal*. Cambridge University Press, Cambridge, UK.
- Proctor, R., Schiebinger, L. (Eds.), 2008. *Agnotology: The Making and Unmaking of Ignorance*. Stanford University Press.
- R Development Core Team, 2004. *R: A language and environment for statistical computing*. Vienna, Austria.
- Richardson, M., 2013. Comment on “The phase relation between atmospheric carbon dioxide and global temperature” by Humlum, Stordahl and Solheim. *Glob. Planet. Change*.
- Richardson, I.G., Cliver, E.W., Cane, H.V., 2002. Long-term trends in interplanetary magnetic field strength and solar wind structure during the twentieth century. *Journal of Geophysical Research* 107.
- Rypdal, K. (2013) Interactive Comment Interactive Discussion, *Earth Syst. Dynam. Discuss.*, 4, C305–C310, 2013
- Santer, B.D., Thorne, P.W., Haimberger, L., Taylor, K.E., Wigley, T.M.L., Lanzante, J.R., Solomon, S., Free, M., Gleckler, P.J., Jones, P.D., Karl, T.R., Klein, S.A., Mears, C., Nychka, D., Schmidt, G.A., Sherwood, S.C., Wentz, F.J., 2008. Consistency of modelled and observed temperature trends in the tropical troposphere. *International Journal of Climatology* 28, 1703–1722.
- Scafetta, Nicola, 2010. Empirical evidence for a celestial origin of the climate oscillations and its implications. *Journal of Atmospheric and Solar-Terrestrial Physics* 72, 951–970.
- Scafetta, Nicola, 2012a. Testing an astronomically based decadal-scale empirical harmonic climate model versus the general circulation climate models. *Journal of Atmospheric and Solar-Terrestrial Physics* 80, 124–137.

Scafetta, Nicola, 2012b. A shared frequency set between the historical mid-latitude aurora records and the global surface temperature. *Journal of Atmospheric and Solar-Terrestrial Physics* 74, 145–163.

Scafetta, Nicola, 2012c. Multi-scale harmonic model for solar and climate cyclical variation throughout the Holocene based on Jupiter–Saturn tidal frequencies plus the 11-year solar dynamo cycle. *Journal of Atmospheric and Solar-Terrestrial Physics* 80, 296–311.

Scafetta, N., West, B.J., 2005. Estimated solar contribution to the global surface warming using the ACRIM TSI satellite composite. *Geophys. Res. Lett.* 32.

Scafetta, N., West, B.J., 2006a. Phenomenological solar signature in 400 years of reconstructed Northern Hemisphere temperature record. *Geophys. Res. Lett.* 33, L17718.

Scafetta, N., West, B.J., 2006b. Phenomenological solar contribution to the 1900–2000 global surface warming. *Geophysical Research Letters* 33.

Scafetta, N., West, B.J., 2007. Phenomenological reconstructions of the solar signature in the Northern Hemisphere surface temperature records since 1600. *Journal of Geophysical Research* 112, D24S03.

Scafetta, Nicola. "Discussion on Common Errors in Analyzing Sea Level Accelerations, Solar Trends and Global Warming." *arXiv:1305.2812* (May 13, 2013a). doi:10.5194/prp-1-37-2013.

Scafetta, N. "Discussion on Common Errors in Analyzing Sea Level Accelerations, Solar Trends and Global Warming." *Pattern Recognition in Physics* 1, no. 1 (May 13, 2013b): 37–57. doi:10.5194/prp-1-37-2013.

Scafetta, N., Humlum, O., Solheim, J.-E., Stordahl, K., (2013c) Comment on "The influence of planetary attractions on the solar tachocline" by Callebaut, de Jager and Duhau. *J. Atmospheric Sol.-Terr. Phys.*

Scafetta, N.: Reply to Benestad's comment on "Discussions on common errors in analyzing sea level accelerations, solar trends and global warming" by

