Tests of Regional Climate Model Validity in the Drought Exceptional Circumstances Report

David R.B. Stockwell

August 5, 2008

Niche Modeling (http://landshape.org/enm)
davids99us@gmail.com

1 Abstract

In a statistical re-analysis of the data from the Drought Exceptional Circumstances Report, all climate models failed standard internal validation tests for regional droughted area in Australia over the last century. The most worrying failure was that simulations showed increases in droughted area over the last century in all regions, while the observed trends in drought decreased in five of the seven regions identified in the CSIRO/Bureau of Meteorology report. Therefore there is no credible basis for the claims of increasing frequency of Exceptional Circumstances declarations made in the report. These results are consistent with other studies finding lack of adequate validation in global warming effects modeling, and lack of skill of climate models at the regional scale.

2 Preamble

The Drought Exceptional Circumstances report (DEC) was commissioned by the Australian Government Ministry for Agriculture and Fisheries and performed by the Centre for Australian Weather and Climate Research (CAWCR), to inform future drought exceptional circumstances (EC) policy in Australia.[1] A press release from the client department released on 6 July 2008 began "Australia could experience drought twice as often and the events will be twice as severe within 20 to 30 years, according to a new Bureau of Meteorology and CSIRO report".[2] Initially concerned with the lack of statistical significance testing in the report, I requested the data to test the significance of their findings. When the data was not immediately forthcoming, a number of bloggers took up the story.[3, 4, 5] We would like to think this attention played a part in the subsequent release of the data facilitating this investigation.

Here I examine support for these claims from the DEC report:

> If rainfall were the sole trigger for EC declarations, then the mean projections for 2010-2040 indicate that more declarations would be likely, and over larger areas, in the SW, SWWA and Vic&Tas regions, with little detectable change in the other regions (Claim 1). Under the high scenario, EC declarations would likely be triggered about twice as often and over twice the area in all regions. In SWWA the frequency and areas covered would likely be even greater (Claim 2).

3 Method

The DEC report projected the incidence of droughts due to global warming to 2040 in seven Australian regions (see report for details [1]). The projections used thirteen global climate models (GCMs) simulating temperature and rainfall and an additional soil moisture level model. These three variables were post-processed for the incidence of droughts during this period. The critical threshold value for drought was defined as the 5th percentile for rainfall and soil moisture, based on all years of available data. This report only examines the rainfall variable because low rainfall is a critical to the EC definition of drought.

Predictive studies are a two-step process: 1) determine the skill of the model in simulating the variable of interest, and 2) extrapolate the model, in this case into the future. Standard tests of model skill are either internal (in-sample) validation, where skill is calculated on data used to calibrate the model, or external (out-of-sample) validation, where skill is calculated on held-back data. As external validation is the higher hurdle, poor internal validation blocks further use of the model. Here internal validation is performed on the thirteen models over the period 1900 to 2007 for each of the seven Australian regions.

The internal validation statistics were:

1. Trend: Significance of the difference between the observed trend (trend-o) and the simulated mean trend (mean-trend-e) in area under drought,

shown as a probability of the null hypothesis (p), calculated as a t-test (t-test) with standard deviation of the simulated trends (sd-trend-e).

2. Correlation: The coefficient of determination (r2), the fraction of variation explained by the model and the probability of false rejection of the null hypothesis of no correlation (corr-sig).

3. Efficiency: The Nash-Sutcliffe coefficient of efficiency (ns-eff), where an efficiency of one is exact prediction, zero is as good as the mean, and negative values indicate less accuracy than the mean.[6]

4. Return Period: The probability of significance of the difference between the observed trend and mean trend projected for the *return period* (returnp-p), the mean time between successive droughts at the given level.

The data was extracted from the calendar year lower 5% rainfall file `http://www.bom.gov.au/climate/droughtec/rain.5pc.tar`, and the projected lower 5% file `http://www.bom.gov.au/climate/droughtec/rain.proj.5pc.tar`.[7] The analysis was performed using the R language and is available at the authors website.[8]

4 Results

The mean, standard deviation and t-tests for trends are recorded in Table 1, and the statistical results for each region are recorded in Tables 2-8. The results were as follows.

1. The mean trend for droughted area differ significantly from the observed trend in every region except Vic&Tas. While drought area decreased in the last century in all regions of Australia except for Vic&Tas and SW-WA, the models simulated increase in droughted area in all regions. The Vic&Tas region has very low observed trend (+1% per year) in droughted area. This means the climate models are significantly biased in the opposite direction to observed drought trends.

2. In almost all cases, the correlation coefficient between simulated and observed values was very low, and not significant. The models on average explained less than 1% of the observed variation in rainfall.

