Kundzewicz's comments on the paper "Credibility of climate predictions revisited" by Anagnostopoulos et al.

Making decision on this paper is not easy. However, after lengthy considerations and discussions I wish to propose a solution.

Recently, I received an interesting personal letter from Stephen Schneider (founder and sole long-term editor of *Climatic Change*), who wrote "When I get a paper that generates controversy and split reviewer advice, I look to be sure that it is mostly differing philosophy rather than technical errors that underlie the dispute." In case of a paradigmatic dispute, Steve tends to accept a paper to generate debate but publishes it with a "springboard" (invited editorial) paper providing a discussion of the topic (rather than being a public criticism of the paper in question). I think that this is quite a good idea.

The paper by Anagnostopoulos et al. and the so far review process unveiled clear communication gaps between climate and water communities and between climate-mainstream experts and climate skeptics. Referees from different camps look differently at the same picture. Results of the review largely depend on the camp and may span a broad range, from "excellent/accept" to "poor/reject". Actually, there is nothing inbetween – either "excellent" or "poor".

Climatologists and hydrologists speak different technical languages. A climatological referee mentions first and second kinds of climatic system predictability that belong to a standard toolbox in the climate community, being largely unknown in the water camp. Technicalities of the Hurst effect and Thiessen polygons (standard menu in hydrological sciences) do not ring the bell in the climate community. There are different disciplinary perspectives on validation / various standards / traditions in disciplines; different attitudes to models vs observations, point vs space?

Authors approach a central issue of considerable controversy. It is fuel to climate change skeptics (climatologists and hydrologists alike), but creating anger among the climate-mainstream camp.

Why not opening a broader debate. Wouldn't it be possible to talk to each other and to try to bring the positions of the camps closer to each other, to specify a list of agreements and disagreements.

Undoubtedly, everyone agrees that GCM model results are directly of no use in hydrological modelling. It is fair to indicate the abuse of GCM results to drive hydrological models. One needs "bias correction" and "downscaling" before entering the realm of hydrological sciences. But aren't these just tricks, curve-fitting exercises, to make a curve fit better to one given set of data? This does not have to hold for a new set of data, hence projections are problematic.

I hope that the authors change a tabloid-type rhetoric towards an objective tone. Do they really have to write "climate change ... has been being taught in schools"? Even disregarding the linguistic problem, it is a real exaggeration. I know of no school with climate change on the curriculum. Teachers of geography or environment may pick up the theme, because there is much interest. Moreover, teachers feel the warming – winters in the old days were much colder than now. This is obvious in Central Europe, for instance.

Authors condemn climate models. Why? They may be poor, hence in need for improvement, but do we have anything better in the no-analogy situaton? Models express the laws of physics. The climate is governed primarily by the Sun, the Earth's orbit and the composition of the atmosphere. Hence the physical properties of the roof of the planetary greenhouse plays a significant role.

I suggest that this paper is published, together with a sprinboard discussion paper. I would not set on the idea of discussions. They may or may not come, but even if a discussion comes, it is temporally disconnected from the main paper published much earlier.

IAHS Press
Centre for Ecology and Hydrology
Wallingford
Oxfordshire OX10 8BB, UK

web: www.iahs.info
tel: +44 1491 692405
fax: +44 1491 692448/692424
e-mail: frances@iahs.demon.co.uk

Disseminating the results of hydrological research and practice worldwide

HYDROLOGICAL SCIENCES *JOURNAL* DES SCIENCES HYDROLOGIQUES

Editors: *Professor Z. W. Kundzewicz Professor D. Koutsoyiannis*

Mr A. Christofides
Dept of Water Resources,
Faculty of Civil Engineering,
National Technical University of Athens,
Heroon Polytechneiou 5,
GR 157 80 Zographou,
Greece

30 November 2009

Dear Mr Christofides

Paper title: Credibility of climate predictions revisited
HSJ MS no: 3318
(please quote this number in correspondence)

Thank you for submitting the above paper to *Hydrological Sciences Journal*.

It has not been easy to make a decision, in view of the conflicting reviews. Attached are three formal reviews and one statement. Please, consider this material and prepare a final draft of the paper. It is planned that your paper is published in HSJ, together with a springboard/discussion paper, as explained in the SpringboardDk file. Please, try to take the reviews onboard and react to recommendations as far as you can.