- Scafetta (2013d), *Pattern Recogn. Phys.*, 1, 105-106, doi:10.5194/prp-1-105-2013, 2013.
- Shaviv, N.J., 2002. Cosmic Ray Diffusion from the Galactic Spiral Arms, Iron Meteorites, and a Possible Climatic Connection. *Phys. Rev. Lett.* 89, 051102.
- Solheim, J.-E., Stordahl, K., Humlum, O., 2011. Solar Activity and Svalbard Temperatures. *Advances in Meteorology* 2011, 1–8.
- Solheim, J.-E., Stordahl, K., Humlum, O., 2012. The long sunspot cycle 23 predicts a significant temperature decrease in cycle 24. *Journal of Atmospheric and Solar-Terrestrial Physics* 80, 267–284.
- Solomon, S., Quin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), 2007. *Climate Change: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, United Kingdom and New York, NY, USA.
- Stauning, P., 2011. Solar activity–climate relations: A different approach. *Journal of Atmospheric and Solar-Terrestrial Physics* 73, 1999–2012.
- Stephenson, G., 1973. *Mathematical methods for science students*, 2nd ed. Longman Scientific & Technical, London.
- Strang, G., 1988. *Linear Algebra and its Application.* Harcourt Brace & Company, San Diego, California, USA.
- Svensmark, Henrik, 1998. Influence of Cosmic Rays on Earth's Climate. *Physical Review Letters* 81, 5027–5030.
- Svensmark, Henrik, 2007. *The chilling stars: a new theory of climate change.* Icon, Thriplow, Cambridge.
- Svensmark, H., Friis-Christensen, E., 1997. Variations of cosmic ray flux and global cloud coverage - a missing link in solar-climate relationships. *J. Atmos. Sol. Terr. Phys.* 59, 1225.
- Veizer, J., 2005. *Celestial Climate Driver: A Perspective from Four Billion Years of the Carbon Cycle.* *Geoscience Canada* 32.

- Von Storch, H., Zorita, E., 2005. Comment on "Hockey sticks, principal components, and spurious significance" by S. McIntyre and R. McKittrick. *Geophysical Research Letters* 32.
- Wahl, E.R., Ammann, C.M., 2007. Robustness of the Mann, Bradley, Hughes reconstruction of Northern Hemisphere surface temperatures: Examination of criticisms based on the nature and processing of proxy climate evidence. *Climatic Change* 85, 33–69.
- Waldmeier, M., 1961. The sunspot-activity in the years 1610-1960. Zurich Schulthess & Company AG.
- Weart, S.R., 2004. *The Discovery of Global Warming*. Harvard University Press.
- Wheelan, C. 2013. *Naked Statistics, STRIPPING THE DREAD FROM THE DATA*, ISBN 978-0-393-07195-5
- Wilks, D.S., 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press, Orlando, Florida, USA.
- Wilks, D.S., 2006. On "Field Significance" and the False Discovery Rate. *Journal of Applied Meteorology and Climatology* 45, 1181–1189.
- Williams, R.W., 1960. Fundamental Formulas of Physics, in: D.H.Menzel (Ed.), Dover books, pp. pp544–562.
- Yndestad, H., 2006. The influence of the lunar nodal cycle on Arctic climate. *ICES Journal of Marine Science* 63, 401–420.

