Journal of Environmental Science and Engineering, 5 (2011) 1053-1062

# An Alternative Means of Assessing Climate Models

D. Bray and H. von Storch
*Institute for Coastal Research, Helmholtz Center, Geesthacht D21502, Germany*

Received: January 18, 2011 / Accepted: March 21, 2011 / Published: August 20, 2011.

**Abstract:** The authors compared two different sets of assessment of the abilities of contemporary climate models. One group is made of experts, and their results are provided in two expert reports, while the other is the subjective assessment made by "physical climate scientists" in general, sampled in a series of three survey questionnaires. The expert group is considerably more optimistic than the general group; the former suggesting progress, while the perception of the latter group is more or less stationary.

**Key words:** Climate modelling, climate model assessment, survey of climate scientists.

## 1. Introduction

Recently there have been some comprehensive reviews of climate models. These include the "Climate Models and Their Evaluation"-chapter in the IPCC report 2007 [1] and CCSP-Report [2]. The authors of the IPCC chapter state the goal as "This chapter assesses the capacity of the global climate models used elsewhere in this report for projecting future climate change". The latter sets out to determine "What are the major components and processes of the climate system that are included in present state-of-the-science climate models, and how do climate models represent these aspects of the climate system?" and the former, to "... assesses the capacity of the global climate models used elsewhere in this report for projecting future climate change". Both works assess model abilities to simulate observed climate, to produce modeled output that compares favorably to observations and/or the ability to simulate observed climate change. In short, a number of expert authors judge the modeling ability of climate science by assessing the output of models, noting both strengths and weaknesses.

H. von Storch, professor, Ph.D., research field: climate science. E-mail: hvonstorch@web.de.

**Corresponding author:** D. Bray, Ph.D., research fields: sociology of science, sustainability, knowledge transfer, sociology of environment. E-mail: dennis.bray@hzg.de.

For the two reports forming the comparative basis for this analysis, one ([2], ibid) has 8 authors listed and additional contributing 18 scientists, so that a total of 26 people was underwriting this work. Thirteen authors are listed as having written consensually the IPCC Chapter by Randall et al. ([1], ibid), another 74 person have been involved as "contributing authors". Additionally a number of scientists have reviewed the drafts and commented. In the IPCC chapter there are approximately 30 specialized topics listed and it should be noted that the 74 authors are not necessarily specialists in all areas.

Some of the general conclusions of the IPCC Chapter 8, Climate Models and Their Evaluation, for example, state that confidence in model estimates have been enhanced, Atmosphere-Ocean General Circulation Models (AOGCMs) provide credible quantitative estimates of future climate change, it is not clear how well current climate models can capture the impact of future warming on the terrestrial carbon balance, there is some evidence that models have improved over the last several years in simulating the annual cycle of the precipitation patterns and that the global statistics of the extreme events in the current climate, especially temperature, are generally well simulated by the current models, for example. The general conclusions of CSSP 2008 include, for example, climate modeling has been steadily improving over the past several decades, climate model simulation of precipitation has improved over time but is still problematic, models forced by observed well-mixed greenhouse-gas concentrations, volcanic aerosols, estimates of variations in solar energy incidence and anthropogenic concentrations are able to simulate the recorded 20th Century global-mean temperature in a plausible way, modern AOGCMs generally simulate continental and larger-scale mean surface temperature and precipitation with considerable accuracy and Cloud feedbacks remain the largest source of uncertainty in climate sensitivity estimates.

One you note that there appears to be no concise statements, and overall evaluations are given in subjective terms rather than any metric terms (i.e. enhanced confidence, improved, plausible, etc.). Furthermore, when considering the group of authors and contributors to the two expert reports [1, 2] as samples, a caveat has to be made–the selection of the participating persons was likely not random among the experts, and the members of the group had no independent opinions as they interacted intensely with each other. In case of the surveys, examined in this paper in some detail, the sampling is more random, and the answers may be expected to be independent of each other.

Therefore, we offer an alternative approach to this issue, not to relying on experts but on a broader group of physical climate scientists. Taking a different perspective, the approach consists of presenting climate scientists' subjective accounts of selected assessments of climate models over the period 1996-2008. Instead of focusing on model and model output comparisons we explicitly ask members of the physical climate science community to comment on various aspects of modelling. Using the results of a time series of three surveys of climate scientists [3, 4] the state of climate models, as perceived by climate scientists working in the "physics of the climate system, are assessed, and changes in the assessments over time are presented. Information was collected using a series of questionnaire surveys".

## 2. Data

The questions in the survey pertaining to the task at hand are of a very simple design. They are not designed to infer anything about the progress of the models , i.e. How much has x, y, or z model improved over the last ten years, as some respondents in 2008 might not have been in the science in 2003 for example. Rather the questions are designed to measure the here and now appraisal of the models at three points in time.

