

# Estimating trends in monthly maximum and minimum temperatures in GCMs for which these data are not archived

**J.H. Ricketts<sup>a</sup> and J.O. Carter<sup>a</sup>**

<sup>a</sup>*Queensland Climate Change Centre of Excellence, Ecosciences Precinct,  
41 Boggo Rd, Dutton Park, Queensland. (Email:james.ricketts@climatechange.qld.gov.au)*

**Abstract:** Information on future trends in maximum and minimum temperatures is central to the problem of plant growth modelling. Whilst these variables were available for seven global climate models in the CMIP3 dataset, they were not archived for fourteen. This paper addresses the problem of producing estimated monthly trends for these two variables from monthly trends in other more available variables.

We are currently engaged in a project called “Consistent Climate Scenarios” which aims to produce a consistent data source for input into agricultural projects for any location in Australia. It is important that the datasets must be “weather-like” and be internally consistent in terms of variation of variables and cross-variable variation at point scale, and spatially and temporally correlated in a weather-like fashion.

The approach is to combine climate change projections from GCMs and historical daily data from the SILO patched point and gridded data sets (<http://www.longpaddock.qld.gov.au/silo/>). The initial phase of the project involves using trends derived from monthly patterns of change provided by CSIRO and made available from their OzClim 3™ web site (<http://www.csiro.au/ozclim>). During the course of this work we needed to incorporate trends from variables which were not available in OzClim. CSIRO also supplied us with pattern-scaled data for a number of additional climate variables including cloud cover (CLT), and precipitable water (PRW).

Our approach to estimating the trends in maximum and minimum temperatures (*Tasmax* and *Tasmin* respectively) is a multistage one, combining genetic programming tool (the free software tool "Eureqa" from Cornell Creative Machines Lab) to explore possible relationships, with expert knowledge to produce a multiple regression model.

Additionally, a simpler method of estimation was applied. Taking the GCMs for which we had data, we computed the ensemble deviation of maximum and minimum temperatures from the ensemble mean at each point and then adjusted the means of the two models of interest using the ensemble deviation.

We find that trends in monthly maximum and minimum average temperature can be predicted over Australia at GCM native resolutions using trends in mean temperature, cloud cover, and precipitable water, and that the resulting statistical model explains better than 95% of variance over all available GCMs. The final statistical model derived in our work is,

$$T_m = P_m^1 \cdot tas_m + P_m^2 \cdot prw_m + P_m^3 \cdot clt_m + P_m^4 \cdot tas_m clt_m + P_m^5 \cdot tas_m prw_m,$$

where  $T_m$  is either maximum or minimum temperature for each month  $m$ ,  $tas_m$ ,  $prw_m$ , and  $clt_m$  are trends per degree of global warming in *mean surface temperature*, *precipitable water*, and *cloud cover* respectively. The five parameters were derived by multiple linear regression using all grid points over continental Australia, at native grid scale from six of seven GCMs for which we were provided the trends.

**Keywords:** Estimation, genetic programming, trend, temperature, GCM

## 1. INTRODUCTION

### 1.1. Consistent Climate Scenarios

We are currently engaged in a project called “Consistent Climate Scenarios” which aims to produce consistent climate data for input into agricultural modeling projects. It is important that the datasets must be “weather-like” and be internally consistent in terms of variation of variables and cross-variable variation at point scale, and spatially and temporally correlated in a weather-like fashion continentally.

The approach is to combine data change projections from GCMs with historical daily data from the SILO patched point and gridded data sets (<http://www.longpaddock.qld.gov.au/silo/>). The initial phase of the project involves using trends derived from monthly patterns of change provided by CSIRO and made available from their OzClim 3™ web site (<http://www.csiro.au/ozclim>). These are then being used to perturb selected daily climate variable series from 1900-2100 at specified locations and output in a variety of formats (e.g. files compatible with the Agricultural Production Systems Simulator (APSIM) (<http://www.asris.csiro.au/downloads/ApsimFactSheet.pdf>)). Four GCMs have been designated as not suited to the Australian region based on recommendations of an expert panel. In all we will supply data derived from nineteen GCMs.