GISP2 temperature

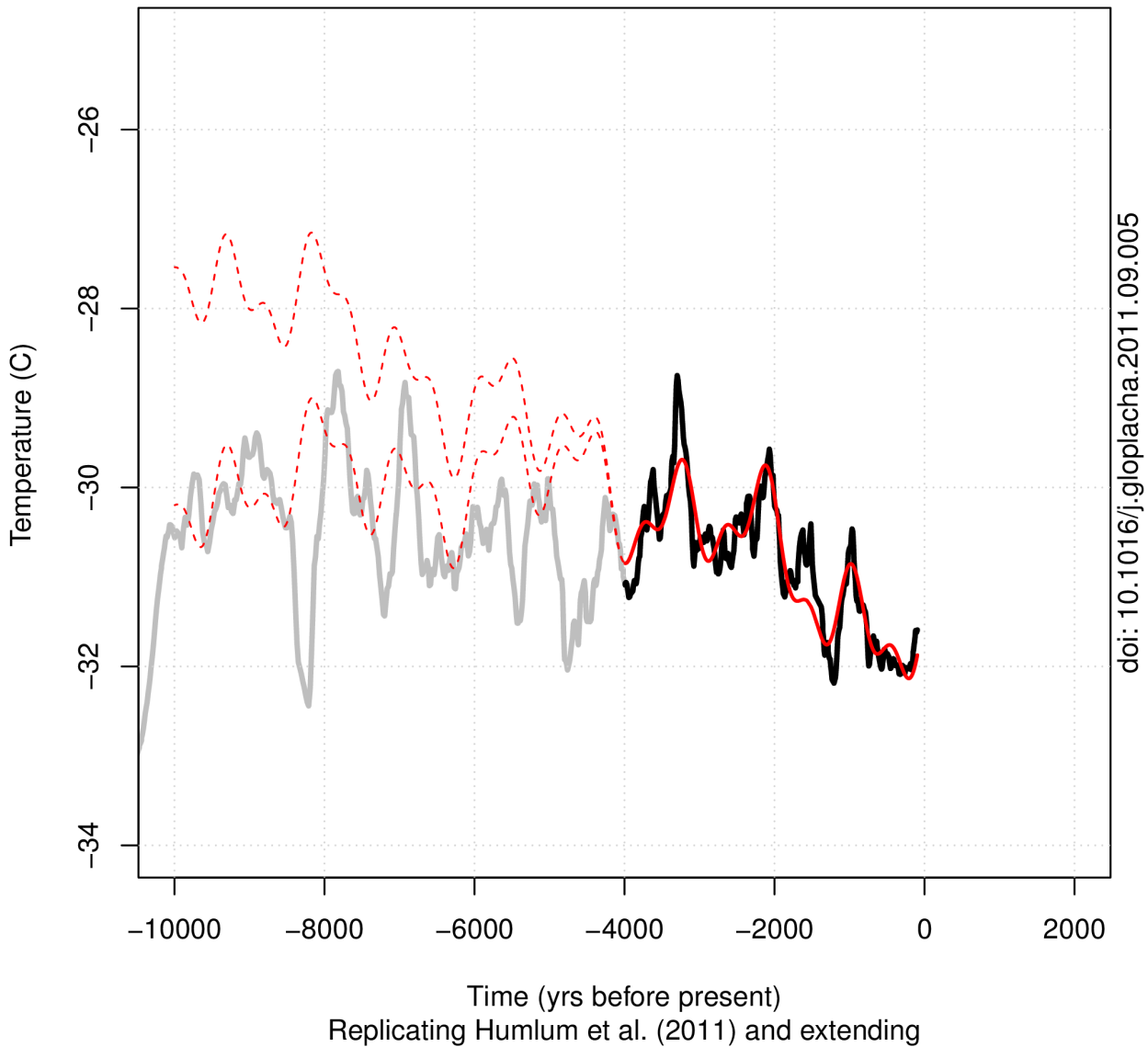


Figure S1. A replication of Humlum et al.'s model for the GISP2-record (solid red) and extensions back to the end of the last glacial period (red dashed). The two red dashed lines represent two attempts to extend the curve fit, one keeping the trend over the calibration interval and one setting the trend to zero. The black curve shows the part of the data showed in Humlum et al. and the grey part shows the section of the data they discarded. The figure can be reproduced from 'replicationDemos' with the function 'Humlum.et.al.2011()'.