All three surveys employed non-probability convenience sampling. Convenience sampling provides an inexpensive approximation of truth. Quite simply, the sample is selected because they are convenient. Sampling special groups (scientists) often results in a comparatively difficult sample selection as no complete list of the population is available or even known. It should also be noted that the three surveys of climate scientists do not constitute a panel study but rather a repeated survey and that this does not hinder any longitudinal analysis. To investigate change (in this case the evaluation of the abilities of climate science models) the goal is to measure that same thing. Repeated surveys collect the data from different samples. Panel studies submit the survey questions to the same people over a period of time. Panel studies, then, follow individuals over time. The purpose of the surveys of climate scientists forming the basis of this analysis was to follow the opinion of a collective body over time. Repeated surveys do not capture the actual change of opinion within the individual scientist, i.e. a shift from on polar perspective to another polar perspective. Instead the goal is to capture the effects of all of the changes. (For a more complete account of sampling, response rate and survey logic, see Ref. [5]).

The response categories available to respondents was a range of 1 to 7 with a value of 1 being an indication of “very inadequate” and a value of 7 being “very adequate” or “very much” and “not at all”. In other words, in a scale of 1 to 7 a value of 1 represents on extreme possibility and a value of 7 represents the opposite extreme position. The values are intentionally subjective. What we attempt to capture is the perceptions of a large body of climate scientists, both users and constructors of models, at the time of the survey. It is not an attempt to address a probability level of the model output as this would be knowledge limited to very few of the sample, i.e. too explicit to ask of a large sample. To have done so would likely have resulted in an extremely low response rate or simply in a series of guesses rather than subjective appraisals. Rather than risk these limitations questions were phrased to be applicable to a larger number of respondents. What we do not attempt to do is to discuss the skills of the models in terms of simulation of x or y.

It has been the case that criticisms were forth coming regarding the response categories in the surveys used in this analysis, with claims that “very much, not at all, very adequate”, etc. are not clearly defined. We would like to point out at this point that such responses are not so dissimilar than those of the IPCC Report or the CSSP. Throughout these reports statements are often presents in the form of “improved but still problematic simulate in a plausible way, considerable accuracy, generally well”, etc., in short, subjective appraisals.

The methodical issues were discussed in some detail in peer-reviewed social science literature [5, 6] so that only a brief summary is presented here. In 1996 the survey was distributed by hard copy. In all, the survey covered 5 countries, Germany, USA, Canada, Denmark and Italy. Response rate was approximately 30% for a total of 546 respondents. In 2003, in an effort to reach a broader audience it was decided to conduct an on-line survey. In all there were 557 respondents but the response rate was, of course, not possible to calculate. This sampling method is known as saturation sampling. Saturation sampling attempts to survey all identifiable targets and overcomes the lack of reliable sampling frames. The low cost of internet research makes this possible.

In 2008, attempting to improve the survey but maintain a large sample size, three lists were employed in constructing the sample. List one included a list of authors, affiliations and email addresses drawn from climate journals with the 10 highest ISI impact ratings for the last 10 years. These are authors of climate related papers in peer reviewed climate related journals. The second list was the list of authors who contributed to Oreskes’ [7] conclusions concerning consensus. A third list was drawn from readily available email lists on institute web sites (i.e. NCAR, MPI, AMS, etc.). Duplicates in the three lists were removed before distribution. The combined invitation list numbered a potential 2677 respondents; defunct email addresses reduced the valid mail out to 2059. Provisions were made so that should someone submit a duplicate form the form identifier resulted in the original being over written. Consequently, for each invitation it was only possible to have one completed survey written to the data set. The response rate for ISI authors list was approximately 27%, for Oreskes’ list, approximately 10%, and from the Institute list, approximately 19%, for a combined response rate of 18% (375 responses).

For the purpose of this analysis a subsample of climate scientists working in the “physics of the climate system” (modelling, model development, data acquisition, theory development’) was drawn. This resulted in the following sample sizes: 1996, 282 scientists, 2003, 288 scientists and 2008, 293 scientists.

