J.-E. Solheim (JES)
Kjell Stordahl (KS)
Ole Humlum (OH)
Harald Yndestad (HY)

janesol@online.no June 14, 2013

Comments to Earth System Dynamics Discussion paper "Agnotology: Learning from mistakes" by R. E. Benestad, H. O. Hygen, R. van Dorland, J. Cook and D. Nuccitelli (BHDCN-Paper)

Our comments regular*, quotations from BHDCN in italics*

Introduction

The accusation made in the BHDCN-paper that we belong to a .. *culture neglecting replication, not sharing methods and data, or not testing the methods..*is completely wrong and fails totally if our papers are read correctly. The claim that we have hidden data, and give incomplete description of methods is wrong. The data used are all from open sources, and our methods are described in our papers. Other researchers may use other methods and reach other conclusions. New data may appear which give different results. We may have done mistakes, and that will of cause be corrected by repeating the investigations by ourselves or other scientists – and other methods may of cause give different results.

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

Page 457

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

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

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

BHDCN: *The core of the analysis carried out by Humlum et al. (2011a) involved curve-fitting and tenuous physics, with a vague idea that the gravity of solar system objects somehow can affect the Earth's climate.*

OH: The analysis is not based on curve fitting, but on wavelet analysis, which is something else. There is no mention of the gravity of the solar system anywhere in the paper, and this point of criticism is apparently also made up by BHDCN.

Page 459:

BHDCN: *They claimed that they could "produce testable forecasts of future climate" by extending their statistical fit, and in fact, they did produce a testable forecast of the past climate by leaving out the period between the end of the last ice age and up to 4000 yr before present. However, they did not state why the discarded data was not used for evaluation purposes, and the problem with their model becomes apparent once their fit is extended to the part of data that they left out.*

We extended their analysis back to the end of the last ice age. Figure 1 shows our replication of part of their results (Table 1). It clearly shows that the curve-fit for the selected 4000 yr does not provide a good description for the rest of the Holocene.

OH: On page 145-46 in the paper discussed we state the following: "The second series, the GISP2 data has merit because of the long time range represented (back to the Eemian interglacial), and because the Greenland air temperature appears to vary in overall concert with the temperature of much of the planet (Chylek and Lohmann, 2005; Brox et al., 2009). Here we chose to focus on the most recent 4000 years of the GISP2 series, as the main thrust of our investigation is on climatic variations in the recent past and their potential for forecasting the near future. In addition, this part of the GISP2 series shows an overall linear temperature trend, which simplifies the following analysis". Here we clearly state that the GISP2 record is much longer than just 4000 years, and explain why we make use of only the last 4000 years.

On page 152 we define limitation of our approach by writing "This suggests that natural variations which are strong now are likely to continue without major changes at least some time into the future, and therefore likely to influence also the future climate. This knowledge on persistent and strong natural variations may be used for an attempt of forecasting, at least some time into the near future".

On page 155 we write "Our empirical experience suggests a realistic forecasting time range of about 10–25% of the total record length. In the case of Greenland, such forecasting suggests that the present post LIA warm period is likely to continue for most of the 21st century, before the overall Late Holocene cooling may again dominate, but this being dependant on the magnitude of the anthropogenic greenhouse enhancement".

By this the limited forecasting range is clearly defined, and a 'hindcasting of more than 6000 years based on the 4000 year record used as indicated by Benestad is highly inappropriate and is not in any way reflecting what we are writing in the paper.

BHDCN*: Furthermore, they failed to acknowledge well-known shortcomings associated with curve-fitting, but rather based their analysis on unjustified fit to a set of Fourier series.*

OH: Once again, we did not employ curve fitting, but wavelet analysis, which is something quite else.

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

There are certain conditions which have to be satisfied, to make a forecasting model. One example: the data (time series) have to be transformed to stationarity in the forecasting process. (When the analyses are finished, the analyzed data are retransformed). This is the case for ARIMA models, wavelet analysis (which is used in Humlum et al (2011a)) etc. Hence, the ability of the time series will in many cases influence of the selection of the time period used for the analyses and model building.