After Scafetta (2012): Fig. 5b

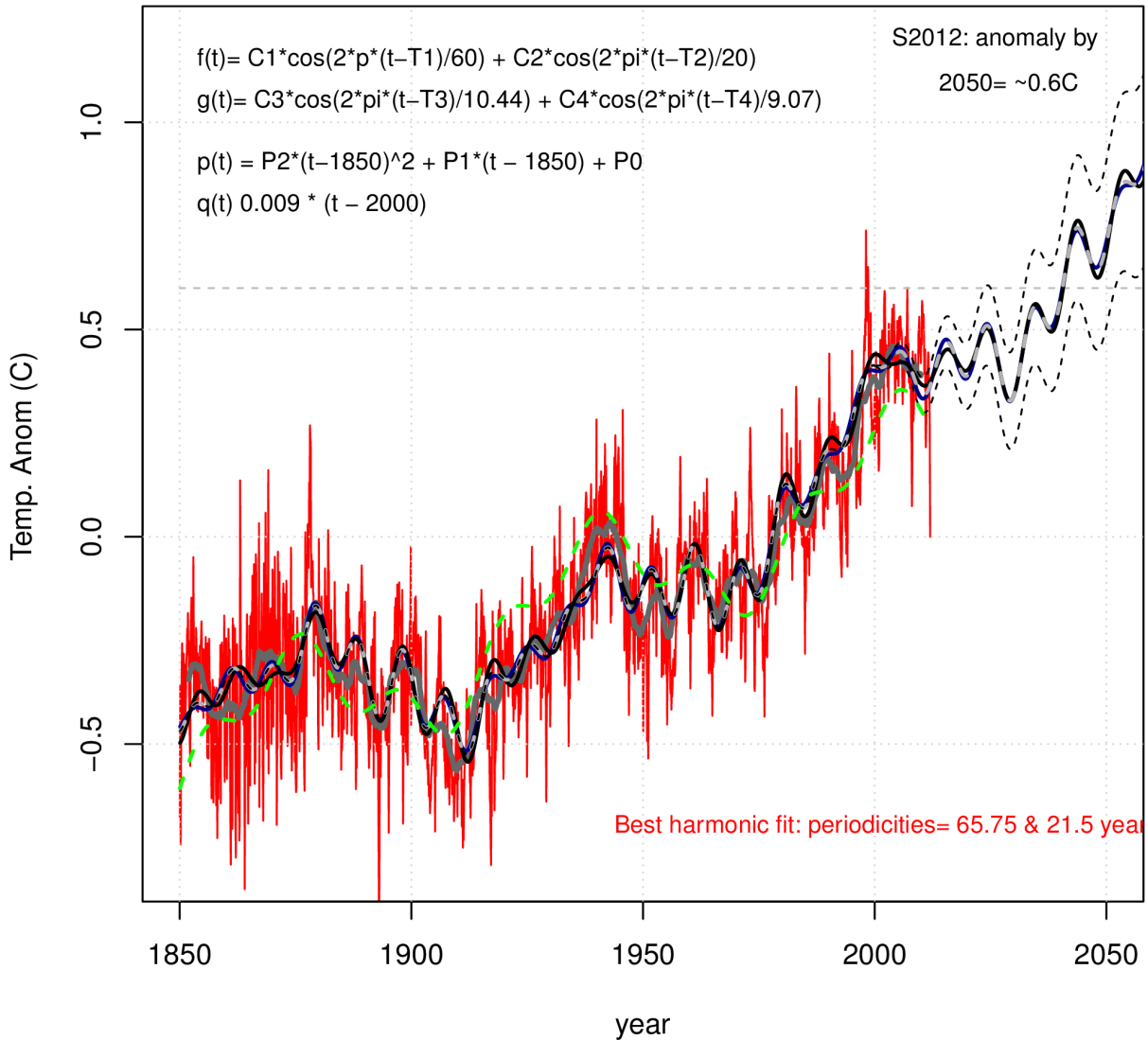


Figure S2. A reproduction of figure 5 in Scafetta, N., 2012a (also available on-line from <http://arxiv.org/pdf/1201.1301v1.pdf>). The reproduction was done calling the function 'Scafetta2012()' in replicationDemos. The grey horizontal dashed line marks the level where Scafetta's curve intersects year 2050 whereas the blue/black/grey curves are based on Scafetta's model using the trend and values of the coefficients reported in his paper. An iterative search for periodicities in the vicinity of those suggested by Scafetta, gave a best fit to the pair of harmonics if they were 21.5 and 65.75 years respectively. Green dashed line show alternative fit based on the fit to 65.75 and 21.5 year periodicities as well as a linear trend that yielded the greatest R2.