Yours sincerely,

Professor Z. W. Kundzewicz

Cc.
Co-Editor
Production Editor
Referees

Review for the Hydrological Sciences Journal

"Credibility of climate predictions revisited" by Anagnostopoulos, D., Koutsoyiannis, A., Christofides, A., Efstradiadis, A. and Mamassis, N. (#HSJ3318)

This manuscript purports to critically assess the consistency between climate model simulations and historic temperature and precipitation observations for a small sample of stations that are distributed unevenly globally. A second analysis is performed using a denser network of stations covering the United States. Although testing of climate model skill is always to be welcomed the present analysis is not sufficiently rigorous to support the headline claim that "local model projections are poor" (beyond what is already known). The following aspects of the study are of particular concern:

1. The distribution of stations employed in the global analysis is highly biased with major gaps evident across Africa, Asia and South America. The station selection appears to have been determined by convenience of access to (online) data. Other well known global (e.g., Legates and Willmott, 1990; New et al., 2002) or North American (e.g., Maurer et al., 2002) would have provided better spatial and temporal coverage. Note also that the analysis covers only land areas.

2. It is unclear whether the climate model simulations employed in the study reflect all known historical forcings. Reference to SRES A2 and IS92a implies that only anthropogenic components were incorporated.

3. Much more detail is needed on the best linear unbiased estimation (BLUE) technique, and the purpose of the Hurst-Kolmogorov coefficient should have been explained.

4. Unless the meteorological station data are transformed to conform with the grid resolution of the climate models there is always a danger of comparing apples with oranges. This is further justification for the use of one of the gridded data sets mentioned in #1. Direct comparison between GCM output for individual (or a few) grid points and the temperature trend at a single meteorological station (such as Durban) is meaningless because of the scale mismatch, and the fact the GCM cannot resolve sub-grid processes such as land cover changes, local topographic influences, etc.

5. Comparing changes between two periods of climate model simulation is fraught with uncertainty due to sampling natural variability (see Kendon et al., 2008). Furthermore, any correlation between time-series of observations and climate model output is meaningless unless a range of initial conditions have been properly sampled.

6. The premise that "climate models have been eluding verification" is unfounded. The IPCC Fourth Assessment Report devoted considerable attention to climate model verification using a diverse set of metrics above and beyond monthly temperature and precipitation indices (see Randall et al., 2007). In any event, the observation that climate models do not represent regional climate / sub-grid variability is widely known (e.g., Osborn and Hulme, 1997) and has long been the rationale for downscaling techniques (Wigley et al., 1990). Likewise, the sweeping assertion that hydrologists and water managers use GCM output uncritically is incorrect (e.g., Xu, 1999).

7. The GCM evaluation is based on a very limited sample compared with the much larger multi-model and perturbed-physics experiments that are now routinely employed (e.g., Murphy et al., 2008). Furthermore, two pairs of GCMs originate from the same climate modelling centre, further reducing the size of the sample.

8. There are a number of presentational issues. For example, the Figure numbering is not sequential from #9 onwards; temperature changes should not be expressed as percentages (Figure 13); the selected GCM outputs are not "predictions" but scenarios.

On this basis of the above, publication is not recommended.

Supporting materials

Kendon, E.J., Rowell, D.P., Jones, R.G. and Buonomo, E. 2008. Robustness of future changes in local precipitation extremes. *Journal of Climate*, **21**, 4280-4297.

Legates, D.R. and Willmott, C.J. 1990. Mean seasonal and spatial variability in gauge-corrected, global precipitation. *International Journal of Climatology*, **10**, 111-127.

Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P. and Njissen, B. 2002. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *Journal of Climate*, **15**, 3237-3251.

Murphy, J.M., Booth, B.B.B., Collins, M., Harris, G.R., Sexton, D.M.H. and Webb, M.J. 2007. A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philosophical Transactions of the Royal Society A*, **365**, 1993-2028.

New, M. G., Lister, D., Hulme, M. and Makin, I. 2002. A high-resolution data set of surface climate for terrestrial land areas. *Climate Research*, **21**, 1-25.

