Stochastic Downscaling of Numerical Climate Model Simulations

Bryson C. Bates, Stephen P. Charles and James P. Hughes*
CSIRO Land and Water, Private Bag PO, Wembley WA 6014, Australia
*Department of Biostatistics, University of Washington, Seattle WA 98195, USA

Abstract Stochastic downscaling and limited area models (LAMs) have been proposed as tools for overcoming the poor performance of general circulation model (GCM) simulations of local to regional scale precipitation under present climate conditions. Although GCMs perform reasonably well in simulating synoptic atmospheric fields, they tend to over-estimate the frequency and under-estimate the intensity of daily precipitation and do not reproduce the statistics of historical records at the spatial scales of interest in regional impact analyses. This paper describes the application of a nonhomogeneous hidden Markov model (NHMM) fitted to observed atmospheric and precipitation data to LAM atmospheric fields for South-West Western Australia. The NHMM is unique amongst stochastic downscaling methods in that it determines the most distinct patterns in a multi-site, precipitation occurrence record rather than patterns in atmospheric fields. These patterns can in turn be defined as conditionally dependent on a range of raw and derived (e.g., gradients, lags and principal components of) daily atmospheric series. We compare a LAM simulation and the downscaled LAM simulation with observed 'winter' precipitation statistics at five stations near Perth, Western Australia. The results show that the downscaled simulations reproduce observed precipitation probabilities and wet and dry spell frequencies at each station. The LAM simulation tends to under-estimate the frequency of dry spells and over-estimate the probability of precipitation and the frequency of wet spells.

1. INTRODUCTION

Interest in regional climate simulation has grown from the increasing demands of the scientific community, policy makers and the general public for realistic assessments of the regional impacts of natural climate variability and possible climate change due to the enhanced greenhouse effect. General circulation models (GCMs) perform reasonably well in simulating the present climate with respect to annual or seasonal averages at large spatial scales (>10^4 km^2), but poorly at the smaller space and time scales relevant to regional impact analyses [Grotch and MacCracken, 1991; Robock et al., 1993; Houghton et al., 1996, p. 44]. Differences between model simulations of precipitation and surface air temperature also seem to be greater at the regional scale [Gates et al., 1996].

The poor performance of GCMs at local and regional scales has led to the development of limited area models (LAMs) in which a fine computational grid over a limited domain is nested within the coarse grid of a GCM. The host GCM provides the large-scale synoptic forcing to the LAM through the LAM's lateral boundaries. The main advantages of LAMs are that they: allow a more accurate representation of orography; simulate regional climate at a higher spatial resolution than the host GCM; and are more economical to run than a GCM with a similar spatial resolution [Walsh and McGregor, 1995].

Although GCMs and LAMs perform reasonably well in simulating synoptic scale atmospheric fields, they tend to over-estimate the frequency and under-estimate the intensity of daily precipitation and thus fail to reproduce the statistics of historical records at the spatial scales of interest in impact analyses [Mearns et al., 1995; Walsh and McGregor, 1995]. Also, LAMs do not resolve the full structure of precipitating systems even at a spatial resolution of 20 km [M.J. Manton, pers. comm., 1996].

The problems highlighted above have led to the development of stochastic downscaling methods for simulating sub-grid scale precipitation. Many downscaling techniques use weather classification schemes for large-scale atmospheric circulation patterns, and then model the daily precipitation process conditional on the derived patterns [e.g., Kalkstein et al., 1990; Bardossy and Plate, 1991, 1992; Bogardi et al., 1993; Matyasovszky et al., 1993a,b; Von Storch et al., 1993; Wilby et al., 1994; Bartholy et al., 1995; Kidson and Watterson, 1995]. These schemes produce weather patterns that are independent of precipitation occurrence data and have had only limited success in reproducing wet and dry spell length statistics [see, e.g., Hay et al., 1991; Wilson et al., 1991, 1992; Zorita et al., 1995].