BHDCN use the word "curve fitting" for modeling the temperature evolution. "Curve fitting" is a misleading expression. It is important to underline that the forecasting method, wavelet analysis, is a rather advanced forecasting method.

Another point: The long-term temperature time series reflect temperature the last glacial period and the last interglacial period. For a forecaster point of view, it is important to analyse the temperature evolution the last part of the interglacial and NOT the last glacial period when temperature forecasts for the last part of the interglacial are developed.

JES: Another strawman argument used by BHDCN appears when they explain that a function can be fitted by a sum of Fourier series having no predictability at all (outside of the data interval), hereby indicating that the periods we find cannot be used for predictions.

This is correct if the series contain just noise, but it is our experience from research on light variations from stars and planets, that it is often possible to find stationary, or nearly stationary periodic signals, which can be predicted to continue at least in the near future. We only used the 3 most significant periods, and we tested the predictive power by deleting a part of the series. This is shown for the Svalbard data as a proof of the validity of the method. Our conclusion was that it was safe for the forecast about 10-25 per cent (of the series length) in the future. So we have done exactly what BHDCN claim we have not.

Extensive Monte Carlo tests (S.O.Kepler 1993, Baltic Astronomy 2, 515) have proved the validity of FT periods with amplitudes above the FAP (false alarm probability) as developed by J.D. Scargle (1982, Astrophysical Journal, 263, 835). We used only the strongest periods fulfilling this criterion.

The ice-core analysis was done for 4000 year input, and we made a forecast for the period 1850-2800. This forecast shows a temperature increase 1900-2000, which is observed***. Our model is then clearly verified outside the range analysed, and may be a good predictor for the next couple of hundred years.*** Our model shows that the temperature may reach a maxium during this century, and that we are close to that maximum now (figure 8).

Page 468

Case 4: ignoring negative tests

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

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

Page 469

BHDCN*: Furthermore, the analysis was not objective, inflating the significance of the results. A more subtle aspect of this study was the number of attempts to find a correlation, and the lack of accounting for all the tests in the evaluation of the significance of the results. There is a good chance of seeing false fortuitous correlations if one examines enough local temperature records.*

JES: We have not inflated anything, just reported the result of our investigation and the methods used. The claim that we selected only local series where we did find correlations is wrong. We selected temperature series from places with long records, away from (large) cities, preferably stations at light houses on the coast. The results from all tests are reported in our paper II (SSH12). In the Svalbard paper (SSH11) the analysis was done in more detail for one location, as we demonstrated that the correlation, which for all places were found for yearly data, was significant only for the winter temperatures.

BHDCN*: When we reconstructed their Table 1 we got nearly the same results, albeit not identical. SSH2011 stated that they based their method for estimating SCL on a publication from 1939 (Waldmeier, 1961), however, more recent work on the estimation of SCLaccount for uncertainties in estimating the true SCL as the sunspot record exhibits stochastic variations around the slow Schwabe cycle. Rather than estimating the SCL from the few data points around the solar minima, Benestad (2005) proposed to use a Fourier truncation to fit the sunspot record and hence use the entire data sample to estimate the SCL.*

In particular, SSH2011's estimate of the SCL for cycle 23 (12.2 yr) was substantially longer than the estimate of 10.5 yr reported by the Danish Meteorological Institute (based on Friis-Christensen and Lassen (1991) and follow-up studies) and 10.8 yr estimatedby Benestad (2005) (Table 1). Such a long cycle is the basis for their projected cooling (a decrease from −11.2 to −17.2 _C with a 95% confidence interval of −20.5 to −14 _C) at Svalbard over solar cycle 24 (starting 2008).

JES: Several errors here: We have not estimated the SCL, but used the official SCL determined by an international committee based on several parameters and reported by NGDC (National Geophysical Data Center). Their SCL is based on the Waldmeier (1961) definition with certain additions, and is specified as follows: " When observations permit, a date selected as either a cycle minimum or maximum is based in part on an average of the times extremes are reached in the monthly mean sunspot number, in the smoothed monthly mean sunspot number, and in the monthly mean of spot groups alone. Two more measures are used at time of sunspot minimum: the number of spotless days and the frequency of occurrence of old and new cycle spot groups".