Osborn, T.J. and Hulme, M. 1997. Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *Journal of Climate*, **10**, 1885-1908.

Randall, D.A., R.A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R.J. Stouffer, A. Sumi and K.E. Taylor, 2007: Climate Models and Their Evaluation. In: *Climate Change 2007: The Physical Science Basis*. Contribution of

Wigley, T.M.L., Jones, P.D., Briffa, K.R. and Smith, G. 1990. Obtaining sub-grid-scale information from coarse resolution General Circulation Model output. *Journal of Geophysical Research*, **95**, 1943-1953.

Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Xu, C.Y. 1999. From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Progress in Physical Geography*, **23**, 229-249.

IAHS Press
Centre for Ecology and Hydrology
Wallingford
Oxfordshire OX10 8BB, UK

web: www.iahs.info
tel: +44 1491 692405
fax: +44 1491 692448/692424
e-mail: frances@iahs.demon.co.uk

Disseminating the results of hydrological research and practice worldwide

HYDROLOGICAL SCIENCES *JOURNAL* DES SCIENCES HYDROLOGIQUES

Editors: *Professor Z. W. Kundzewicz Professor D. Koutsoyiannis*

REFEREE'S REPORT: *Hydrological Sciences Journal* MS no. 3318

Paper title: Credibility of climate predictions revisited

Authors: Anagnostopoulos et al., Greece

Referee name:

Please note that the contents of the manuscript remain confidential until published. Reviews are anonymous unless reviewers wish their names to be made known to the author(s). Would you like your name to be revealed to the author(s)?
NO

Aggregate assessment – How do you rate this paper in absolute terms?

Poor to fair	Good	Very good to excellent
		XXXXXX

Is the subject of the article	No	Possibly	Yes	Comments
Within the scope of the *Journal*?			XXX	

Please summarize, in one or two sentences, the main contribution and novelty, if any, of this paper:

Is the paper a new, original and valuable contribution to hydrological theory, methodology, modelling, education, etc?			XXX	
Is the paper a new, original, and valuable contribution to factual information about the hydrology of a particular region?			XXX	
If the reply to either of the above questions is positive, is the paper of sufficiently wide interest to merit publication in an international journal?			XXX	

Is the paper technically sound and free of errors of fact or logic?			XXX	
Are the objectives clear? Is the material clearly presented?			XXX	
Is the methodology appropriate?				
Are the assumptions and the analysis valid and adequately justified?			XXX	
Are the interpretations and conclusions sound and justified by the data?			XXX	
Are the data of appropriate quality?			XXX	

	No	Possibly	Yes	Comments
Is the quality of the language satisfactory?			XXX	
Does the title of this paper clearly and sufficiently reflect its contents?			XXX	
Are the references adequate, up-to-date, and relevant?	XXX			See below
Are the approach, results and conclusions intelligible from the abstract alone?			XXX	
Are the key words informative, appropriate and complete?	XXX			I do not find any key words
Are the illustrations of adequate quality, legible and understandable?			XXX	

Could the paper be shortened without detriment to the material presented in it (e.g. by removal of poor, irrelevant, excessive, or redundant material)? ***NO***

Please indicate such material in the manuscript. Are all illustrations and/or tables necessary? ***YES***

If not, could some of them be removed? Alternatively, could the information in the paper be more clearly or concisely conveyed by the use of tables or figures?

Please add any other specific comments you may have—if necessary, continuing on a separate sheet (sheets). Since the authors are requested to indicate on their revised papers where the reviewers' comments have been taken into account, it would help if you number any comments you may have.

I only have a few comments with respect to what is a very much needed assessment of the skill of multi-decadal global model predictions, and is a valuable follow up paper to their earlier model/observational paper which had less sites. My comments are:

1. **The author uses statistical terminology that needs to be defined within the paper in order to make the results clearer to readers who are not specialists on this topic. This includes briefly defining the terms such as the Hurst coefficient; the Thiessen coefficient; the BLUE technique, and Hurst-Kolmogorov behaviour.**
2. **The last part of the sentence at the bottom of page 8 is unclear; "…is less biased in favour of the models". What does this mean? This needs to be rewritten.**
3. **The authors leave out references that support their conclusions of the difficulty of skilfully predicting the climate system on yearly and decadal time scales. These include, as just two examples,**