In contrast, the nonhomogeneous hidden Markov model (NHMM) of Hughes [1993] and Hughes and Guttorp [1994a,b] determines the most distinct patterns in a multi-site, precipitation occurrence record rather than patterns in atmospheric circulation. These patterns can in turn be defined as conditionally dependent on a range of atmospheric variables. Recent work with historical atmospheric and precipitation data in South-West Western Australia (SWWA) suggests that the NHMM provides credible reproductions of spell length statistics [Charles et al, 1996].

This paper describes an application of a NHMM fitted to observed atmospheric and precipitation data to a LAM simulation for SWWA. We compare the LAM simulation and the downscaled LAM simulation with observed 'winter' precipitation statistics at five stations near Perth, Western Australia. The results show that the downscaled simulations reproduce observed precipitation probabilities and wet and dry spell frequencies at each station. The LAM simulation tends to under-estimate the frequency of dry spells and over-estimate the probability of precipitation and the frequency of wet spells.

2. STUDY AREA AND DATA

The region experiences a 'Mediterranean' climate with mild, wet winters and hot, dry summers. Eighty percent of annual precipitation falls in the period from May to October. This is primarily due to the successive passage of cold fronts lodged between high pressure systems at latitudes 30 to 35°S. Mean daily minimum and maximum temperatures at Perth for the coldest month (July) are 9 and 17.3°C while those for the hottest month (February) are 17.9 and 29.9°C.

Atmospheric data were obtained from the Bureau of Meteorology (Australia) on a Lambert conformal grid of 1100 GMT values for the period from 1978 to 1992. These data were chosen as they are close to the mid-point of the daily precipitation recording period for SWWA (24 hours to 0100 GMT, 9 am local time). The available variables included mean sea level pressure (MSLP) and at 850 hPa and 500 hPa levels the geopotential height (GPH), air temperature, dew point temperature (T_d), and U- and V- wind components. The data were interpolated to a rectangular 3.75° longitudinal by 2.25° latitudinal grid. Twenty-four variables were derived from this data set [Charles et al., 1996, Table 1]. These included the raw variables listed above, the north-south and east-west gradients of the raw variables, and lagged raw variables.

Daily precipitation data for thirty precipitation stations were obtained from the same source for the corresponding period. The location of the stations is shown in Figure 1.

The entire data set was analysed on a seasonal basis; 'winter' (May-October) and 'summer' (November-April). Only 'winter' results will be presented here.

3. DOWNSCALING MODEL

A hidden Markov model is a doubly stochastic process: there is an underlying (unobserved or hidden) stochastic process that can only be observed through another set of stochastic processes that produce the sequence of observed outcomes [Rabiner and Juang, 1986]. Assume that there are a finite number of hidden states, N. At each clock time, t, a new state may be entered according to the state transition probability distribution which

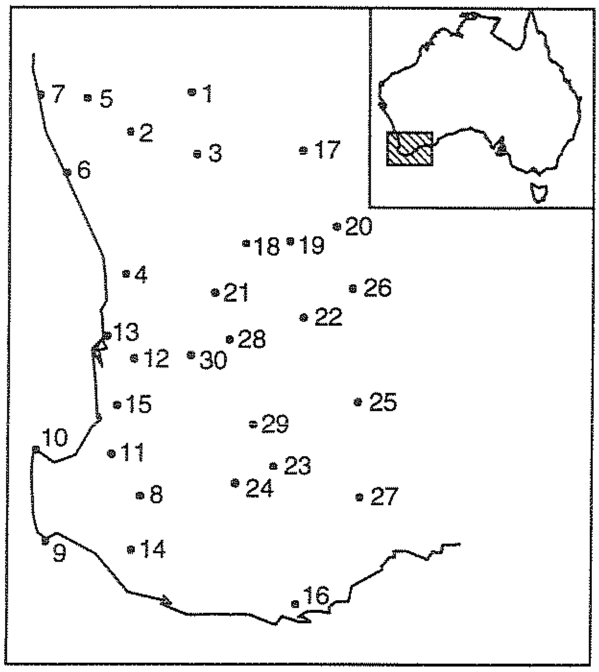

Figure 1: Location of precipitation stations.

depends on the state at time t–1. After each transition, an observable outcome is produced according to a probability distribution that depends on the current state.