Scafetta (2012): Table 1

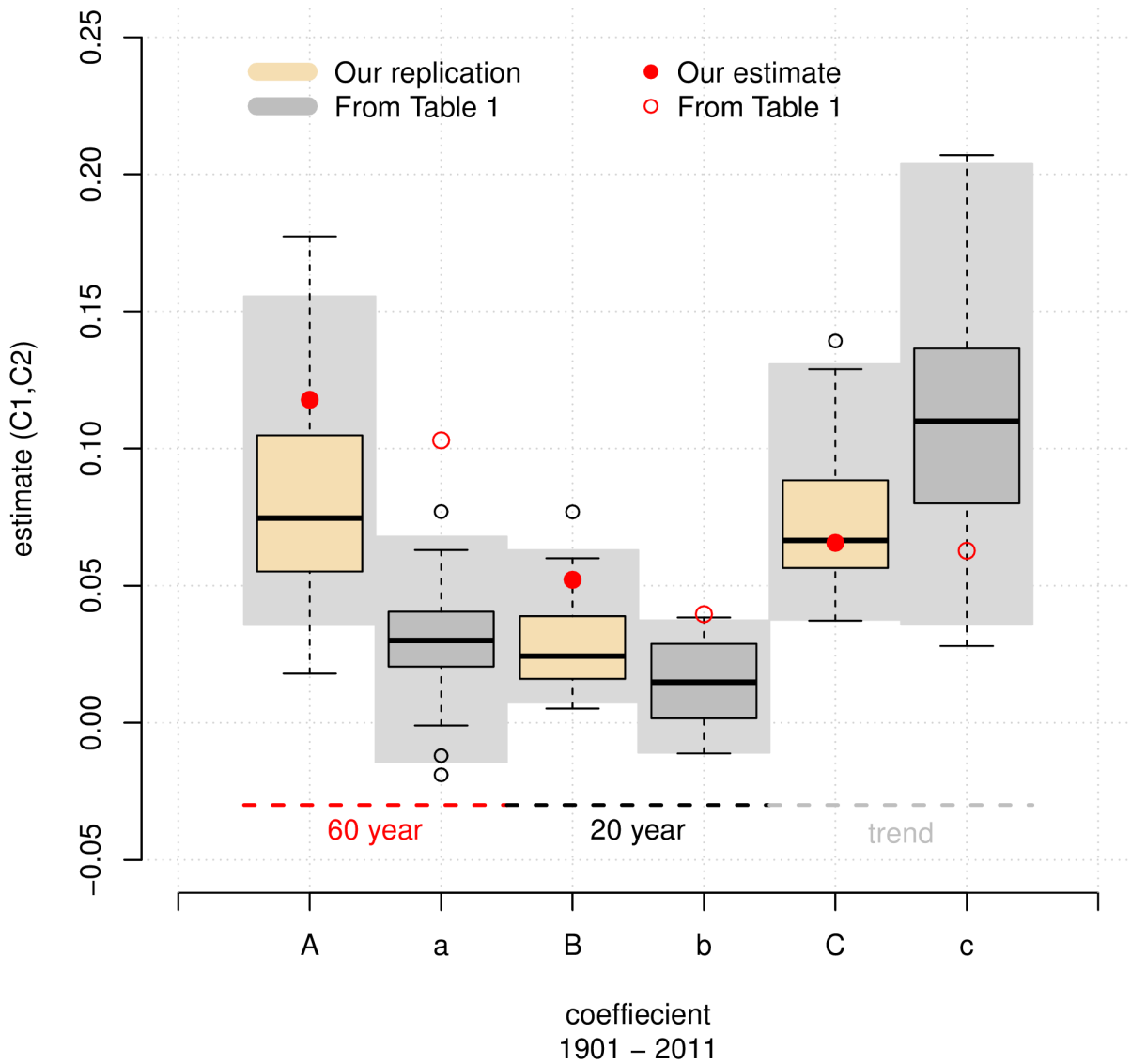
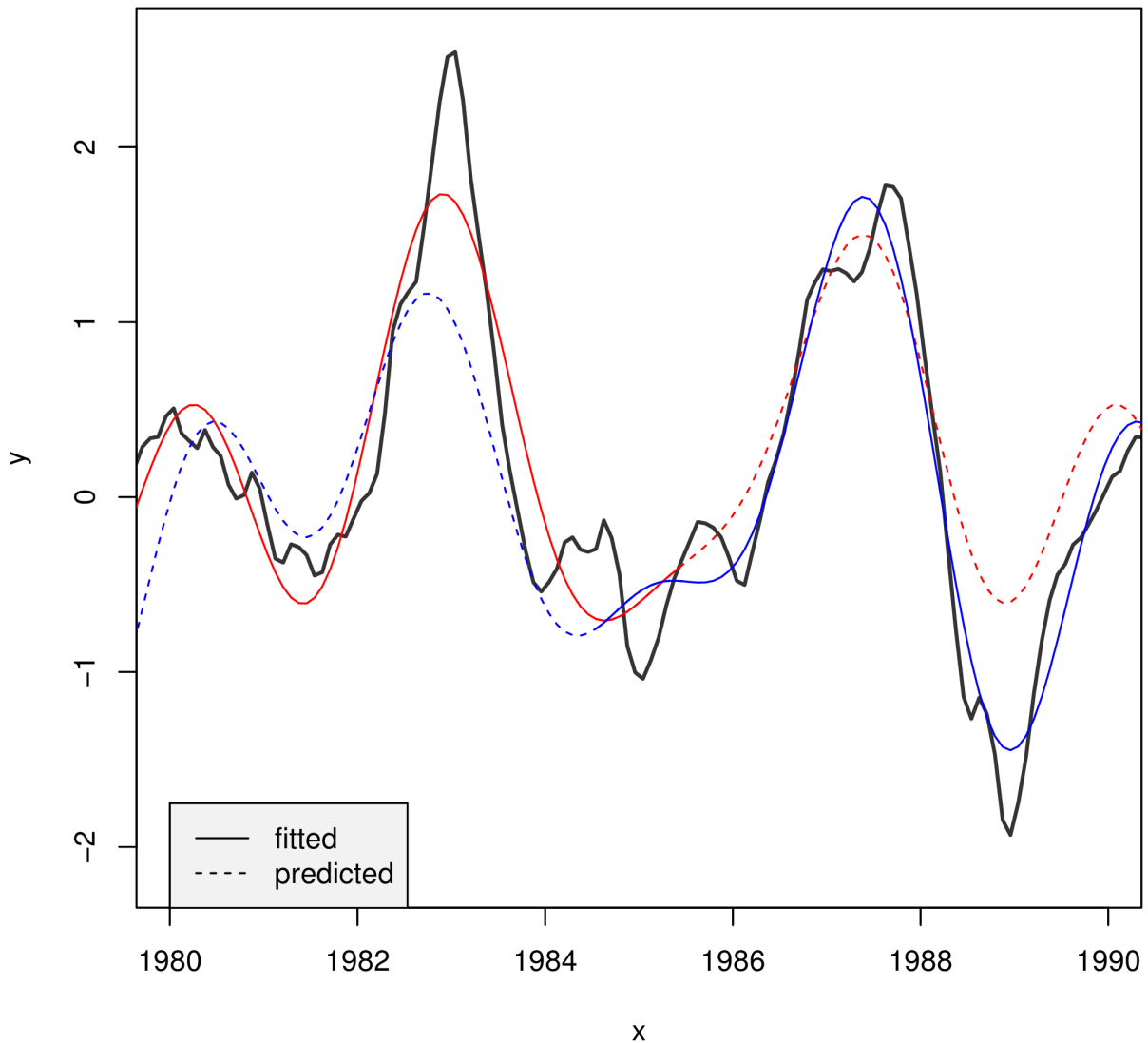