   Rial, J., R.A. Pielke Sr., M. Beniston, M. Claussen, J. Canadell, P. Cox, H. Held, N. de Noblet-Ducoudre, R. Prinn, J. Reynolds, and J.D. Salas, 2004: Nonlinearities, feedbacks and critical thresholds within the Earth's climate system. Climatic Change, 65, 11-38

   National Research Council, 2005: Radiative forcing of climate change: Expanding the concept and addressing uncertainties. **Committee on Radiative Forcing Effects on Climate Change, Climate Research Committee, Board on Atmospheric Sciences and Climate, Division on Earth and Life Studies, The National Academies Press, Washington, D.C., 208 pp.**

   The paper would be even stronger if literature was referenced which highlights the complexity of the climate system.

Finally, this paper is a very important scientific contribution and needs to be widely read. Undoubtedly there will be those who do not even want to see such evaluations of the climate models appear in the literature. However, the publication of this paper will provide the data and statistical analyses which they can use to seek to refute the findings of this paper, if they can.

Overall evaluation – The paper should be:

Accepted as it stands, apart from editorial changes.	
Accepted after minor revision.	**XXX**
Subject to major revision. If revised paper is re-submitted, it needs to be reconsidered and re-reviewed.	
Rejected outright.	

The paper should be sent to another referee before terminating the review process (e.g. in the case of potentially contentious elements). If possible, please suggest the name (and e-mail) of a reviewer.	
If you have recommended major revision and re-submission, would you be willing to review the revised manuscript?	
Would you be willing to edit the language, should this paper be accepted?	

IAHS AISH

IAHS Press
Centre for Ecology and Hydrology
Wallingford
Oxfordshire OX10 8BB, UK

web: www.iahs.info
tel: +44 1491 692405
fax: +44 1491 692448/692424
e-mail: frances@iahs.demon.co.uk

Disseminating the results of hydrological research and practice worldwide

HYDROLOGICAL SCIENCES *JOURNAL* DES SCIENCES HYDROLOGIQUES

Editors: *Professor Z. W. Kundzewicz* *Professor D. Koutsoyiannis*

REFEREE'S REPORT: *Hydrological Sciences Journal* MS no. 3318

Paper title: Credibility of climate predictions revisited

Authors: Anagnostopoulos et al., Greece

Referee name:

Please note that the contents of the manuscript remain confidential until published. Reviews are anonymous unless reviewers wish their names to be made known to the author(s). Would you like your name to be revealed to the author(s)?

Aggregate assessment – How do you rate this paper in absolute terms?

Poor to fair	Good	Very good to excellent
X		

Is the subject of the article	No	Possibly	Yes	Comments
Within the scope of the *Journal*?			X	

Please summarize, in one or two sentences, the main contribution and novelty, if any, of this paper:

Is the paper a new, original and valuable contribution to hydrological theory, methodology, modelling, education, etc?	X			This paper is misleading as it is based on a wrong assumption related to the climate system predictability.
Is the paper a new, original, and valuable contribution to factual information about the hydrology of a particular region?	X			See the above
If the reply to either of the above questions is positive, is the paper of sufficiently wide interest to merit publication in an international journal?				N/A

Is the paper technically sound and free of errors of fact or logic?	X			General statements that the reliability of climate model predictions "is typically not assessed" and that "climate models have been eluding verification" are not true. (NB: terminologically it is cleaner to say that model simulations are verified, while the models themselves are evaluated.) A very significant amount of studies (e.g. under Coupled Model Intercomparison Project), to say nothing about the IPCC regular effort on model