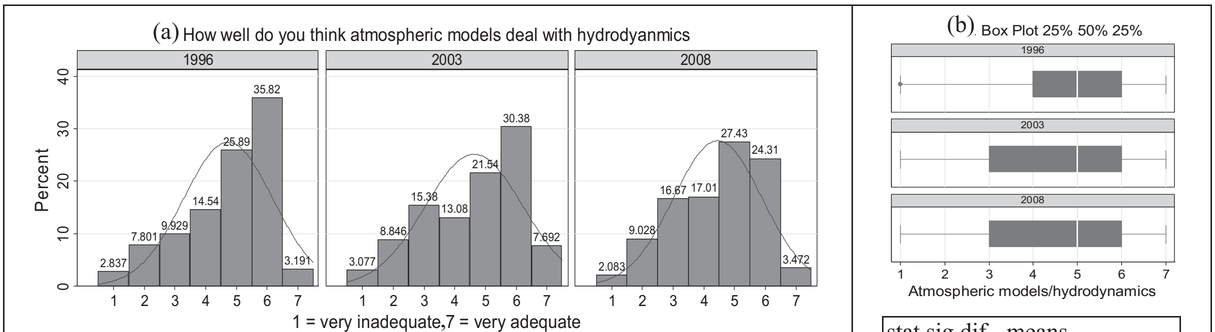
## 3. Components of Models

In all three years scientists were asked to respond to the following sets of questions: How well do you think atmospheric climate models can deal with the following processes: (1) hydrodynamics, (2) radiation, (3) vapour in the atmosphere, (4) the influence of clouds, (5) precipitation and (6) atmospheric convection; For all questions, the possible response categories ranged from 1 very inadequate to 7 very adequate. In a second set of questions pertinent to all three surveys, scientists were asked if they agreed or disagreed with the statements "The current state of scientific knowledge is developed well enough to allow for a reasonable assessment of the effects of: (1) turbulence, (2) surface albedo, (3) land surface processes, and (4) green house gases (specified in 2008 as "emitted from anthropogenic sources")". Response rate categories ranged from 1 strongly disagree to 7 strongly agree.

The responses concerning perceptions of abilities and the state of knowledge are presented in Figs. 1-10. Each figure contains a histogram with the probability density function, a box plot, indicating the median and concentration of perceptions. It is assumed the middle 50 percentile would exclude any radical (activist or skeptic) appraisals and therefore constitute somewhat of a conservative consensus of opinion. Means between years are shown as statistically significant or not.

Fig. 1 presents the analysis of scientists' perceptions of how well atmospheric models can deal with hydrodynamics. According to the responses, the perception of the ability of climate models to deal with hydrodynamics has not changed much, and if anything, has diminished over the years, the means between 1996 and 2008 demonstrating a statically significant difference. However the magnitude of the difference between the means is not great and a mean of approximately 4 or 5 suggests the perception of a need for considerable improvement. A visual analysis of the distributions over time however seems to indicate the perception of a declining ability of models to deal with hydrodynamics. An increase in doubt is also evident in the box plot with the middle 50th percentile moving towards a more negative assessment after 1996.

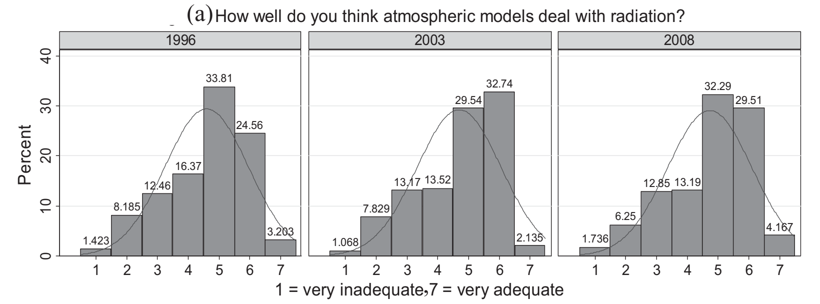
Fig. 2 addresses the perceptions of climate scientists towards the ability of atmospheric models to deal with radiation. Here there are no statically significant differences among the means. The means again, however, suggest that climate modellers perceive a significant amount of room for improvement. Of course, as with any aspect of any science, skepticism and uncertainty are an inherent trait of any assessment. The box plots also indicate little change in assessment over time. In summary, any indicator of change in the perception of the ability of atmospheric models, according to the respondents to the surveys, to deal with radiation has remained somewhat static between 1996 and 2008, and while the mean assessment is no overly high, the shape of the distribution as well as the box plots, suggest a reasonable level of satisfaction with the ability of climate models to deal with radiation.



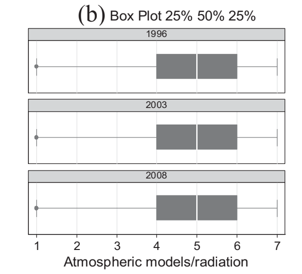

| | 1996 | 2003 | 2008 |
|---|---|---|---|
| n | 282 | 260 | 288 |
| mean | 4.73 | 4.63 | 4.45 |

| stat sig dif - means | |
|---|---|
| 1996-2003 | - |
| 1996-2008 | 0.023 |
| 2003-2008 | - |

**Fig. 1 Perceptions of the ability of atmospheric models to deal with hydrodynamics.**




| | | | |
|---|---|---|---|
| n | 281 | 281 | 288 |
| mean | 4.59 | 4.69 | 4.73 |

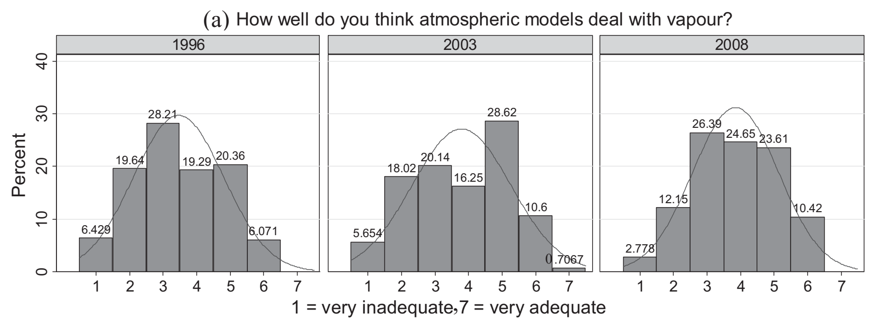



| stat sig dif - means | |
|---|---|
| 1996-2003 | - |
| 1996-2008 | - |
| 2003-2008 | - |