Figure S3 compares the amplitudes for the 60 (a) and 20-year harmonics (b) and trends (c) from the GCMs (box) and the HadCRUT3v (red symbols). Capital letters on the x-axis refer to re-calculated regression coefficients whereas the lower-case letters refer to those in Scafetta (2012). The comparison shows that the amplitudes of the 20 and 60-year variability found in the observational record (filled circles) are within the range simulated by the GCMs. The grey boxes are values copied from Table 1 in Scafetta, (2012a), whereas the yellow boxes are re-computed with no constraints on the phase. The boxes mark the middle 50% of the GCM results (i.e. the interquartile range). See Table 1 for description of the functions used to generate this figure. Here the values for a, b, and c in Scafetta's table 1 have been divided by 0.1, 0.04, and 0.1 respectively in order to provide comparable values. The grey backgrounds

represent the 90% confidence intervals. The whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. Scafetta.tab1()

NINO3.4 curve fit



Scafetta (2012): ...distinguishes a mere curve fitting exercise.. p. 9

Figure S4. An example demonstrating that for a small interval of the ENSO cycle, it is possible to find two cycles which seem to be part of a regular oscillation. This demonstration is produced with the call 'ENSO.example()', which also shows a comparison between the fits and the rest of the data – for which they fail to track the ENSO evolution.

Table S1. An overview of functions in the 'replicateDemos' package, a short description of their demonstrations/replication and the reference to the paper for which they are relevant.