				evaluation are thus negated. The statement that increasing confidence in the role of anthropogenic influence on climate is based solely on models negates other important aspects of the attribution part of the state-of-the-art climate change science. Compared to these major problems with interpretation of facts, other errors, e.g. mixing (p.3) 21st century scenarios (such as SRES A2) with 20th century simulations (such as 20C3M), are really technical and minor.
Are the objectives clear? Is the material clearly presented?			X	
Is the methodology appropriate?	X			Testing models against observationally based data (i.e. past and present climate) is a key part of model evaluation. However, the direct point-to-point (or model-gridbox-to-station) comparison – a part of the analysis undertaken by the authors – is not appropriate due to spatial scale differences predetermining local biases and hardly characterizing reliability of model simulations at larger scales. It is also well known that the direct use of state-of-the-art AOGCMs' outputs to drive local (e.g. hydrological) models without any downscaling is not appropriate.
Are the assumptions and the analysis valid and adequately justified?	X			The fundamental problem of the paper is that the first and the second kinds of climate system predictability seem to be misunderstood by the authors. Due to unforced (model-generated) variability arising from non-linearity of the climate system, trajectories of simulations with the same AOGCM cannot coincide in the phase space if the initial conditions are different. Therefore, the correlation between model and observation time series is a poor choice for a model simulation verification as the observations represent but a single realization. Comparisons should utilize climate statistics rather than year-to-year values that manifest natural variability. If the authors have a look at the behaviour of just two different ensemble members simulated by any of the models they are evaluating, they would not need to bother themselves with collecting observational data to demonstrate the poor correlation.
Are the interpretations and conclusions sound and justified by the data?	X			Having the wrong assumptions, the conclusion that the continental or global climatic projections are not credible is not supported by the analysis undertaken. On the other hand, it

				should be admitted that as a direct input to hydrological models “to make local predictions” state-of-the-art AOGCMs are indeed of limited usefulness. But this is not a news, and this is why downscaling techniques are used for local scale projections.
Are the data of appropriate quality?		X		The quality of the observational and model data used in this paper do not look inappropriate.

Referee’s report HSJ ***(continued)***

	No	Possibly	Yes	Comments
Is the quality of the language satisfactory?			X	
Does the title of this paper clearly and sufficiently reflect its contents?			X	
Are the references adequate, up-to-date, and relevant?	X			The relevance of some references is not evident (e.g. Collins, 2002; Kolmogorov,1940; Stroeve et al., 2007)
Are the approach, results and conclusions intelligible from the abstract alone?		X		
Are the key words informative, appropriate and complete?	X			N/A
Are the illustrations of adequate quality, legible and understandable?			X	

Could the paper be shortened without detriment to the material presented in it (e.g. by removal of poor, irrelevant, excessive, or redundant material)? Please indicate such material in the manuscript. Are all illustrations and/or tables necessary? If not, could some of them be removed? Alternatively, could the information in the paper be more clearly or concisely conveyed by the use of tables or figures?

Please add any other specific comments you may have—if necessary, continuing on a separate sheet (sheets). Since the authors are requested to indicate on their revised papers where the reviewers’ comments have been taken into account, it would help if you number any comments you may have.

Overall evaluation – The paper should be:

Accepted as it stands, apart from editorial changes.	
Accepted after minor revision.	
Subject to major revision. If revised paper is re-submitted, it needs to be reconsidered and re-reviewed.	
Rejected outright.	X

The paper should be sent to another referee before terminating the review process (e.g. in the case of potentially contentious elements). If possible, please suggest the name (and e-mail) of a reviewer.	Ron Stouffer rjs@gfdl.noaa.gov
If you have recommended major revision and re-submission, would you be willing to review the revised manuscript?	NA
Would you be willing to edit the language, should this paper be accepted?	No

Authors' response to editor and reviewer comments on "A comparison of local and aggregated climate model outputs with observed data" (former "Credibility of climate predictions revisited")

March 2010

Editor

"I hope that the authors change a tabloid-type rhetoric towards an objective tone. Do they really have to write "climate change ... has been being taught in schools"? Even disregarding the linguistic problem, it is a real exaggeration. I know of no school with climate change on the curriculum. Teachers of geography or environment may pick up the theme, because there is much interest. Moreover, teachers feel the warming – winters in the old days were much colder than now. This is obvious in Central Europe, for instance."

We have removed the entire tabloid-type first paragraph, and a large part of the abstract. We have replaced the provocative title of the paper to a neutral one. We have also made numerous related small changes throughout the paper. We think that the result is a really cold, objective-toned text, and we agree that it is now much better.