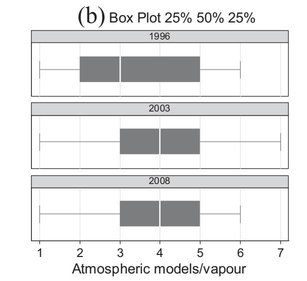
**Fig. 2 Perceptions of the ability of atmospheric models to deal with radiation.**

Fig. 3 looks at the perceptions of the ability of climate models to deal with vapour. Concerning the assessment of the ability of climate models to deal with vapour, both the means and the box plots indicate a marked improvement between 1996 and 2003, however, since 2003 there is no perception of improvement. The means also indicate that the ability to deal with vapour is not perceived of as being as good as the ability to deal with radiation or hydrodynamics and leans more towards the negative end of the scale.

Fig. 4 is the climate scientists' appraisal of the ability of models to deal with clouds, a renowned problem issue for climate modelling.

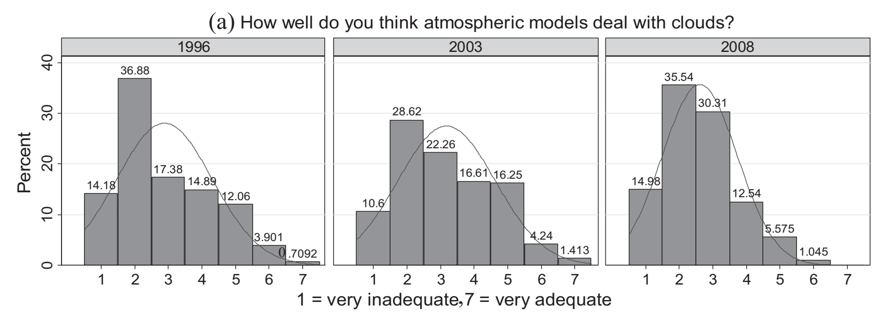



| | | | |
|---|---|---|---|
| n | 280 | 283 | 288 |
| mean | 3.46 | 3.79 | 3.85 |

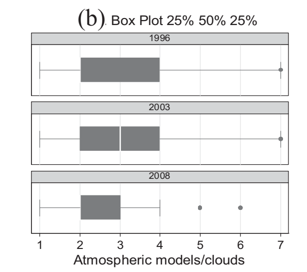



| stat sig dif - means | |
|---|---|
| 1996-2003 | 0.006 |
| 1996-2008 | 0.000 |
| 2003-2008 | - |

**Fig. 3 Perceptions of the ability of atmospheric models to deal with vapour.**




| | | | |
|---|---|---|---|
| n | 282 | 283 | 287 |
| mean | 2.88 | 3.18 | 2.61 |




| stat sig dif - means | |
|---|---|
| 1996-2003 | 0.016 |
| 1996-2008 | 0.012 |
| 2003-2008 | 0.000 |

**Fig. 4 Perceptions of the ability of atmospheric models to deal with clouds.**

The weakness in the ability of climate models to deal with clouds is evident in both the histograms and the box plots. The means indicate that between 1996 and 2003 there was a statistically significant difference between the means, albeit of a small magnitude. However, the perception of the ability to deal with clouds dropped to a new low (when considering the years 1996, 2003, 2008) in 2008. The concentration of dissatisfaction is indicated in the compression and location of the 50th percentile in the box plots and the perception of the lack of confidence in climate models to deal with clouds is very evident in the means. One should note however, that this set of means differs little from the means in the assessment of the ability of models to deal with vapour.

Fig. 5 addresses issues with climate models and their ability to deal with precipitation. As with clouds, the means indicate that the perception among climate modellers is that climate models to not have a very high ability to deal with precipitation. The statistically significant difference between the means of 2003 and 2008 seems to indicate that in 2008 climate scientists perceived the ability of climate models to deal with precipitation to be worse than the previous years in which similar measures were made. A mean of only 2.86 suggests a very high level of dissatisfaction with the ability of the models. The box plots indicate that although the 50th percentiles of the samples seems to have remained stationary over the period of years (at a level of low assessment), by 2008 even the tail representing the upper percentile has declined in the assessment of the models.

Fig. 6, the last variable directly assessing the ability of models, deals with the perceptions of climate modellers regarding the ability of atmospheric models to deal with atmospheric convection.

Again, the mean indicates well below acceptable levels, and between 2003 and 2008 climate modellers have lowered their perception of the ability of climate models to deal with atmospheric convection. There is also a notable shift evident in the box plots.