3. In all cases the efficiency was negative. This indicates that the climate models simulate drought area worse than the mean.

4. In almost all cases the difference between the means of the return periods was significant. This indicates the frequency of droughts in the models has no relationship to the actual frequency of droughts. The models miroch-h and ncar-ccsm were exceptions in some regions.

Except in the few cases noted above, the model simulations have no resemblance to patterns of observed droughtedness in the last century. We conclude the models have failed internal validation and no further testing is warranted.

5 Discussion

Due to general failure of climate models at simulating the statistical characteristics regional droughted areas (internal validation), there is no basis for belief in the claim (1) of increasing frequency of EC declarations.

With respect to claim 2, the report represented that they had performed analysis under high global warming scenarios. In fact, they simply quoted the upper 10% of model projections, not alternative warming scenarios (Section 4.1).

> Low, mean and high scenarios are given, where the mean is the 13-model average, and the low and high scenarios are the lowest and highest 10% of the range of model results, respectively.

Irrespective of the lack of validity of the underlying climate projections, there is no logical connection between the extreme values of models, and simulations using different global warming scenarios. Claim 2 is therefore both illogical and invalid.

No importance should be placed in the apparent skill of some models at some statistics in some regions. For example, the climate model miroc_h from the Centre for Climate Research, Japan, shows good agreement with observed return period in most regions. However, it has not been subjected to more rigorous external validation, and has it performed poorly in other tests including a novel pattern-matching approach used in a study including three co-authors of DEC report.[9]

We make the following suggestions:

1. Hennessey *et al.***:** Studies of complex variables like droughts should be conducted with statisticians to ensure the protocol meets the objectives of the study. *A priori* study designs might examine the level of aggregation of data that would reduce noise and provide adequate power, and suitable statistics and tests would be chosen to minimize bias.

2. Drought modeling: The percentage of droughted area appears to be a 'bounded extreme value, peaks over threshold' statistic. The distribution resembles a Pareto (power) law, with a smaller mode where the predicted extent of drought approaches 100% (shown for SW-WA on Figure 1). Recasting the drought modeling problem into known statistical methods might salvage some data from the DEC report. Aggregating the percentage area under drought to the whole of Australia might reduce the boundedness of the distribution, and might also improve the efficiency of the models.

3. Policy makers: Regional effects studies using climate models should be heavily discounted. Scientists across many disciplines have expressed concern with the improper use of models in the environmental sciences.[10] It is claimed that the modeling studies in the IPCC report showed general ignorance of basic validation practices in climate science.[11] However, validation practices are moot when current climate models are not generally reliable.[12, 13] Policy making based on science should not be influenced by studies that appear scientific but do not adhere to the generally accepted validation practices.

6 Acknowledements

I wish to thank the following people for helpful comments: Ian Castles, Peter Gallagher, Steve McIntyre, Geoff Sherrington.

References

[1] K. Hennessy, R. Fawcett, D. Kirono, F. Mpelasoka, D. Jones, J. Batholsa, P. Whetton, M. Stafford Smith, M. Howden, C. Mitchell, and N. Plummer. An assessment of the impact of climate change on the nature and frequency of exceptional climatic events. Technical report, CSIRO and the Australian Bureau of Meteorology for the Australian Bureau of Rural Sciences, 33pp., 2008.

[2] DAFF08/084B. Droughts to be more severe and occur more often in the future. 2008.

[3] Andrew Bolt. Why won't the csiro come clean on its scare report? *http://blogs.news.com.au/heraldsun/andrewbolt*, 2008.

[4] Peter Gallagher. Csiro's (ab)use of intellectual property. *http://www.petergallagher.com.au*, 2008.

[5] Steve McIntyre. Csiro adopts phil jones' stonewall tactic. *http://www.climateaudit.org*, 2008.

[6] J. E. Nash and J. V. Sutcliffe. River flow forecasting through conceptual models part i — a discussion of principles. *Journal of Hydrology*, 10(3):282–290, 1970.

[7] Bureau of Meterorology. Exceptional circumstances report. *http://www.bom.gov.au/climate/droughtec*, 2008.

[8] David R. B. Stockwell. Niche modeling. *http://landshape.org/enm*, 2008.

[9] R. Suppiah, K.J. Hennessy, P.H. Whetton, K. McInnes, I. Macadam, J. Bathols, J. Ricketts, and C.M. Page. Australian climate change projections derived from simulations performed for the ipcc 4th assessment report. 131-152. *Australian Meteorological Magazine*, 56(3), 2007.