The NHMM of Hughes [1993] and Hughes and Guttorp [1994a,b] may be defined by a state transition probability matrix (**A**), a precipitation occurrence probability distribution (**B**) and an initial state distribution (π):

$$A_{ij} = P(S_j^t \mid S_i^{t-1}, \mathbf{X}^t), \ (t = 1, \dots, T; \ i, j = 1, \dots, N) \quad (1)$$

$$B_{jk} = P(R_k^t \mid S_j^t), \ (k = 1, \dots, M) \quad (2)$$

$$\pi_i = P(S_i^{t=1} \mid \mathbf{X}^{t=1}) \quad (3)$$

where S_j^t denotes the jth (unobserved or hidden) weather state at time t, R_k^t the probability of precipitation occurrence at station k and time t, and $\mathbf{X}^t$ the derived atmospheric data at time t. Thus the compact notation $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ may be used to represent the NHMM.

Charles et al. [1996] describe the fitting of the NHMM to historical atmospheric data and the precipitation data for the thirty stations shown in Figure 1. A brief summary follows. Classification trees and other exploratory analyses were used to determine which atmospheric variables had the strongest relationship with precipitation occurrence over the region. These variables were considered candidates for inclusion in the NHMM. Maximum likelihood estimation was used to estimate $\lambda = (A, B, \pi)$ for a series of models with different numbers of weather states and different sets of derived atmospheric variables. The estimation problem is to maximise $P(\mathbf{R}|\lambda)$ where $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_T)$. The resulting model fits were evaluated using an approximation to the Bayes factor, $2l + \log|\mathbf{V}|$, where l is the negative log-likelihood and $\mathbf{V}$ is the covariance matrix of the parameters [Kass and Raftery, 1995]. This criterion is useful for comparing models with different combinations of atmospheric variables. The weather state sequence was obtained from the selected model using the Viterbi algorithm [Forney, 1978]. This permitted the assignment of the daily observations to their respective states.

Table 1 reports the results of the model fitting. For the six and seven state NHMMs, the selected atmospheric variables were mean MSLP, north-south MSLP gradient and 850 hPa east-west gradient. Comparison of the precipitation occurrence patterns associated with each state with their corresponding composite MSLP fields suggested that the six state NHMM has a high degree of physical realism [Charles et al., 1996, Figure 2]. Moreover, the patterns associated with the six states are distinct whereas the patterns for two of the seven states are almost indistinguishable. Thus the six state, three atmospheric variable NHMM was used to downscale climate model simulations.

Table 1: NHMM Selection for 'Winter' Precipitation.

No. of States	Atmospheric Variables*	Degrees of Freedom	$2l + \log\|\mathbf{V}\|$
4	—	132	59255
5	—	170	57194
6	—	210	56487
6	1,4	270	53974
6	1,4,8	300	53744
7	—	252	55831
7	1,4	336	53307
7	1,4,8	378	53287

* 1 = mean MSLP; 4 = north-south MSLP gradient; 8 = east-west 850 hPa GPH gradient.

4. LIMITED AREA MODEL

We used a recent simulation from the CSIRO Division of Atmospheric Research Limited Area Model (DARLAM) with a 125 km grid spacing. Synoptic forcing at the lateral boundaries of the computational grid was provided by the Mk II version of the CSIRO nine-level GCM (CSIRO9 GCM). The spatial resolution of the GCM is roughly 500 km, and DARLAM currently uses the same number of vertical levels as the GCM. Walsh and McGregor, [1995] report that the timings were 30 s per model day for the GCM runs and 130 s per model day for DARLAM simulations over the Australasian region (77 × 77 grid squares) on a CRAY Y/MP computer. They estimated that running CSIRO9 GCM on the same machine at a resolution of about 125 km would require about 1900 s per model day.