"Authors condemn climate models. Why? They may be poor, hence in need for improvement, but do we have anything better in the no-analogy situaton? Models express the laws of physics. The climate is governed primarily by the Sun, the Earth's orbit and the composition of the atmosphere. Hence the physical properties of the roof of the planetary greenhouse plays a significant role."

Since we have removed all such rhetoric, there is now no condemnation of climate models in the paper; only the presentation of our method, its justification, and the results. In the two concluding paragraphs of the paper, we now also attempt to reply whether we have anything better.

Reviewer A

"1. The distribution of stations employed in the global analysis is highly biased with major gaps evident across Africa, Asia and South America. The station selection appears to have been determined by convenience of access to (online) data. Other well known global (e.g., Legates and Willmott, 1990; New et al., 2002) or North American (e.g., Maurer et al., 2002) would have provided better spatial and temporal coverage. Note also that the analysis covers only land areas. "

This is correct but it does not undermine the results of the study. We find that GCM projections are not good at the stations we selected. We think that this conclusion is valuable. In addition, the consistency of results across the stations makes it unlikely that results at Africa, Asia, South America, or over the sea, would be dramatically different. Indeed, the station selection has been determined by convenience of access to data. There is no reason why this would bias the sample against GCM outputs; on the contrary, it could bias it in favour, because climate modelers could calibrate their models so that they match known data.

"2. It is unclear whether the climate model simulations employed in the study reflect all known historical forcings. Reference to SRES A2 and IS92a implies that only anthropogenic components were incorporated. "

TAR model runs use historical forcings up to 1989, and only extend using scenarios from 1990 and beyond. Therefore, for periods up to 1989, choice of scenario does not matter, whereas for later periods there is no significant difference between different scenarios for the same model. For AR4 models, we used 20C3M. This was not entirely clear in the first draft; we have now explained it better (second paragraph of section "Methodology and data").

"3. Much more detail is needed on the best linear unbiased estimation (BLUE) technique,

and the purpose of the Hurst-Kolmogorov coefficient should have been explained. "

We provided detail on the BLUE technique and on the Hurst coefficient and the Hurst-Kolmogorov behaviour (third and fourth paragraphs of section "Methodology and data").

"4. Unless the meteorological station data are transformed to conform with the grid resolution of the climate models there is always a danger of comparing apples with oranges. This is further justification for the use of one of the gridded data sets mentioned in #1. Direct comparison between GCM output for individual (or a few) grid points and the temperature trend at a single meteorological station (such as Durban) is meaningless because of the scale mismatch, and the fact the GCM cannot resolve sub-grid processes such as land cover changes, local topographic influences, etc. "

This is a common argument, also used by Reviewer C. The preceding paper, Koutsoyiannis et al. (2008), has also been criticised on these terms by various science blogs. However, we think that the argument is incorrect, and we have provided a substantial new section which justifies this (section "Justification of the methodology", subsection "Scale of comparison"). At the experimental level, we treat this argument by making the comparison at a large scale (contiguous USA). At the theoretical level, while daily temperatures can indeed differ significantly at a distance of 200 km (which is comparable to the size of the grid), the maximum temporal resolution we use is monthly; and differences in monthly temperature at points that close will be almost identical. The same applies to precipitation at over-year scales. There could, however, be a systematic bias; for example, a point being consistently 1°C higher than a nearby one or than the grid box average. This bias may show in our comparison, since we make unbiased estimation, but one of the metrics we use is the correlation coefficient, which ignores bias and therefore does not have this problem.

"5. Comparing changes between two periods of climate model simulation is fraught with uncertainty due to sampling natural variability (see Kendon et al., 2008). Furthermore, any correlation between time-series of observations and climate model output is meaningless unless a range of initial conditions have been properly sampled. "

This is another common argument, which we also think incorrect. The previous argument has to do with spatial scale; this has to do with temporal scale. The two are related, so we treat this at the same section as the previous one. At the experimental level, we address this by making comparisons at the climatic scale. At the theoretical level, we explain that, since the climatic scale is derived from the annual scale, it is not possible for the former to be correct when the latter is incorrect; this could hold only under certain circumstances, which do not hold in our case.