[10] Daniel B. Botkin, Henrik Saxe, Miguel B. Araujo, Richard Betts, Richard H. W. Bradshaw, Tomas Cedhagen, Peter Chesson, Terry P. Dawson, Julie R. Etterson, Daniel P. Faith, Simon Ferrier, Antoine Guisan, Anja Skjoldbord Hansen, David W. Hilbert, Craig Loehle, Chris Margules, Mark New, Matthew J. Sobel, and David R. B. Stockwell. Forecasting the effects of global warming on biodiversity. *BioScience*, 57(3):227–236, 2007.

[11] Green K.C. and Armstrong J.S. Global warming: Experts' opinions versus scientific forecasts. Technical report, National Center for Policy Analysis, 2008.

[12] D. Koutsoyiannis, A. Efstratiadis, N. Mamassis, and A. Christofides. On the credibility of climate predictions. *Hydrological Sciences Journal*, 53(4):671–684, 2008.

[13] Patrick Frank. A climate of belief. *Skeptic*, 14(1), 2008.

7 Tables

	trend-o	mean-trend-e	sd-trend-e	t-test	p
MDB	−0.03	0.02	0.05	3.44	0.00
NSW	−0.03	0.02	0.06	2.83	0.02
NWAust	−0.03	0.00	0.04	2.86	0.01
Qld	−0.10	0.02	0.05	7.50	0.00
SW-WA	0.05	0.07	0.04	2.30	0.04
SWAust	−0.02	0.02	0.04	3.44	0.00
VicTas	0.01	0.03	0.06	1.03	0.32

Table 1: t-test of difference in mean of predicted trends to the observed mean of droughted area

	r2	corr-sig	ns-eff	returnp-p
cgcm-t47	0.00	0.78	−1.40	0.01
cgcm-t63	0.00	0.84	−0.97	0.03
csiro-mk3.0	0.00	0.62	−1.36	0.08
csiro-mk3.5	0.00	0.95	−1.16	0.11
giss-aom	0.02	0.20	−0.67	0.01
giss-er	0.02	0.17	−1.04	0.02
iap	0.01	0.33	−1.42	0.00
inmcm	0.00	0.96	−1.14	0.01
ipsl	0.01	0.25	−0.89	0.01
miroc-h	0.00	0.68	−1.47	0.23
miroc-m	0.00	0.76	−1.26	0.05
mri	0.01	0.45	−1.07	0.11
ncar-ccsm	0.01	0.24	−1.20	0.56

Table 2: Internal validation statistics for MDB

	r2	corr-sig	ns-eff	returnp-p
cgcm-t47	0.00	0.57	−2.20	0.01
cgcm-t63	0.00	0.75	−1.15	0.00
csiro-mk3.0	0.00	0.54	−1.53	0.01
csiro-mk3.5	0.00	0.58	−1.34	0.03
giss-aom	0.01	0.26	−1.12	0.01
giss-er	0.01	0.30	−1.25	0.00
iap	0.01	0.31	−1.65	0.00
inmcm	0.00	0.85	−1.61	0.00
ipsl	0.01	0.45	−1.17	0.00
miroc-h	0.01	0.44	−1.70	0.01
miroc-m	0.01	0.43	−1.11	0.01
mri	0.00	0.69	−1.24	0.01
ncar-ccsm	0.01	0.34	−1.67	0.01

Table 3: Internal validation statistics for NSW

	r2	corr-sig	ns-eff	returnp-p
cgcm-t47	0.00	0.61	−1.28	0.00
cgcm-t63	0.04	0.05	−0.77	0.00
csiro-mk3.0	0.00	0.75	−1.61	0.00
csiro-mk3.5	0.02	0.16	−1.59	0.00
giss-aom	0.03	0.06	−1.77	0.01
giss-er	0.00	0.95	−1.75	0.00
iap	0.00	0.52	−1.91	0.00
inmcm	0.00	0.64	−2.01	0.00
ipsl	0.00	0.63	−2.16	0.00
miroc-h	0.01	0.37	−1.66	0.27
miroc-m	0.00	0.57	−2.26	0.00
mri	0.01	0.40	−1.94	0.00
ncar-ccsm	0.00	0.60	−0.59	0.45

Table 4: Internal validation statistics for NWAust

	r2	corr-sig	ns-eff	returnp-p
cgcm-t47	0.00	0.79	−1.05	0.01
cgcm-t63	0.01	0.39	−0.72	0.03
csiro-mk3.0	0.00	0.75	−1.01	0.01
csiro-mk3.5	0.00	0.63	−1.34	0.04
giss-aom	0.00	0.74	−2.03	0.01
giss-er	0.01	0.30	−1.11	0.00
iap	0.01	0.44	−1.25	0.00
inmcm	0.01	0.33	−1.38	0.00
ipsl	0.01	0.27	−1.72	0.00
miroc-h	0.00	0.67	−1.58	0.07
miroc-m	0.00	0.69	−1.39	0.00
mri	0.00	0.67	−0.95	0.01
ncar-ccsm	0.00	0.96	−0.68	0.39