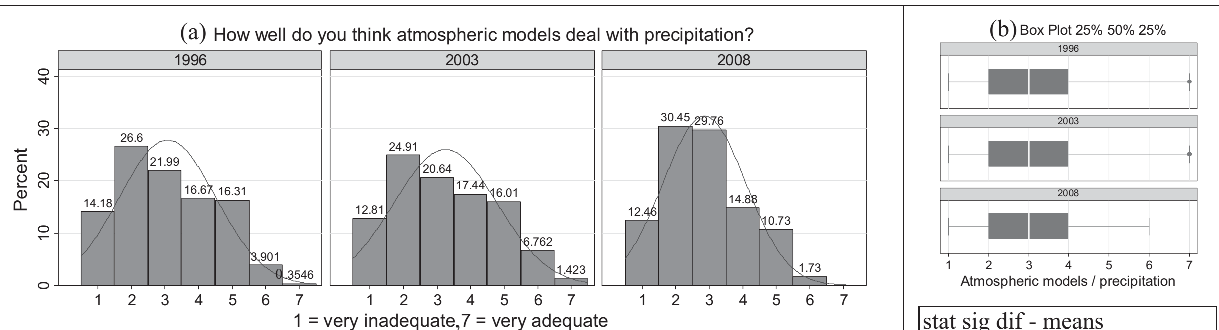
## 4. Processes

The next set of figures deal with climate modellers' perceptions of the state of knowledge pertaining to the effects of turbulence, albedo, land surface processes and green house gases from anthropogenic sources.

Fig. 7 represents the level of agreement from climate scientists to the statement "The current state of knowledge is developed well enough to allow for a reasonable assessment of the effects of turbulence".

With an overall low assessment, the data indicates that by 2008, the assessment of the knowledge pertaining to the effects of turbulence is lower than in 1996. This is perhaps a realization of the complexity of the issue. The box plots indicate a similar shift.

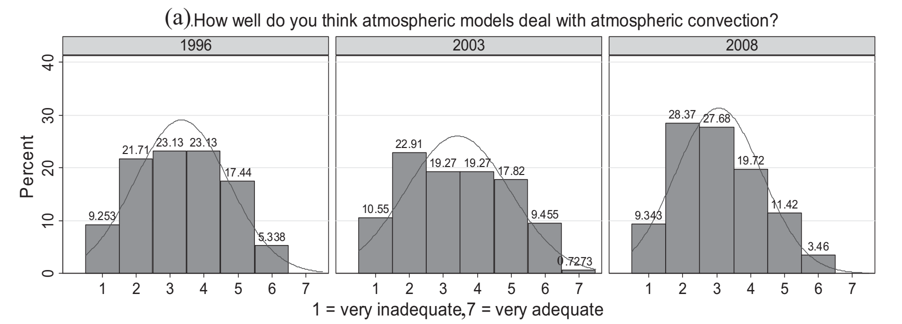
Fig. 8 concerns the assessment of the effects of albedo. The mean of 1996 is very close to the mean of

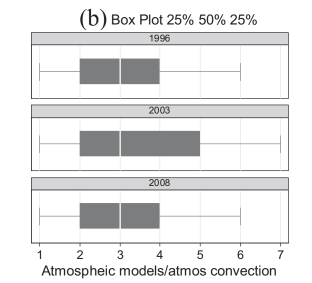



| | 1996 | 2003 | 2008 |
|---|---|---|---|
| n | 282 | 281 | 289 |
| mean | 3.07 | 3.25 | 2.86 |

| stat sig dif - means | |
|---|---|
| 1996-2003 | - |
| 1996-2008 | - |
| 2003-2008 | 0.001 |

**Fig. 5 Perceptions of the ability of atmospheric models to deal with precipitation.**

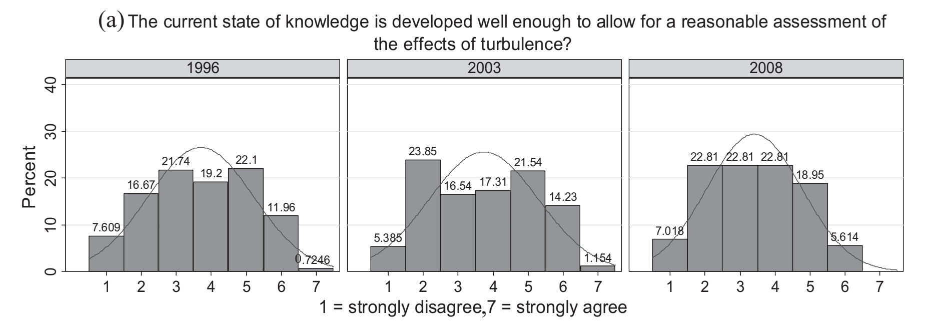


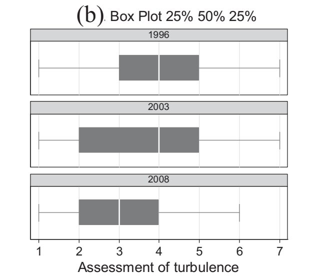




| | | | |
|---|---|---|---|
| n | 281 | 275 | 289 |
| mean | 3.34 | 3.42 | 3.06 |

| stat sig dif - means | |
|---|---|
| 1996-2003 | - |
| 1996-2008 | 0.012 |
| 2003-2008 | 0.002 |

