Available online at www.sciencedirect.com
SciVerse ScienceDirect
International Journal of Forecasting 27 (2011) 1000–1003
international journal of forecasting
www.elsevier.com/locate/ijforecast

Discussion

Commentary on "Validation and forecasting accuracy in models of climate change"

Noel S. Keenlyside

Leibniz Institute of Marine Sciences, IFM-GEOMAR, Germany
Geophysical Institute, University of Bergen, Norway

Climate is continually changing, and over the last century a significant level of warming was observed at a global scale. The scientific consensus is that this is largely due to anthropogenic greenhouse gas emissions (IPCC, 2007). The results of complex climate models have played an important role in reaching this consensus. However, like all models, these suffer from systematic errors (IPCC, 2007). Understanding these errors is important, as it may lead not only to their reduction, but also to a better representation of forecast uncertainty. As Fildes and Kourentzes (FK) realise, the application of forecasting methodologies can be of significant value here; their article is one of the first efforts in this vein, and a very welcome one.

FK begin by introducing climate prediction using state-of-the-art climate models and reviewing their traditional method of validation. Unlike other classes of forecasts, climate prediction has been assessed by the ability, not to predict climate, but rather to reproduce it (as is explained further below). There is, however, a rapidly increasing level of interest in performing near-term (or decadal) forecasts (i.e., predictions of 10–30 years ahead) (Meehl et al., 2009), to which traditional verification methodologies can be applied more easily. FK take on this new challenge, and, in the second part of the study, apply standard methodologies to near-term forecasts from one complex model (Smith et al., 2007).

Two of FK's main conclusions are that the complex climate model does not perform better than simple benchmark statistical models on a decadal timescale, and that structural deficiencies can be identified in the complex model results. While I commend FK for their effort, the robustness of these conclusions is questionable. In addition, there are two other points related to their article which warrant further discussion. First, is the verification of decadal forecasts relevant for longer climate change projections? Second, to what extent do external factors (i.e., greenhouse gas emissions) determine decadal changes in regional climate? I will now expand on these comments, beginning with a brief clarification of some important distinctions between decadal and centennial scale predictions; see Keenlyside and Ba (2010) for a more in depth discussion.

DOI of original article: 10.1016/j.ijforecast.2011.03.008.
E-mail address: noel.keenlyside@gfi.uib.no.

doi:10.1016/j.ijforecast.2011.07.002

Climatic variation results from processes which are external to the Earth's climate system, internal to it, or a combination of both. Anthropogenic caused changes in greenhouse gas concentrations are considered to be an external factor. Internal climate variability arises from interactions within the atmosphere itself, as well as from interactions with other components of the climate system, such as the ocean. These interactions give rise to variability on a wide range of timescales. In particular, pronounced decadal variability exists in the North Atlantic and Pacific oceans, with impacts over land (Latif, Collins, Pohlmann, & Keenlyside, 2006).

Both externally forced and internal variability may be predictable, but this predictability arises from external factors in the former case and the initial conditions in the latter. Long-term climate change projections, such as those which were performed for the fourth IPCC assessment report (IPCC, 2007), are designed to capture the predictability resulting from external factors, which are dominant primarily on centennial and longer timescales. These models are able to reproduce the broad-scale warming over oceans and enhanced warming over land and in high latitudes which have been observed over the last century (see Fig. 1(a) and (b)) (IPCC, 2007). As FK point out, there is only one realisation of climate and the future warming driven by increases in greenhouse gas concentrations due to anthropogenic activity causes remains to be fully observed. This makes the application of standard forecast verification methodologies difficult. In spite of both this and FK's assertions, it is the models' ability to reproduce the climate change and variability observed to date that gives us confidence in their use for informing policy.

At decadal timescales and more regional levels, internal climate variability becomes more important, although the influence of external factors is not negligible (Hawkins & Sutton, 2009). Thus, decadal prediction demands both realistic initial conditions and accounting for external factors. The simulations performed for the IPCC fourth assessment report only account for external factors, and are not designed to capture changes on these timescales. These simulations consistently fail to reproduce the full structure of the observed changes over the last thirty years, with the observed patterns resembling modes of internal climate variability (Fig. 1(c) and (d)) (Keenlyside & Ba, 2010).

One motivation of FK is the assumption that the verification of decadal predictions will provide insights into longer-term climate projections. However, this is not necessarily the case: as has just been explained, different processes act on different timescales. Additional feedback systems, for example those involved in the carbon cycle, may also influence and even accelerate global warming (Cox et al., 2004). Nevertheless, as the climate system is non-linear, it is likely that the information gleaned from decadal (and short time-scale) predictions will be able to help inform climate change prediction and lead to better models (Palmer, Doblas-Reyes, Weisheimer, & Rodwell, 2008).

FK assess the decadal prediction skill of externally forced simulations by comparisons with statistical benchmark models. Based on this assessment, they conclude that the (externally forced) models are not suitable for regional decadal predictions, which highlights the need for initialised predictions (Hawkins & Sutton, 2009). These conclusions are consistent with the consensus discussed above, but this does not mean that predictions which only account for external factors cannot inform policy at a regional level, particularly on centennial timescales. It is a question of the signal-to-noise ratio. If an external factor dominates the internal variability, as is the case for particularly strong or persistent forcing, then external-forced variability can dominate, even at a regional level. However, the model uncertainty is large at these scales (Hawkins & Sutton, 2009).