Table 5: Internal validation statistics for Qld

	r2	corr-sig	ns-eff	returnp-p
cgcm-t47	0.01	0.44	−2.09	0.08
cgcm-t63	0.01	0.37	−1.73	0.16
csiro-mk3.0	0.04	0.04	−0.78	0.08
csiro-mk3.5	0.01	0.35	−1.40	0.10
giss-aom	0.00	0.76	−1.72	0.03
giss-er	0.04	0.05	−1.21	0.08
iap	0.00	0.64	−1.00	0.35
inmcm	0.00	0.55	−1.40	0.10
ipsl	0.01	0.45	−1.10	0.03
miroc-h	0.00	0.83	−1.03	0.34
miroc-m	0.01	0.41	−2.11	0.18
mri	0.00	0.62	−0.80	0.69
ncar-ccsm	0.00	0.82	−1.25	0.15

Table 6: Internal validation statistics for SW-WA

	r2	corr-sig	ns-eff	returnp-p
cgcm-t47	0.01	0.37	−1.01	0.00
cgcm-t63	0.02	0.18	−1.30	0.00
csiro-mk3.0	0.00	0.64	−1.77	0.00
csiro-mk3.5	0.01	0.34	−2.00	0.00
giss-aom	0.00	0.97	−1.37	0.00
giss-er	0.01	0.46	−1.57	0.00
iap	0.01	0.34	−1.96	0.00
inmcm	0.04	0.04	−1.46	0.00
ipsl	0.00	0.71	−1.24	0.01
miroc-h	0.00	0.59	−1.10	0.47
miroc-m	0.00	0.65	−1.82	0.00
mri	0.01	0.45	−1.25	0.01
ncar-ccsm	0.00	0.70	−1.24	0.00

Table 7: Internal validation statistics for SWAust

	r2	corr-sig	ns-eff	returnp-p
cgcm-t47	0.00	0.92	−1.03	0.01
cgcm-t63	0.01	0.29	−1.12	0.01
csiro-mk3.0	0.01	0.26	−1.15	0.00
csiro-mk3.5	0.00	0.53	−1.03	0.01
giss-aom	0.04	0.03	−0.57	0.01
giss-er	0.01	0.40	−1.50	0.01
iap	0.01	0.31	−0.73	0.02
inmcm	0.00	0.58	−1.41	0.02
ipsl	0.03	0.06	−0.69	0.00
miroc-h	0.02	0.15	−0.32	0.31
miroc-m	0.01	0.33	−1.54	0.01
mri	0.02	0.20	−1.11	0.10
ncar-ccsm	0.01	0.25	−0.75	0.15

Table 8: Internal validation statistics for VicTas

8 Figures

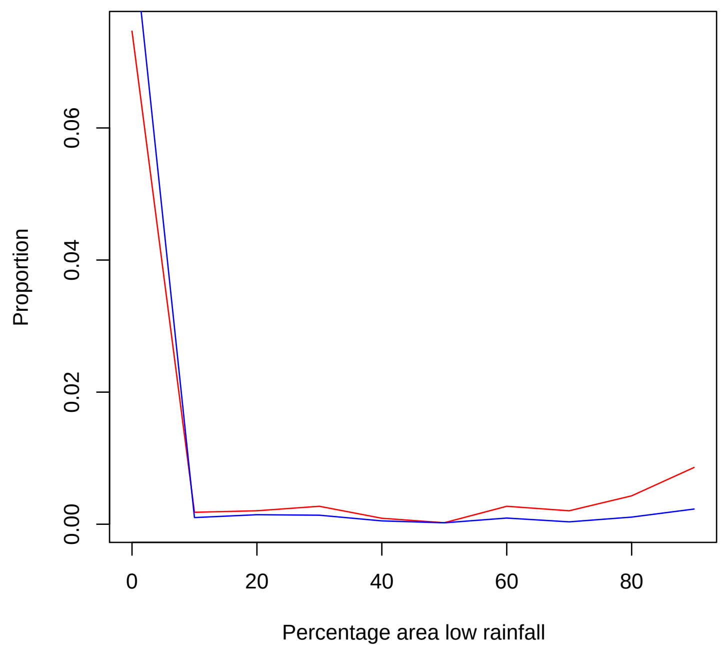

Figure 1: Distribution of simulated (1900-2007 in blue) and projected (2008-2040 in red) droughted areas in SW-WA.