**Fig. 6 Perceptions of the ability of atmospheric models to deal with atmospheric convection.**

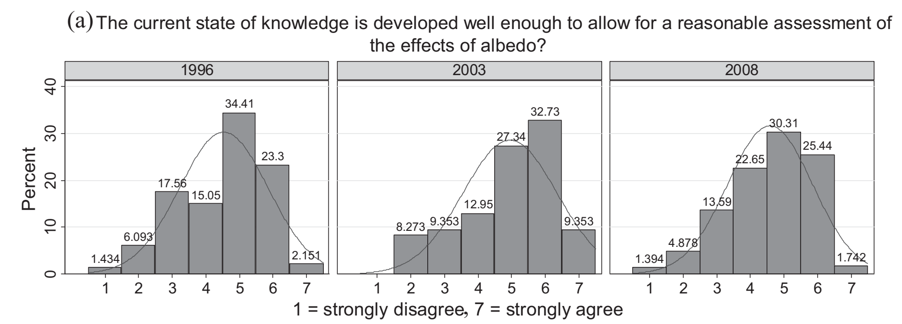


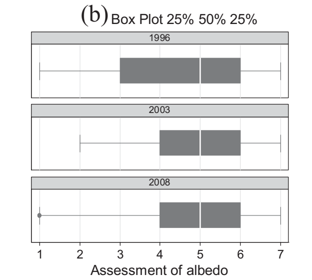




| | | | |
|---|---|---|---|
| n | 276 | 260 | 285 |
| mean | 3.70 | 3.73 | 3.41 |

| stat sig dif - means | |
|---|---|
| 1996-2003 | - |
| 1996-2008 | 0.015 |
| 2003-2008 | 0.010 |

**Fig. 7 Assessment of the knowledge pertaining to the effects of turbulence.**




| | | | |
|---|---|---|---|
| n | 279 | 278 | 287 |
| mean | 4.53 | 4.95 | 4.59 |

| stat sig dif – means | |
|---|---|
| 1996-2003 | 0.000 |
| 1996-2008 | - |
| 2003-2008 | 0.001 |

**Fig. 8 Assessment of the knowledge pertaining to the effects of albedo.**

2008, suggesting little in the way of advancement, and with the mean value of approximately 4, also suggesting much room for improvement. However, the understanding of the effects of albedo is reported to be considerably higher than some of the other variables discussed to this point.

Fig. 9, assessing the state of the knowledge pertaining to the effects of land surface processes, shows a consistently low assessment over the years with a statistically significant slight improvement between 1996 and 2003. The box plots indicate an all but unchanging assessment across the years.

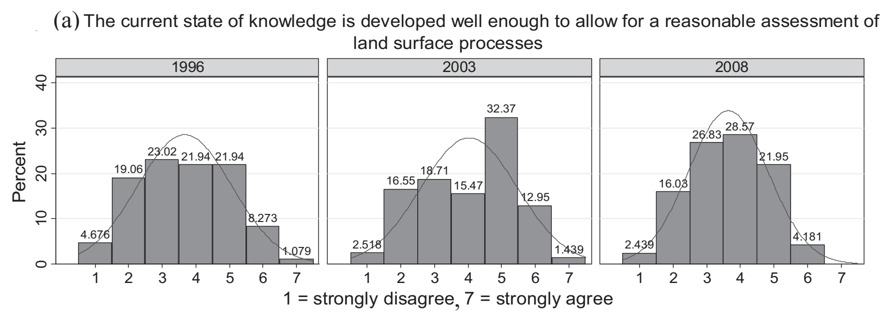
Fig. 10, "Assessment of the Knowledge Pertaining to the Effects of Green House Gases from Anthropogenic Sources", addresses the crux of much of the contemporary debate concerning global warming.

The data in Fig. 10 suggests that over the years, the only statistically significant change in the assessment has been between 1996 and 2003, with 2003 being a slightly higher mean. The box plots indicate no change in the location of the 50th percentile over the years in question. The elevation of the arguments put forth substantiating androgenic induce climate change do not seem to coincide with the evolution of knowledge.

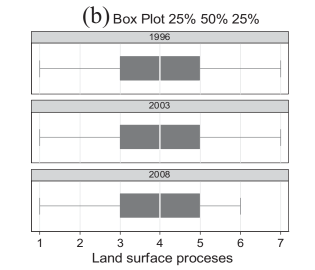
## 5. Projections

This section of the paper looks at the basements made in the surveys concerning the "capacity of the global climate models used elsewhere in this report for projecting future climate change" as per "Climate Models and Their Evaluation" Ch.8. in the IPCC report 2007. Unfortunately, the question employed in the 996 and the 2003 surveys do not coincide directly with the questions posed in the 2008 survey. In 1996 and 2003 one single measure was employed regarding the ability of climate models to make projections of future conditions: respondents were presented with the statement "Climate models can accurately predict climatic conditions of the future' and asked to base an answer on a 7 point scale ranging from strongly agree to strongly disagree".