The length of 12.2 yrs for cycle 23 is simply the length calculated by NGDC two years *after* the cycle was finished. We consider this as the OBSERVED CYCLE

LENGTH. This must be a far better estimate than the length estimated by DMI and Benestad (2005), who estimated the length *before* the cycle had ended. It should also be remarked that the solar cycle 24 started officially in 2009. It is not possible to check the estimates of Benestad (2005) for SCL, since table 1 in that paper is nonexistent.

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

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

BHDCN*: SSH2011 used a weighted regression to account for errors of the mean temperature estimates over the periods corresponding to solar cycles. Hence they accounted for errors in the mean estimate, but neglected the errors associated with the SCL, which are more substantial than the errors in the mean seasonal or annual temperature over 10 yr segments*.

JES: There are two possible errors in SCL that affects the analysis.

The first is in the temperature averages. The definition of winter is the months DJF. If the SCL starts in Jan or Feb, the month of Dec (prev. yr) should belong to the previous cycle. In these cases there will be one monthly temperature associated with the wrong cycle. This happened only once for the series, since cycle 17 started 1944.2. Similar for the year average temperature. We did not split years, so temperatures for cycles starting .5 or later was calculated from the next year, similar if it ended after .5: temperatures for the rest of the year was included in the cycle. We assume that these errors are included in the sigma calculated for the mean temperature. The temperature averages were taken over 10, 11 or 12 years, as SCL to the nearest year, not in 10 yrs segments.

The second error is in the temperature forecast, where an error in SCL, should be added to the estimated error. We assume that the SCL by NGDC is calculated with the precision of 0.05 yrs, which means an error 1/20 of the temperature change for one year longer/shorter SCL. For Svalbard this would be ±0.05C for the yearly average and ±0.1C for the winter temperature. This is small if we compare with the estimated errors (sigma) in the forecast which for the year average is σ = 0.7C and the winter σ = 0.9C. We disagree with BHDCN that the error in SCL is substantial and larger than the error in the mean and seasonal temperatures.

BHDCN*: They also applied a bootstrapping approach to estimate the errors in the correlation coefficients (between −0.52 and −0.97), as they argued that there is no analytical expression to do so…the claim made by SSH2011 that there is no analytical expression for estimating confidence intervals for correlation is false.*

KS: The statement by BHDCN that there is no analytical expression for estimating confidence intervals for correlation is false, is wrong.

An approach for estimation of the confidence interval for the correlation coefficient which do not give exact values, is the following:

The probability distribution function f(r) of the correlation coefficient ,r, is given in: http://mathworld.wolfram.com/CorrelationCoefficientBivariateNormalDistribution.html.

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

2.5% = f(r(low))

97.5% = f(r(high))

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

The calculations of the confidence interval based on the mentioned Z-transform is described in: http://support.sas.com/documentation/cdl/en/procstat/63104/HTML/default/viewer.htm#procstat_corr_sect018.htm

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

Page 470

*Their estimate of the errors in the correlation involved 1000 picks of random paired sub-samples from the SCL and temperatures, where the same pair sometimes were picked more than once. A more appropriate strategy would be to carry out a set of Monte-Carlo simulations accounting for the errors due to the SCL (_S) and mean temperature estimates (_**T**).*

KS: BHDCN raise a discussion about usage of the bootstrap methodology contra a special Monte Carlo analysis for estimating correlation statistics. This type of discussion should be directed to other statistical bodies and not to a specific paper where bootstrap methodology is used. The bootstrap methodology has been used of statisticians for many years now, and is accepted as a methodology for giving deeper insight in statistical data analysis.

Page 471

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

KS: The BHDCN approach is to apply Monte Carlo analysis where the standard deviation of the difference between the SCL estimates from SSH2011 and Benestad (2005) are used to the SCL simulations. However, the different quality of the SCL estimates – especially the weakness in Benestad (2005) SCL estimates, destroy the quality of these simulations.