R function	Description	Reference
Humlum.et.al.2011()	Extends the curve-fit beyond the fitted interval for Greenland temperature.	Humlum et al. (2011a)
Humlum.et.al.2012()	Repeats lagged correlation between differentiated series - where trends are effectively removed.	Humlum et al. (2013)
Solheim.et.al.2011()	Replicates analysis based on data given in tables: SCL, temperature, correlations, and bootstrapping.	Solheim et al. (2011)
LoehleScafetta2011()	Some replication and some demonstrations showing how curves segments are made up of harmonics, and how a fit within an interval fails to describe the remaining part of the curve.	Loehle and Scafetta (2011)
Scafetta2010()	Some replication and some demonstrations showing how different temperatures change results, and how different noise processes may seem to contain long-term cycles.	Scafetta (2010)
ENSO.example()	Example showing how the El Niños during the 1980s resembled sinousoids, and that calibrating a curve-fit on one half will give a good match for the other half. But the fit fails outside this segment.	Scafetta (2012a)
resonance()	Simulation of how a damped oscillator forced with noisy signal produces oscillations with fixed frequencies.	Loehle and Scafetta (2011); Scafetta (2012a, 2010)
Scafetta2012()	Replicates Fig 5b in Scafetta (2012a)	Scafetta (2012a)
Scafetta.tab1()	Replicates and visualises Table 1 in Scafetta (2012a).	Scafetta (2012a)
Scafetta2006()	The R-script used by Benestad & Schmidt (2008), modified to be part of the R-package.	Scafetta and West (2005, 2006a, 2006b)
MM04()	The R-script used to carry out the analysis of Benestad (2004)	McKittrick and Michaels (2004)
Douglas2007()	Replication and evaluation of the GCM confidence interval based on data in tables.	Douglass et al., (2007)
paleaproxy()	Comparison between cosmogenic Be-10 isotope proxies, CO ₂ , and temperature from the Vostoc ice core.	Veizer (2005)
DJF()	Examine the winter temperature at Svalbard and the forecast made by SSH11	Solheim et al., (2011)
testLTP()	Test the LTP assumption and the way long-term trends affect the auto-correlation.	(Cohn and Lins, 2005; Franzke, 2012)

Table S2. Overview of papers which have attempted to identify and attribute cycles in Earth's climate to external causes.

Author/year	Journal	Periodicity (yr)	Comment
Humlum et al. (2011a)	Glob. Plan. Change	16.8, 25.7, 68.4, 556, 1186, 2804	
Humlum et al. (2011b)	Advances in Meteorology	71.7, 24.9, 15.3, (74.3, 24.5, 17.1)	
Scafetta and West (2006b)	GRL	7.3-14.7, 14.7 - 29.3	35-60% and 20-40% of 1900-2000 and 1980-2000 warming respectively
Scafetta and West (2007)	JGR	11	up to ~50% of warming since 1950
Scafetta (2012a)	J. Astr. Terr. Phys	9.1,10-10.5,20-21,60-62	at least 60% of warming since 1970
Loehle and Scafetta (2011)	Open Atm. Sci. J.	20, 60	
Scafetta (2012c)	J. Astr. Terr. Phys	9.98, 10.9, 11.86, 61, 115, 130	
Yndestad (2006)	J. Marine Science	6, 18, 74	

Table S3. Overview of the contrarily papers reviewed here and a summary of the type of error identified.