### 1.2. Patterns of change

Pattern scaling of global climate models (GCMs) is a climate projection technique which was first suggested by Santer (Santer et al. 1990), and which was used in the IPCC First Assessment Report to generate scenarios from CO<sub>2</sub> doubling experiments (Mitchell et al. 1990). The method has been developed and used ever since. Further refining (Jonas et al. 1996) who used a signal to noise method, pattern scaling is given by (Mitchell et al. 1999) as a linear least squares method over a sequence of periods. CSIRO’s implementation of the technique, used in OzClim, regresses GCM derived mean annual global temperatures against grid point mean monthly values over the 21<sup>st</sup> century GCM run, on a month of year basis. That gives for every grid point, 12 separate month-wise estimates of trend in a climate variable expressed as change per degree of predicted mean annual global warming. Then, following (Mitchell 2003) these are scaled against output from the MAGICC simple climate model (SCM) to give predicted changes at years of interest. OzClim is further documented in (Ricketts, Page 2007). Pattern scaled files are provided using either of two conventions. Temperature trends are provided as absolute trends per degree of global warming. Rainfall trends are provided as trend per degree of global warming divided by the model baseline average over the 1975-2004 interval, expressed as a percentage trend. The former are known as “trend files”, the latter as “tpc files”.

During the course of this work we needed to incorporate trends from variables which were not available in OzClim. CSIRO also supplied us with additional pattern-scaled data for a number of other climate variables (**Table 1**) including cloud cover (CLT), and precipitable water (PRW).

### 1.3. The CMIP3 data set

The global climate model trends and averages, the GCM global mean annual temperatures, and the temperatures corresponding to the SRES emissions from MAGICC were supplied by CSIRO. Their source of the GCM data was PCMDI’s web site which is a repository of data submitted for the third climate model inter-comparison project (CMIP3). To this was added CSIROm3.5. All GCMs have had precipitation and mean temperature data submitted as monthly averages. Only seven groups submitted mean monthly averages of maximum and minimum land surface temperatures (known as *tasmax* and *tasmin* respectively).

**Table 1:** Variables provided by CSIRO, some as linear trends of change per degree of global warming; some as “tpc” – linear trends expressed as a percentage of the model reference mean.

Climate Variable	Mnemonic name	Trend type
Mean temperature	Tas	Linear trend (trend)
Maximum temperature	Tasmax	Linear trend (trend)
Minimum temperature	Tasmin	Linear trend (trend)
Rainfall	Pr	Percentage trend (tpc)

Total column water vapor/Precipitable water	PRW	Percentage trend (tpc)
Cloud fraction	CLT	Linear trend (trend)
Screen height specific humidity	HUSS	Percentage trend (tpc)
Screen height relative humidity	HURS	Percentage trend (tpc)
Windspeed	WSP	Percentage trend (tpc)
Downward shortwave solar radiation	RSDS	Percentage trend (tpc)

#### 1.4. Infilling

There are few papers in which the specific problem estimating trends in monthly averaged *tasmax* and *tasmin* from GCM data has been documented. Furthermore, in this case the trends are not time trends, but trends per degree of global warming. One possibility, the “ratios method”, is to use the ensemble mean trends from those models for which data is available and compute a mean ratio of each of these to the mean trend of GCM average temperature (*tas*), applying the derived ratio across all remaining GCMs. Another is simply to apply the trend of the mean temperature.

GCMs evolve differently from each other in the way they distributed such variables as cloud cover, water vapour and rainfall. However, given the underlying physics, it seemed reasonable to seek for simple correlations between them. One option would be to perform a stepwise multiple regression, but this approach would not provide combinations of variables and so it was decided to use a two stage, exploratory process.

#### 1.5. Genetic programming, Eureqa, and automated solution seeking

Eureqa is a fairly new, publically available product from Cornell Creative Machines Lab <http://creativemachines.cornell.edu/eureqa> which is aimed at automated numerical regression methods and optimization, and is based on genetic programming (GP), first introduced by John Koza in 1992, (Koza 1992). Input to Eureqa consists of rows of variable values. Searches are specified by writing a search function. For instance  $Tasmin = f(tas, hurs, huss, wsp, prw, clt, rsds)$  would specify that *Tasmin* be predicted by some combination of the RHS variables. The output of Eureqa is a list of candidate functions ranked by fitness (error/complexity), a plot of the selected solution against the target values, some fitting statistics of the selected solution and lastly, a plot of the Pareto coverage front formed by the components of the fitness, error against function complexity (Figure 1).

## 2. THE RATIOS METHOD

All seven models for which trends in *tasmax* and *tasmin* are available, were used to estimate an ensemble mean ratio of trends in maximum temperature to trends in minimum temperature.