Another aspect is the quality of the Monte Carlo simulations itself. The Monte Carlo simulations are based on simulations of the temperature. For applying simulation of the temperature evolution, BHDCN calculates the temperature variation by estimation of the mean temperature and adding the estimated standard deviation for each observation in the period. BHDCN does not mention that there are significant uncertainties in the estimates of the mean temperature and in the standard deviations. Hence for each of the 80.000 simulations, these uncertainties are generated in every run. Compared to the bootstrap analyses, where each temperature sample is based on real observed data and not constructed observations based on uncertainties in estimated mean and standard deviations.

The conclusion is rather simple, the bootstrap analyses presented in SSH2011 has to be recommended.

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

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

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

BHDCN*: Solheim et al. (2012) expanded the correlation exercises between SCL and temperature to include several locations in the North Atlantic region. The fact that several of these give similar results can be explained from the spatial correlation associated with temperature anomalies on time scales greater than one month. Their analysis involved 6–11 degrees of freedom, depending on the length of the available record, but since they applied their analysis to both SCL with zero and one-period lag, in addition to a number of locations, they would need to account for the problem of multiplicity and apply e.g. the Walker test. The failure to do so will give misleading results.*

JES: We are aware of the spatial correlations, in fact that was one of our conclusions that the effect was strongest (with $r^2 > 0.5$) for stations in or near the North Atlantic – see table 1). We reduced the degree of freedom with one, since we investigated two lags (0 and 1).

Page 472

BHDCN*:The main problem with the analysis presented by SSH2011 was the lack of a convincing physical basis, inappropriate hypothesis testing, the inflation of significance, and a small data sample insufficient to support the conclusions.*

JES: Our aim with the paper was to produce a forecast, not to give a physical explanation. The model is clearly defined. If it works will be proven by observations the next 5-10 years. That the effect (of SCL) clearly is strongest in and near the North Atlantic, invites to speculate on a physical reason related to the ocean currents, and heat transported from warmer locations where the Sun warms the deep ocean.

Unfortunately the length of data series are limited. Still we tested the method by deleting the last observation in each series, and compared it with predictions based on the remaining. All observed (unused values) were in the 95% bracket.

Page 478

Case 9: looking at wrong scales

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

OH: We set out to investigate the short-term variations (DIFF12; using two 12-month windows), as the long term variation is the integral of these short-time variations. On the time scale investigated in our study, changes of atmospheric CO_2 follows after corresponding changes in temperature. For that reason the changes in global air temperature cannot be the result of changes in atmospheric CO_2, while the opposite possibility must be considered viable. If one expands the time window considered from 1 yr to 3, 5 or 8 years, our conclusion still holds true. In short, Benestad et al. attempts to give the reader the false impression that there are mathematical errors in our analysis, which is not the case.

BHDCN: *Humlum et al. chose to analyse a short series from 1980 describing the global analysis of the CO_2 concentrations rather than the almost identical series from Mauna Loa going back to 1958.*

OH: We decided to analyse the shorter global data series for atmospheric CO_2, rather than the somewhat longer Mauna Loa series, because we wanted to use a global average rather than observations from one single point (Mauna Loa). If one instead uses the Mauna Loa series for analysis, a result entirely similar to that published by us is achieved. Presumably BHDCN are well aware of this, but nevertheless leaves the casual reader with the wrong impression that our analysis is wrong.

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

Page 483

Case 13 contamination by other factors

Beck (2010) has repeated his analysis in more detail, and most of the objections mentioned by BHDCN are resolved. He has selected only places where weather conditions are known (wind from the sea, or from isolated islands, ocean crossings etc.). From an estimated total sum of more than 200 000 single samples collected since 1800 in the Northern and Southern hemispheres, he selected 97 404 samples from 901 stations compiled in 87 data files. A thorough literature review revealed that these data are only partly known to climate science and had been rejected and ignored without a thorough validation of the quality, sampling and analyzing methods used. The new dataset contains high quality data including vertical profiles, which can be easily compared to today´s standards. Two periods approximately 70 years apart are identified with slowly rising atmospheric CO_2 levels up to values close to 400 ppm.