"6. The premise that "climate models have been eluding verification" is unfounded. The IPCC Fourth Assessment Report devoted considerable attention to climate model verification using a diverse set of metrics above and beyond monthly temperature and precipitation indices (see Randall et al., 2007). In any event, the observation that climate models do not represent regional climate / sub-grid variability is widely known (e.g., Osborn and Hulme, 1997) and has long been the rationale for downscaling techniques (Wigley et al., 1990). Likewise, the sweeping assertion that hydrologists and water managers use GCM output uncritically is incorrect (e.g., Xu, 1999). "

We have removed the statement that "climate models have been eluding verification", because this is not the main focus of the paper. We have also improved the abstract and we no longer claim that the reliability of GCM outputs is not assessed. Instead, we have added a new section ("Alternative evaluation methods", under "Justification of the methodology") in which we explain that alternative evaluation methods, such as perturbed-physics ensembles and model intercomparison, are no substitute for our method. As to whether regional climate is expected to be represented well by climate models, see above.

"7. The GCM evaluation is based on a very limited sample compared with the much larger multi-model and perturbed-physics experiments that are now routinely employed (e.g.,

Murphy et al., 2008). Furthermore, two pairs of GCMs originate from the same climate modelling centre, further reducing the size of the sample. "

Concerning multi-model and perturbed-physics experiments, see above. Concerning the size of the sample: is the reviewer suggesting that we have been right that these particular GCM outputs are not good estimates of future output, but that other GCM outputs might? If yes, then our research is still valid, since it permits one to distinguish bad GCM outputs from others that might be good.

"8. There are a number of presentational issues. For example, the Figure numbering is not sequential from #9 onwards; temperature changes should not be expressed as percentages (Figure 13); the selected GCM outputs are not "predictions" but scenarios. "

We have changed the numbering of Figures properly so that references are sequential. Former Figure 13 (now Figure 11) is correct; the percentages are not temperature changes; they are differences between modeled changes and observed changes. Whether GCM outputs are predictions, projections, or scenarios, needs some discussion, but since this is not very relevant to the paper we replaced most occurences of "prediction".

Reviewer B

"1. The author uses statistical terminology that needs to be defined within the paper in order to make the results clearer to readers who are not specialists on this topic. This includes briefly defining the terms such as the Hurst coefficient; the Thiessen coefficient; the BLUE technique, and Hurst-Kolmogorov behaviour. "

We provided detail on the BLUE technique and on the Hurst coefficient and the Hurst-Kolmogorov behaviour (third and fourth paragraphs of section "Methodology and data"). We also described what the Thiessen method is about in a short sentence (second paragraph of "Comparison at a large scale" under "Methodology and data").

"2. The last part of the sentence at the bottom of page 8 is unclear; "...is less biased in favour of the models". What does this mean? This needs to be rewritten. "

We replaced the phrase with "... is less forgiving," which should be clearer.

"3. The authors leave out references that support their conclusions of the difficulty of skilfully predicting the climate system on yearly and decadal time scales. These include, as just two examples,

Rial, J., R.A. Pielke Sr., M. Beniston, M. Claussen, J. Canadell, P. Cox, H. Held, N. de Noblet-Ducoudre, R. Prinn, J. Reynolds, and J.D. Salas, 2004: Nonlinearities, feedbacks and critical thresholds within the Earth's climate system. Climatic Change, 65, 11-38

National Research Council, 2005: Radiative forcing of climate change: Expanding the concept and addressing uncertainties. Committee on Radiative Forcing Effects on Climate Change, Climate Research Committee, Board on Atmospheric Sciences and Climate, Division on Earth and Life Studies, The National Academies Press, Washington, D.C., 208 pp.

The paper would be even stronger if literature was referenced which highlights the complexity of the climate system. "

We have added the first of these references and discuss it at the final section of the paper ("Conclusions and discussion").