At each native model grid-point (*p*), for every month (*m*), from all available models (*G*) we compute the ensemble mean ratio of *Tasmax* to *Tasmin* for each month and at each grid-point. *N* is the number of models in the ensemble.

$$Ratio_{m,p} = \frac{\sum_G \left( \frac{Tasmax}{Tasmin} \right)_{G,m,p}}{N} \quad (1)$$

Once we have these then for every model we are infilling (*I*) we use the provided mean temperature trend at each month and grid-point  $Tmean_{I,m,p}$  from which we then can compute the estimated temperatures using the model mean temperature and the ensemble ratios.

$$Tasmin_{I,m,p} = \frac{2 \times Tmean_{I,m,p}}{1 + Ratio_{m,p}} \quad (2)$$

$$Tasmax_{I,m,p} = Ratio_{m,p} \times Tasmin_{I,m,p} \quad (3)$$

### 3. REGRESSION APPROACH

The aim of this work was to produce if possible, a single multiple regression model applicable to trends in both maximum and minimum temperatures.

The regression approach involved consideration of theoretical relationships followed by a two phase exploration using Eureqa, after which a multiple linear regression analysis was used to produce a statistical model.

We perform a full multiple linear regression on a month-by-month basis using all of the data points from all of the GCMs at grid scale. The aim of the multiple linear regression analysis is to obtain a single statistical model for both *Tasmax* and *Tasmin* and all months, hence, it is acceptable for some coefficients to be zero for some months. Finally cross validation is performed to ensure reasonable estimates of the homogeneity of prediction.

#### 3.1. Consideration of theoretical relationships

It is known that in general, trends in *tasmin* under global warming are likely to be greater than trends in *tasmax*, depending on trends in water related variables. (E.g.) Trends in water vapour and cloudiness are often partially correlated (Caesar 2006). Thus it is reasonable to test whether estimates of the trends in temperatures are better predicted if the interaction of the partially correlated variables is taken into account. Mean temperature, being defined as the average of maximum and minimum temperatures, must therefore be predictive. FAO-56, (Allen 1996) identifies wind speed, humidity, and radiation as being related variables.

Steps one and two use data from a representative sample of locations based on current weather stations. The aim of step one is to identify a set of climate variables which appear to be of interest after the theoretical considerations, possibly eliminating some climate variables and expanding the set of usable climate models. The aim of step two is to identify the most influential climate variables and their interactions this time using the expanded set of GCMs and performing the analysis on a monthly basis. From this we get a list of variables which are influential in at least some months and a list of interactions which can also be considered.

The potential climate variables of interest were trends in *tasmin*, *tasmax*, *tas*, *hurs*, *huss*, *wsp*, *prw*, *clt*, *rsds*. However *hurs*, *huss*, *wsp*, and *rsds* were not available for all of the target GCMs. The GCMs for which we hold trend files for all variables of potential interest are (MIROC-H, MIROC-M, GISS-AOM, INMCM3, CSIRO-Mk3.5, BCCR and CSIRO-Mk3.0)

#### 3.2. Using Eureqa to explore empirical relationships

We are interested in relationships that hold across models and months, and spatially. Firstly a small representative sample of locations was defined, representative of a variety of climates. Then we select the trends (or percentage trends) per degree of global warming for all months using bilinear interpolation from the trend (or tpc) files at each of these point locations and submit these to an analysis by Eureqa.

Data was loaded into Eureqa from csv files. Eureqa makes available the various mathematical operators, (e.g. add, subtract) and functions (e.g. sin logistic), termed “building blocks”, and these are selected from an extensive list. Also the *fitness metric* (e.g. Mean Absolute Error, AIC), a weighting scheme can be specified. We use Mean Squared Error and no weighting, and confined the building blocks to add, subtract, multiply and divide. The aim is to identify the most influential climate variables and their interactions, performing the analysis on a monthly basis and a minimal set of arithmetic operators. From this we get a list of variables which are influential in at least some months and a list of interactions which can also be considered. Together these are building blocks of a linear additive model.

We ran Eureqa against the specification  $Tasmax$  or  $Tasmin = f(tas, pr, prw, clt)$  for data from each month for ten minutes and then from the four most preferred suggestions from Eureqa we recorded the variables or products of variables that formed the components of each equation. This gave us a list of just six components that came up at least a quarter of the time:

1. *tas* – 100% of the time
2. *tas* \* *clt* – 75%
3. *prw* – 75%
4. *clt* – 50%
5. *pr* – 25%
6. *tas* \* *prw* – 25%.