Instead of rejecting Beck´s work with hand waving arguments, BHDCN should look carefully at the series Beck has selected as significant, and compare them with modern measurements.

Case 14: incomplete account of the physics

Miskolzi calculates by a high precision line-by-line code the IR optical depth as function of time based on observed radiosonde data for pressure, temperature, humidity and CO_2. The result was that the optical depth is approximately constant the last 50 yrs. And since IR radiation does not care what the absorbers are (CO_2 or H_2O) – it at least tells us that more CO_2 has not changed the radiative transfer. What the physics is to explain this is another question, but data from The International Satellite Cloud Climatology Project (ISCCP) shows that the column of water vapor at 680 -310 mb elevation has decreased from 0.6 to 0.4 cm since 1958 (down 30%). This may have compensated for the extra CO_2 in the atmosphere.

A recent paper by Q.-B. Lu (Inter. Journ. Mod. Phys. B, 27, 1350073) concludes: "Remarkably, a statistical analysis gives a nearly zero correlation coefficient (R=-0.05) between corrected global surface temperature data by removing the solar effect and CO_2 concentration during 1850-1970. In striking contrast, a nearly perfect linear correlation with coefficients as high as 0.96-0.97 is found between corrected or uncorrected global surface temperature and total amount of stratospheric halogenated gases during 1970-2012."

Based on the Lu paper, one get the impression that CO_2 is not important for our climate at its present level.

Page 486

Case 15: differences in preprocessing of data

BHDCN*: Although a connection between solar activity and Earth's climate is plausible, there is no trend in the recent solar indices that can explain the current global warming.*

Use of smoothed values of SCL is discussed also in SSH2012 and we conclude that they are of little value. A test of unsmoothed observed values values give good correlations as discussed in case 4 when a delay of one cycle is introduced.

It may be that the lack of trend in the recent solar indices as written above – combined with the reduction of human-made halogenated gases has resulted in the lack of trend in global temperature observed the last 15 years, as concluded by Lu (2013).

Page 488

Case 17 misinterpretation of spectrographic methods

BHDCN: " *claims that the giant planets exert influence on earths´ climate*"

HY: No. The paper title is "…influence on Arctic climate". Earth climate is a reference to global climate. Arctic climate is a reference to a regional climate, which is something else.

BHDCN " *Yndestad claims to have identified the lunar nodal cycle*".

HY: The lunar nodal cycle has been known since James Bradley in the 17th century, and H G Darwin in the 19th century, and later identified in climate indicators by a number scientists.

BHDCN: "*This study is too based on harmonic analysis*".

HY: No. The study is based on a wavelet analysis of the longest oceanographic data series in the world. Wavelet analysis is not the same as harmonic analysis. It looks like BHDCN has done a misinterpretation of methods.

BHDCN: "*Sea-ice and the local Arctic climate is strongly affected by winds and ocean currents*"

HY: The study compares data series of the Earth axis position, Arctic ice extent in the Barents Sea, Arctic ice extent in Greenland Sea, Kola section water temperature in the Barents Sea, wind at the Røst island, NAO winter index. The estimated periods in the Arctic oceans, are confirmed by other scientists. The same periods are identified in sea level, temperature and salinity in inflow of North Atlantic water to the Norwegian sea, and published in: Yndestad H. 2008 William R Turrell, Vladimir Ozhigin. Lunar nodal tide effects on variability of sea level, temperature, and salinity in the Faroe-Shetland Channel and the Barents Sea. Deep-Sea Research I. 55 (2008) 1201-1217.

BHDCN: "*Arctic climate involves dynamics with a pronounced non-linear chaotic character"*

HY: The wavelet analysis in this study shows that the identified periods have phase-reversals, connected to longer harmonic periods. Phase reversals indicates a time-variant and non-linear system.

BHDCN: "*Case 17 misinterpretation of spectrographic methods*"

HY: There is no reference to the misinterpretation in this text by BHDCN.