The FK study primarily aims to introduce methodologies for assessing *initialised* decadal prediction. However, the study has major weaknesses, which partly reflect an intrinsic difficulty: to achieve robust skill assessment using only a very limited number of independent events. The forecast skill of the various benchmarks (FK, Table 2) and the results of the forecast encompassing tests (FK, Table 4) are very sensitive to the period chosen and the forecast horizon. While this may suggest that there is a degree of overfitting, it certainly suggests the existence of quite large confidence intervals on the skill scores. Thus, it is still not clear whether the statistical models are significantly better than the initialised GCM based forecasts. The lack of robustness of the benchmarks also leads to questions as to whether combining the complex model with one of the statistical models would

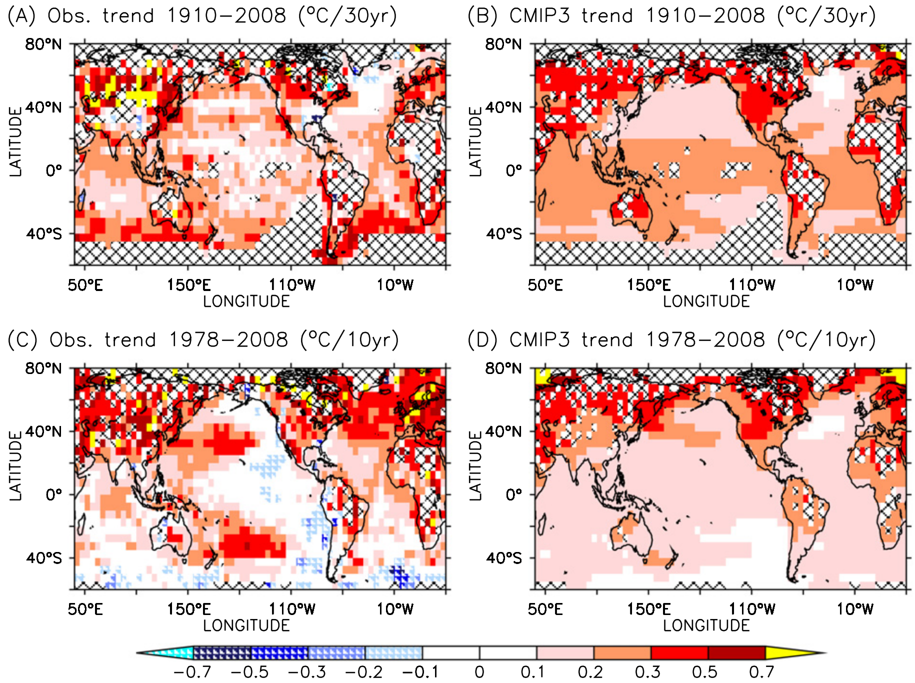

Fig. 1. Centennial and interdecadal surface temperature linear trend patterns, both as observed (Brohan et al., 2006) and from the ensemble mean of 21 climate model simulations which only account for external factors (IPCC, 2007). Regions with insufficient observations (<70%) over the period considered are hatched. The model data are from the Coupled Model Intercomparison Project-3 (CMIP3) database (http://www-pcmdi.llnl.gov/ipcc/aboutipcc.php), and were used for the fourth assessment report of the intergovernmental panel on climate change (IPCC). Note that different units were used for upper and lower panels. (A colour version of this figure is available in the online version of the paper.)

lead to a robust (i.e., out-of-sample) improvement. It may be possible to enhance the robustness of FK's results by adopting pentad or decadal means instead of annual means (i.e., focusing on potentially predictable modes of decadal variability).

The production and verification of decadal predictions is an important and exciting development in climate prediction. The ability to better predict changes over the coming decades and associated uncertainties will increase the confidence in the prediction of climate change, particularly if temporary weakenings of global warming, such as that which occurred between the 1940s and the 1970s, could be foreseen (Keenlyside, Latif, Jungclaus, Kornblueh, & Roeckner, 2008). The upcoming fifth IPCC assessment report will include a limited number of decadal prediction experiments. These experiments should be seen as a pilot study. Quite apart from the forecast verification issues raised by FK, many challenges, both practical and theoretical, still need to be overcome (Meehl et al., 2009). Efforts like those of FK are to be encouraged, as they will contribute to meeting these challenges.

References

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., & Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *Journal of Geophysical Research*, *111*, D12106.

Cox, P. M., Betts, R. A., Collins, M., Harris, P. P., Huntingford, C., & Jones, C. D. (2004). Amazonian forest dieback under climate-carbon cycle projections for the 21st century. *Theoretical and Applied Climatology*, *78*, 137–156.

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*, 1095–1107.

IPCC (2007). *Climate change 2007: the physical science basis.* Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change.

Keenlyside, N. S., & Ba, J. (2010). Prospects for decadal climate prediction. *Wiley Interdisciplinary Reviews: Climate Change*, *1*, 627–635.

Keenlyside, N. S., Latif, M., Jungclaus, J., Kornblueh, L., & Roeckner, E. (2008). Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature*, *453*, 84–88.

Latif, M., Collins, M., Pohlmann, H., & Keenlyside, N. (2006). A review of predictability studies of Atlantic sector climate on decadal time scales. *Journal of Climate*, *19*, 5971–5987.

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., et al. (2009). Decadal prediction. *Bulletin of the American Meteorological Society*, *90*, 1467–1485.

Palmer, T. N., Doblas-Reyes, F. J., Weisheimer, A., & Rodwell, M. J. (2008). Toward seamless prediction: calibration of climate change projections using seasonal forecasts. *Bulletin of the American Meteorological Society*, *89*, 459-–470.

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., & Murphy, J. M. (2007). Improved surface temperature prediction for the coming decade from a global climate model. *Science*, *317*, 796–799.