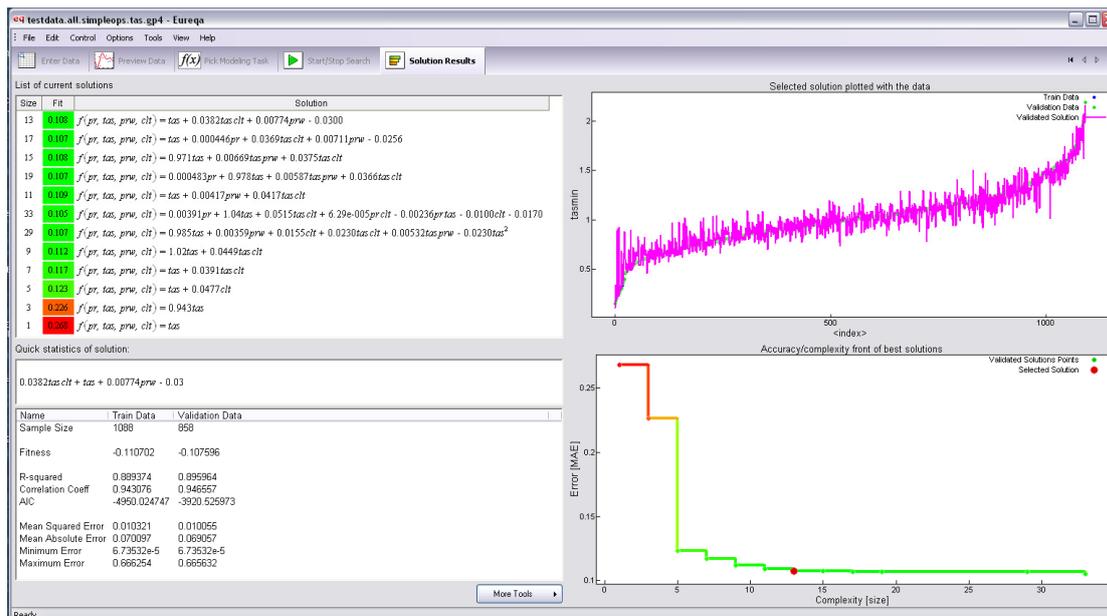


Figure 1: Sample Eureqa™ output at point of termination of *tasmin* run.

### 3.3. Multiple regression

Using a Python script data was then extracted from every land location for each of the seven selected GCMs at GCM native resolution for every month ( $m$ ) from the *trend* files for the three temperature variables (*tas*, *tasmax* and *tasmin*), and *tpc* files for rainfall and precipitable water (*pr* and *prw*), as well as the two interaction terms,  $tas * clt$  and  $tas * prw$ .

The script then computed the coefficients ( $P^1 - 6$ ) and diagnostics for 12 monthly estimated multiple linear regressions of the form.

$$T_m = P_m^1 \cdot tas_m + P_m^2 \cdot prw_m + P_m^3 \cdot clt_m + P_m^4 \cdot tas_m clt_m + P_m^5 \cdot tas_m prw_m + P_m^6 \cdot pr_m \quad (4)$$

Since that run pooled all of the GCM data, we call it the “pooled regression” result. A similar analysis was performed for each GCM separately, to assess the generality of the regression equation. These are the “individual regression” results.