Humlum et al. (2011a)	Curve-fit; selective use of data; ignore part of the data; insufficient model evaluation; missing contextual information; implausible physics; cycle matching; false dichotomy	Analytical set-up questioned	Start with wrong premise.
Humlum et al. (2011b)	Curve-fit; insufficient model evaluation; missing contextual information; implausible physics; cycle matching; false dichotomy	Analytical set-up questioned	Start with wrong premise.
Yndestad (2006)	Implausible physics; single local variable; cycle; insufficient model evaluation; missing contextual information; cycle matching; false dichotomy	Analytical set-up questioned	Start with wrong premise.
Loehle and Scafetta (2011)	Curve-fit; lack of model evaluation; missing contextual information; ; false dichotomy	Analytical set-up questioned	Start with wrong premise.
Scafetta (2012a)	Implausible physics; curve-fitting; questionable statistics; biased analysis; unjustified assumptions; missing contextual information	Analytical set-up questioned	Start with wrong premise and faulty analysis.
Scafetta (2012b)	Curve-fit; lack of model evaluation; missing contextual information; false dichotomy	Analytical set-up questioned	Start with wrong premise.
Scafetta (2012c)	Curve-fit; lack of model evaluation; missing contextual information; false dichotomy	Analytical set-up questioned	Start with wrong premise.
Solheim et al., (2011)	Selective use of data; implausible physics; missing contextual information; questionable statistics	Analytical set-up questioned	Faulty analysis.
Solheim et al., (2012)	Selective use of data; implausible physics; missing contextual information; questionable statistics; false dichotomy	Analytical set-up questioned	Faulty analysis.
Scafetta and West (2005)	questionable presumptions; missing model evaluation; missing contextual information; false dichotomy	Analytical set-up questioned	Start with wrong premises.
Scafetta and West (2006a)	questionable presumptions; missing model evaluation; missing contextual information; false dichotomy	Analytical set-up questioned	Start with wrong premises.
Scafetta and West (2006b)	questionable presumptions; missing model evaluation; missing contextual information; false dichotomy	Analytical set-up questioned	Start with wrong premises.
Scafetta and West (2007)	questionable presumptions; missing model evaluation; missing contextual information	Analytical set-up questioned	Start with wrong premises.
Beck (2008)	Questionable representation in data; missing contextual information;	Analytical set-up questioned	Not addressing the correct question.
Humlum et al. (2013)	Biased set-up; analysis of irrelevant temporal scales; missing contextual information	Analytical set-up questioned	Addressing wrong question
Veizer (2005)	Unconvincing statistics; missing contextual information; false dichotomy	Analytical set-up questioned	Not addressing the correct question.
Friis-Christensen and Lassen (1991)	Questionable processing of data; missing contextual information; vague physics; false dichotomy	Analytical set-up questioned	Start with wrong premise
Lassen and Friis-Christensen (1995)	Questionable processing of data; missing contextual information; vague physics; false dichotomy	Analytical set-up questioned	Start with wrong premise
Svensmark (1998)	Questionable processing of data; missing contextual information; vague physics; false dichotomy	Analytical set-up questioned	Start with wrong premise
Svensmark and Friis-Christensen (1997)	Questionable processing of data; missing contextual information; vague physics; false dichotomy	Analytical set-up questioned	Start with wrong premise
Svensmark (2007)	Questionable processing of data; missing contextual information; vague physics; false dichotomy	Analytical set-up questioned	Start with wrong premise
Shaviv (2002)	Selective use of data; missing contextual information; vague physics; false dichotomy	Analytical set-up questioned	Start with wrong premise
Courtillot et al. (2007)	Selective use of data; missing contextual information; questionable statistics; false dichotomy	Analytical set-up questioned	Start with wrong premise and faulty analysis.
Cohn and Lins (2005)	Circular reasoning; questionable choice of null-model; missing contextual information	Analytical set-up questioned	Start with wrong premise
Franzke (2012)	Circular reasoning; questionable choice of null-model; missing contextual information	Analytical set-up questioned	Start with wrong premise
Douglass et al. (2007)	Mix-up of statistical concepts; missing contextual information	Statistical errors	Faulty analysis.
McKittrick and Michaels (2004a)	Failure to account for effective degrees of freedom; missing contextual information.	Statistical errors	Faulty analysis.
McKittrick and Michaels (2004b)	Failure to account for effective degrees of freedom; missing contextual information.	Statistical errors	Faulty analysis.
McIntyre and McKittrick (2005)	Failure to see the mathematical context; false dichotomy	Misconceived mathematics	Faulty analysis.
McKittrick and Michaels, (2007)	Failure to account for effective degrees of freedom; missing contextual information.		Faulty analysis.
Scafetta (2010)	Neglect of a number of factors; missing contextual information	Implausible physics	Start with wrong premises.
Mörner (1992)	Vague physics; missing contextual information; no proper science document; false dichotomy	Implausible physics	Start with wrong premises.
Miskolczi (2007)	Missing heat fluxes; missing contextual information; questionable presumptions	Implausible physics	Start with wrong premises and faulty analysis.
Miskolczi (2010)	Missing heat fluxes; missing contextual information; questionable presumptions	Implausible physics	Start with wrong premises and faulty analysis.
Gerlich and Tschuschner (2009)	Misplaced analogy; missing contextual information	Implausible physics	Start with wrong premises and faulty analysis.
Scafetta (2013a)	Incorrect account of published work	Misrepresentation of publications	Faulty analysis.
Scafetta (2013b)	Incorrect account of published work	Misrepresentation of publications	Start with wrong premises and faulty analysis.