Some readers of this article will no doubt comment on the use of the word "predict". In 1996 the IPCC had not devoted much effort in distinguishing between the words prediction and projection, and to date some climate scientists still do not make a distinction [5]. In 2008, however, the use of the term in climate science, after considerable discussion, had been clarified

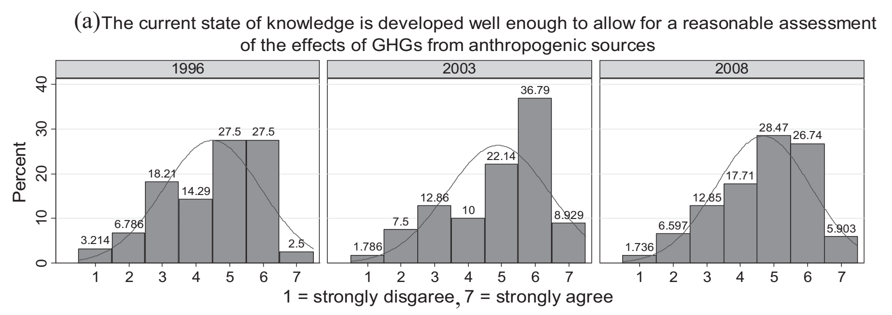



| n | 278 | 278 | 287 |
|---|---|---|---|
| mean | 3.67 | 4.03 | 3.64 |

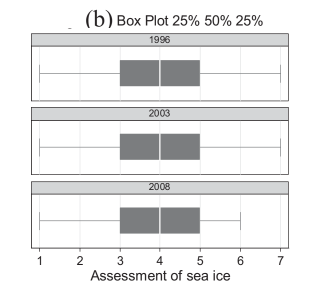



| stat sig dif - means | |
|---|---|
| 1996-2003 | 0.002 |
| 1996-2008 | - |
| 2003-2008 | 0.000 |

**Fig. 9 Assessment of the knowledge pertaining to the effects of land surface processes.**




| n | 280 | 280 | 288 |
|---|---|---|---|
| mean | 4.49 | 4.89 | 4.68 |



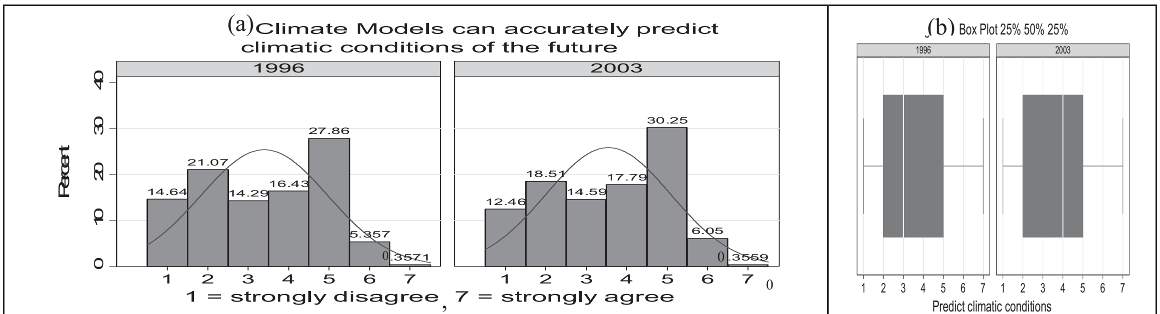

| stat sig dif - means | |
|---|---|
| 1996-2003 | 0.001 |
| 1996-2008 | - |
| 2003-2008 | - |

**Fig. 10 Assessment of the knowledge pertaining to the effects of green house gases from anthropogenic sources.**

by the IPCC. Consequently in the 2008 survey, climate scientists were asked to assess the ability to make “projections” and the question was afforded more detail. In 2008 climate scientists were asked: “How would you rate the edibility of global climate models to: (1) model temperature values for the next 10 years; (2) model temperature values for the next 50 years; (3) model precipitation values for the next 10 years; (4) for the next 50 years; (5) model extreme events for the next 10 years; (6) for the next 50 years”.

Fig. 11 shows the assessment of climate model abilities to predict climatic conditions of the future as made by climate scientists in 1996 and 2003. Neither the shape of the distributions, the means nor the box plots for 1996 and 2003 show much in the way of differences. Overall, there is not a significant amount of faith in the ability of climate models to predict the climatic conditions of the future. Only the median has changed slightly towards a more positive estimate in 2003.

Fig. 12 presents more detailed data as collected in 2008. The data shows, as would be expected, model