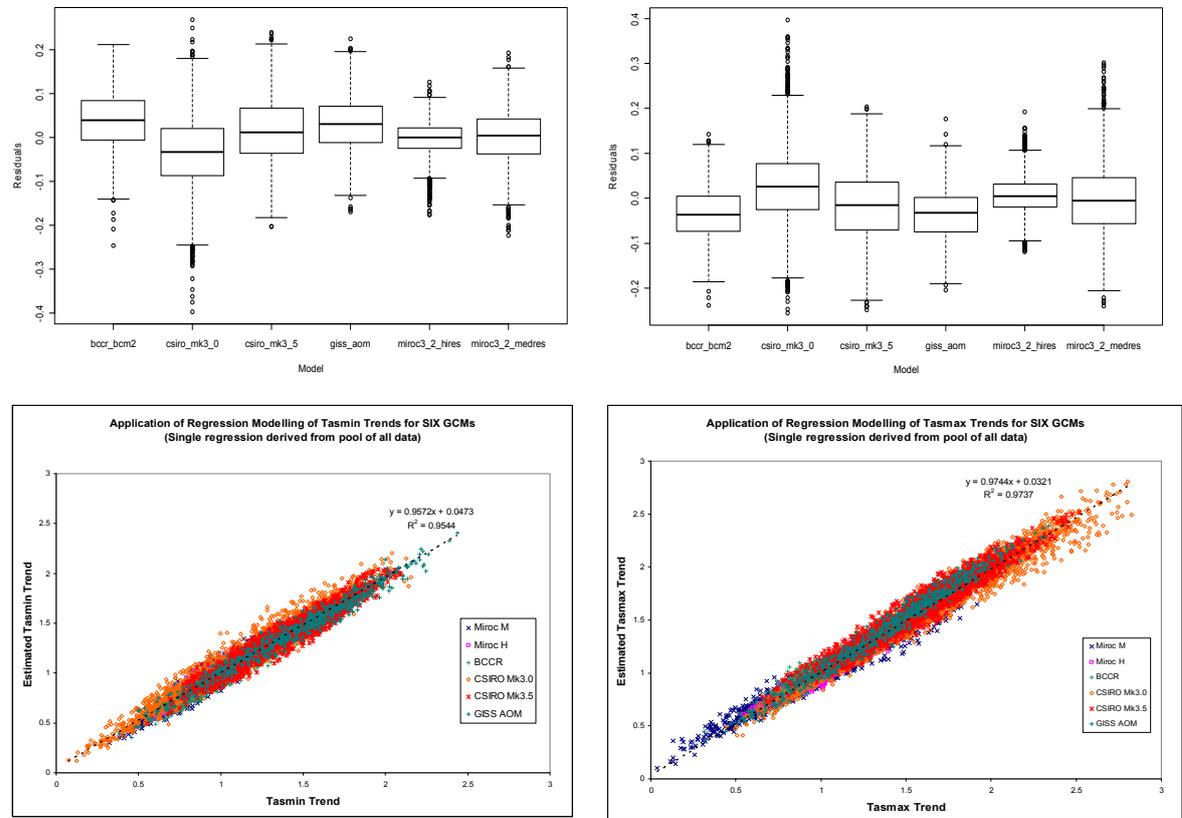
The coefficient for rainfall ( $P^6$ ) was never significantly different from zero and so this term was dropped. The regression coefficients for GCM *tasmax* vs predicted *tasmax* were computed for each of the seven GCMs. The mean was 0.9312 with a standard deviation of 0.1361. The coefficient for INMCM3 was 0.6353, more than two standard deviations from the mean. This GCM was dropped and the regression re-performed.

## 4. RESULTS AND DISCUSSION

Using R, ANOVA with orthogonal contrasts was performed on the residuals for both the pooled regressions and the individual regressions for each of *tasmin* and *tasmax*. For the *pooled regression*, results show that months are not statistically different from each other but that the GCMs (models) do differ in that the variance of their residuals varies (see Figure 2). A similar analysis of the *individual* regressions shows no statistical difference between GCMs or months. Hence, one can conclude that the regression relationship hold across all tested GCMs and is adequate to the task of predicting trends in maximum and minimum temperatures for models where that data is not available, but there is variation in the strength of the relationships between GCMs.

**Estimates of trends in minimum temperatures (Tasmin)**

**Estimates of trends in maximum temperatures (Tasmax)**



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Model	5	5.636	1.127	326.2371	<0.0001***
Month	1	0.002	0.002	0.6988	0.4032
Model:Month	5	0.202	0.040	11.7148	<0.0001***
Residuals	13764	47.553	0.003		

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Model	5	5.133	1.027	271.8476	<0.0001***
Month	1	0.005	0.005	1.3689	0.242
Model:Month	5	0.375	0.075	19.8599	<0.0001***
Residuals	13764	51.977	0.004		

**Figure 2:** Results for analysis of the residuals of each of the six GCMs used to compute the regression model. Top row, box and whisker plots of the residuals distributed by GCM (Model). Second row plots of estimates against the given values. Bottom row, the results of ANOVA showing that the variances of the residuals differ by model, but not by month.