Reviewer C

"General statements that the reliability of climate model predictions "is typically not assessed" and that "climate models have been eluding verification" are not true. (NB: terminologically it is cleaner to say that model simulations are verified, while the models

themselves are evaluated.) A very significant amount of studies (e.g. under Coupled Model Intercomparison Project), to say nothing about the IPCC regular effort on model evaluation are thus negated. The statement that increasing confidence in the role of anthropogenic influence on climate is based solely on models negates other important aspects of the attribution part of the state-of-the-art climate change science. Compared to these major problems with interpretation of facts, other errors, e.g. mixing (p.3) 21st century scenarios (such as SRES A2) with 20th century simulations (such as 20C3M), are really technical and minor. "

We have removed the statement that "climate models have been eluding verification", because this is not the main focus of the paper. We have also improved the abstract and we no longer claim that the reliability of GCM outputs is not assessed. Instead, we have added a new section ("Alternative evaluation methods", under "Justification of the methodology") in which we explain that alternative evaluation methods, such as perturbed-physics ensembles and model intercomparison, are no substitute for our method.

The statement that increasing confidence in the role of anthropogenic influence on climate is based solely on models has also been removed (all that paragraph has been removed).

We have not mixed 21st century scenarios with 20th century simulations; what we did was indeed not entirely clear in the first draft; we have now explained it better (second paragraph of section "Methodology and data").

"Testing models against observationally based data (i.e. past and present climate) is a key part of model evaluation. However, the direct point-to- point (or model-gridbox-to- station) comparison – a part of the analysis undertaken by the authors – is not appropriate due to spatial scale differences predetermining local biases and hardly characterizing reliability of model simulations at larger scales. It is also well known that the direct use of state-of-the-art AOGCMs' outputs to drive local (e.g. hydrological) models without any downscaling is not appropriate. "

See reply to Reviewer A point 4.

"The fundamental problem of the paper is that the first and the second kinds of climate system predictability seem to be misunderstood by the authors. Due to unforced (model-generated) variability arising from non-linearity of the climate system, trajectories of simulations with the same AOGCM cannot coincide in the phase space if the initial conditions are different. Therefore, the correlation between model and observation time series is a poor choice for a model simulation verification as the observations represent but a single realization. Comparisons should utilize climate statistics rather than year-to-year values that manifest natural variability. If the authors have a look at the behaviour of just two different ensemble members simulated by any of the models they are evaluating, they would not need to bother themselves with collecting observational data to demonstrate the poor correlation. "

See reply to Reviewer A point 5.

"Having the wrong assumptions, the conclusion that the continental or global climatic projections are not credible is not supported by the analysis undertaken. On the other hand, it should be admitted that as a direct input to hydrological models "to make local predictions" state-of-the-art AOGCMs are indeed of limited usefulness. But this is not a news, and this is why downscaling techniques are used for local scale projections. "

See reply to Reviewer A point 4.

"The relevance of some references is not evident (e.g. Collins, 2002; Kolmogorov,1940; Stroeve et al., 2007) "

We agree with this, and we have removed Collins (2002) and Stroeve et al. (2007), and improved the relevance of Kolmogorov (1940).

References

Kendon, E.J., Rowell, D.P., Jones, R.G. and Buonomo, E. 2008. Robustness of future changes in local precipitation extremes. Journal of Climate, 21, 4280-4297.

Koutsoyiannis, D. (2009) A random walk on water, *Hydrol. Earth Syst. Sc. Discussions* **6**, 6611–6658.

Legates, D.R. and Willmott, C.J. 1990. Mean seasonal and spatial variability in gauge-corrected, global precipitation. International Journal of Climatology, 10, 111-127.

Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P. and Njissen, B. 2002. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. Journal of Climate, 15, 3237-3251.

Murphy, J.M., Booth, B.B.B., Collins, M., Harris, G.R., Sexton, D.M.H. and Webb, M.J. 2007. A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. Philosophical Transactions of the Royal Society A, 365, 1993-2028.

New, M. G., Lister, D., Hulme, M. and Makin, I. 2002. A high-resolution data set of surface climate for terrestrial land areas. Climate Research, 21, 1-25.

Osborn, T.J. and Hulme, M. 1997. Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. Journal of Climate, 10, 1885-1908.

Randall, D.A., R.A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R.J. Stouffer, A. Sumi and K.E. Taylor, 2007: Climate Models and Their Evaluation. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Wigley, T.M.L., Jones, P.D., Briffa, K.R. and Smith, G. 1990. Obtaining sub-grid-scale information from coarse resolution General Circulation Model output. Journal of Geophysical Research, 95, 1943-1953.

Xu, C.Y. 1999. From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. Progress in Physical Geography, 23, 229-249.