| | 1996 | 2003 |
|---|---|---|
| n | 280 | 281 |
| mean | 3.39 | 3.54 |

| stat sig dif - means | |
|---|---|
| 1996-2003 | - |

**Fig. 11 The perceptions of the predictive abilities of models: 1996-2003.**

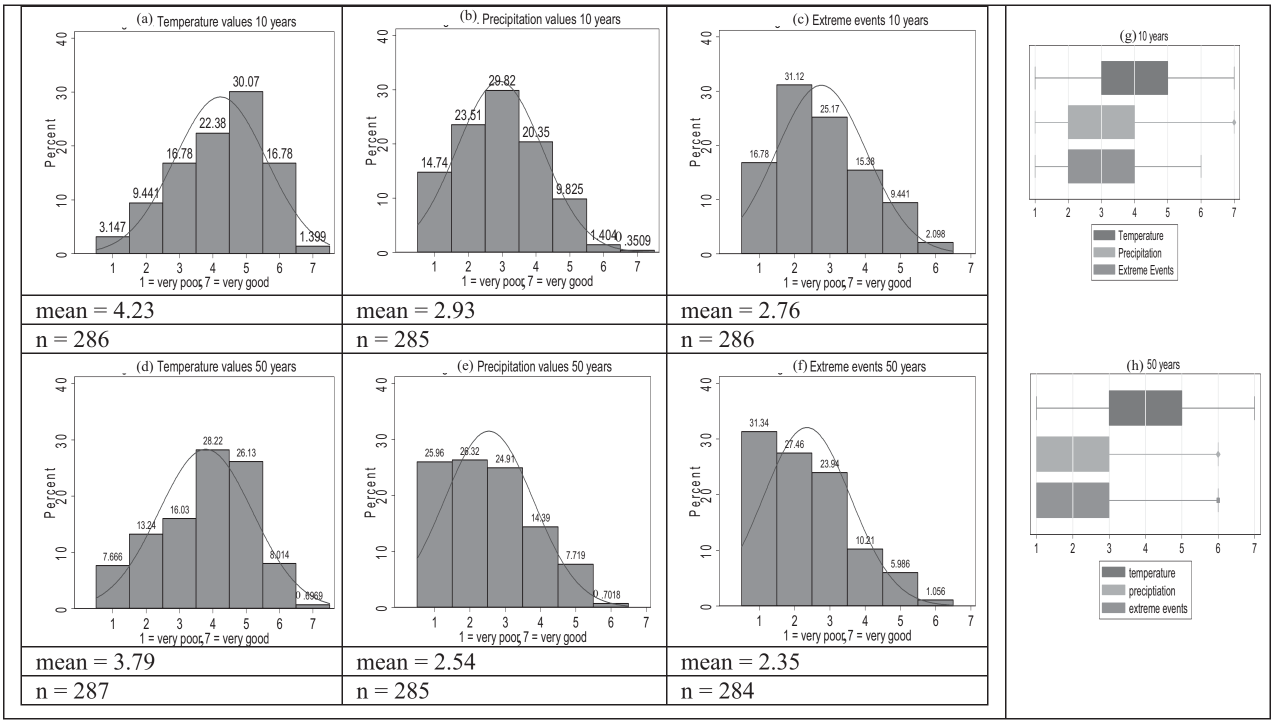



**Fig. 12 The perceptions of the abilities of global climate models 2008.**

abilities to estimate future values is assessed as being better for the shorter period of time with the estimation of temperature values for the next 10 years being evaluated the highest. Also the ranking of skill, with temperature highest, precipitation middle and extreme events lowest, is as expected from the assessment of the components of the models. However, overall, the assessments of global climate models to model future climate are not estimated to be very good.

## 6. Conclusions

The three surveys among physical climate scientists do not contest the product of the two expert reports, but they broaden the perspective of the assessment. The analyses of the two expert groups, which are specific to the attributes of climate models, their assessment, both in method and in content, are very different in scope and method from the empirical analysis of a larger group of non-experts (even if physical climate scientists). Where all coincide is perhaps in the general conclusion suggesting the necessity for much more work and room for significant improvement. However, the conclusions reached by the two expert groups also differ significantly from the conclusion reached surveying a much larger sample of scientists involved in climate modelling: The experts assign a much higher perception of credibility to the abilities of climate models, both in terms of advancement and accuracy. The reasons for this difference are unknown, and speculating about them at this time may not be helpful given the difficult discussion in the present politicized atmosphere.

## References

[1] D.A. Randall, R.A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, et al., Climate models and their evaluation, in: S.D. Qin, M. Manning, Z. Chen, K.B. Averyt, M. Tignor, H.L. Miller (Eds.), Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007, p. 11.

[2] D.C. Bader, C. Covey, W.J. Gutowski Jr., I.M. Held, K.E. Kunkel, R.L. Miller, et al., CCSP, 2008: Climate Models: An assessment of strengths and limitations, A Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research, Department of Energy, Office of Biological and Environmental Research, Washington, D.C., USA, 2008, p. 124.

[3] D. Bray, H. von Storch, The Perspectives of Climate Scientists on Global Climate Change, GKSS 2007/11, GKSS, Geesthacht, Germany, 2007, pp. 1-120.

[4] D. Bray, H. von Storch, CliSci 2008: A Survey of the Perspectives of Climate Scientists Concerning Climate Science and Climate Change, GKSS 2010/9, GKSS, Geesthacht, Germany, 2010, pp. 1-121.

[5] D. Bray, Consensus among climate scientists revisited, Environmental Science and Policy 13 (2010) 340-350.

[6] D. Bray, H. von Storch, ‘Prediction’ or ‘Projection’? The nomenclature of climate science, Science Communications 30 (2009) 534-543.

[7] N. Oreskes, Beyond the ivory tower: The scientific consensus on climate change, Science 306 (2004) 1686.