The question arises as to whether estimating trends in maximum and minimum temperatures using the regression method is better than simply using the trends in the mean temperature, or the ratios method. An examination of the correlation coefficients between each of the variables and the three different estimators shows that the regression method is greatly superior, and that just using the mean temperature trend would be simpler, and as accurate as the ratios method (see **Table 2**). This was confirmed by ANOVA.

We find that trends in monthly maximum and minimum average temperature can be predicted over Australia at GCM native resolutions using trends in mean temperature, cloud cover, and precipitable water, and that the resulting statistical model explains better than 95% of variance for *tasmin* and better than 97% of variance for *tasmax* over tested GCMs. The final statistical model derived in our work is,:

$$T_m = P_m^1 \cdot tas_m + P_m^2 \cdot prw_m + P_m^3 \cdot clt_m + P_m^4 \cdot tas_m clt_m + P_m^5 \cdot tas_m prw_m \quad (5)$$

Where  $T_m$  is either maximum or minimum temperature for each month  $m$ ,  $tas_m$ ,  $prw_m$ , and  $clt_m$  are trends per degree of global warming in mean surface temperature, precipitable water, and cloud cover respectively. The five parameters were derived by multiple linear regression using all grid points over continental Australia, at native grid scale from six of seven GCMs for which we were provided the trends.

Given that one GCM was eliminated from the estimation set, the question arises as to how many of the 12 GCMs being estimated may not be well estimated. We use the binomial theorem. Using the inverse of the cumulative distribution, we conclude that the likely number is no more than four ( $p < 0.05$ ).

The improved estimates in temperature trends from this work will enable more of the CMIP3 data set to be considered in multi-model analyses of extreme temperature trends, and used in statistical downscaling. This methodology can also be adapted to infill other variables that are not consistently available in GCM data sets.

**Table 2:**  $R^2$  values for correlation analyses between trends in maximum and minimum temperature and three predictors, (a) trends in mean temperatures, (b) trends computed using the ratios method, and (c) trends using the regression model reported.

Prediction Method	Maximum temperature	Minimum temperature
Mean Temperature ( <i>tas</i> )	0.9003	0.7982
Ratios	0.8987	0.8163
Regression Method	0.9737	0.9544

## 5. ACKNOWLEDGMENTS

This work was performed as part of the Consistent Climate Scenarios Project (CCSP) funded by the Australian Department of Agriculture, Fisheries and Forestry under their Australia's Farming Futures program. We thank CSIRO Marine and Atmospheric Research for providing the OzClim™ data. The assistance of Dorine Bruget, Stuart Burgess, Ken Day, and Andrej Panjkov and two anonymous reviewers are gratefully acknowledged.

## REFERENCES

- Allen, R.G, Pereira, L.S, Raes, D., Smith, M, 1998. *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56*,, FAO - Food and Agriculture Organization of the United Nations, Rome, ISBN 92-5-104219-5
- Caesar, J, Alexander, L. Vose, R. 2006, "Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set", *J Geophysical Res.*, vol. 111, D05101.
- Jonas, M., Fleischmann, K., Ganopolski, A.V., Krabec, J., Sauer, U., Olendrzyński, K., Petoukhov, V.K. & Shaw, R.W. 1996, "Grid point surface air temperature calculations with a fast turnaround: Combining the results of image and a GCM", *Climatic Change*, vol. 34, no. 3-4, pp. 479-512.
- Koza, J.R. 1992, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, Massachusetts.
- Mitchell, J.F.B., Manabe, S., Meleshko, V. & Tokioka, T. 1990, "Equilibrium climate change - and its implications for the future", *Climate change: the IPCC scientific assessment*, , pp. 131-172.
- Mitchell, J.F.B., Johns, T.C., Eagles, M., Ingram, W.J. & Davis, R.A. 1999, *Towards the Construction of Climate Change Scenarios*, Springer Netherlands.
- Mitchell, T.D. 2003, "Pattern Scaling. An Examination of the Accuracy of the Technique for Describing Future Climates", *Climatic Change*, vol. 60, no. 3, pp. 217-242.
- Ricketts, J.H. & Page, C.M. 2007, A Web Based Version of OzClim for Exploring Climate Change Impacts and Risks in the Australian Region. In Oxley, L. and Kulasiri, D. (eds) MODSIM 2007 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2007, pp. 560-566.
- Santer, B.D., Wigley, T.M.L., Schlesinger, M.E. & Mitchell, J.F.B. 1990, *Developing Climate Scenarios from Equilibrium GCM Results*, Hamburg.