We extracted the 1200 GMT fields from the DARLAM simulation as these offered the closest synchronisation with the 1100 GMT historical data used to fit the NHMM. The modelled data were interpolated to the same 3.75° by 2.25° rectangular grid from which the mean MSLP, north-south MSLP gradient and 850 hPa east-wet gradient were obtained. These DARLAM-derived atmospheric variables were used as input to the NHMM to produce sequences of multi-station, daily precipitation occurrence. The NHMM was not fitted to the DARLAM data.

5. RESULTS

A downscaling model should be able to reproduce observed precipitation statistics, such as the distributions of storm durations and interarrival times, that have not been directly incorporated into the model. In the following we compare the ability of the DARLAM and NHMM models to reproduce observed rainfall frequencies, storm durations and storm interarrival times.

Figure 2 compares observed, DARLAM and downscaled DARLAM 'winter' precipitation probabilities for five stations near Perth. These stations fall within a single DARLAM grid square where the simulated probability of rain during 'winter' is 0.66. This probability is about 0.15 higher than the probabilities observed at each of the five stations. However, the probabilities obtained by applying the NHMM fitted to historical data to the DARLAM simulation match the observed probabilities.

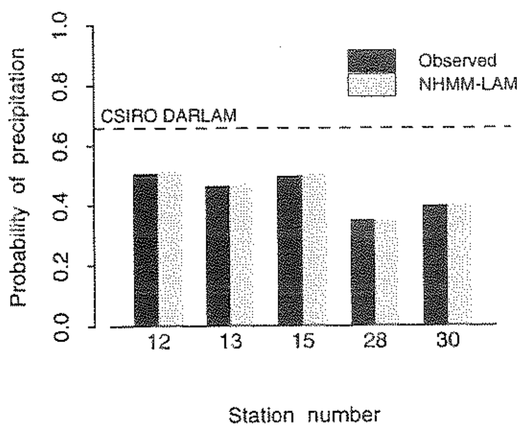

Figure 2: Comparison of observed, DARLAM and downscaled DARLAM 'winter' precipitation probabilities for five precipitation stations near Perth, Western Australia.

Figure 3 compares the cumulative histogram of observed wet spell lengths with the cumulative distribution functions derived from DARLAM and the downscaled DARLAM simulation. Observe that

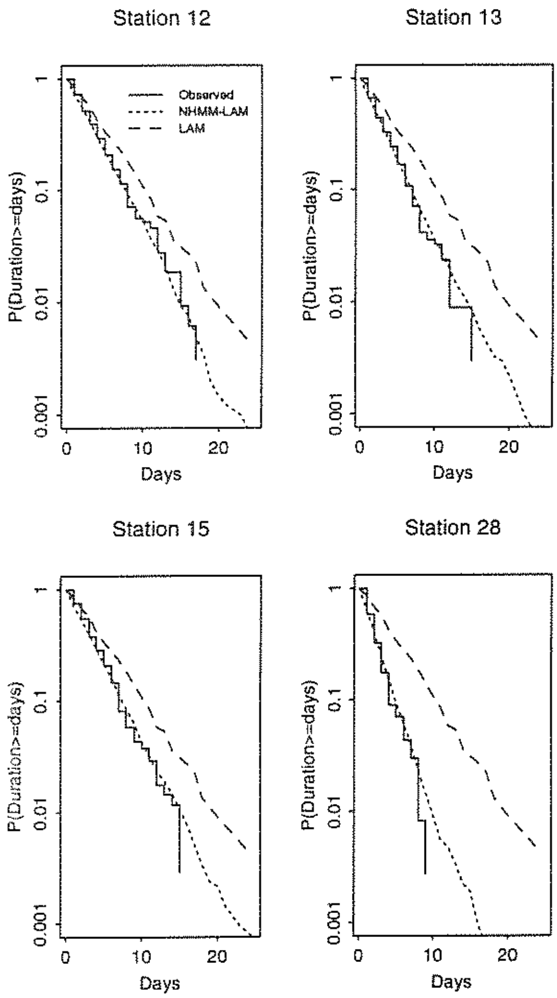

Figure 3: Comparison of observed, DARLAM and downscaled DARLAM wet spell frequencies for four precipitation stations near Perth, Western Australia.

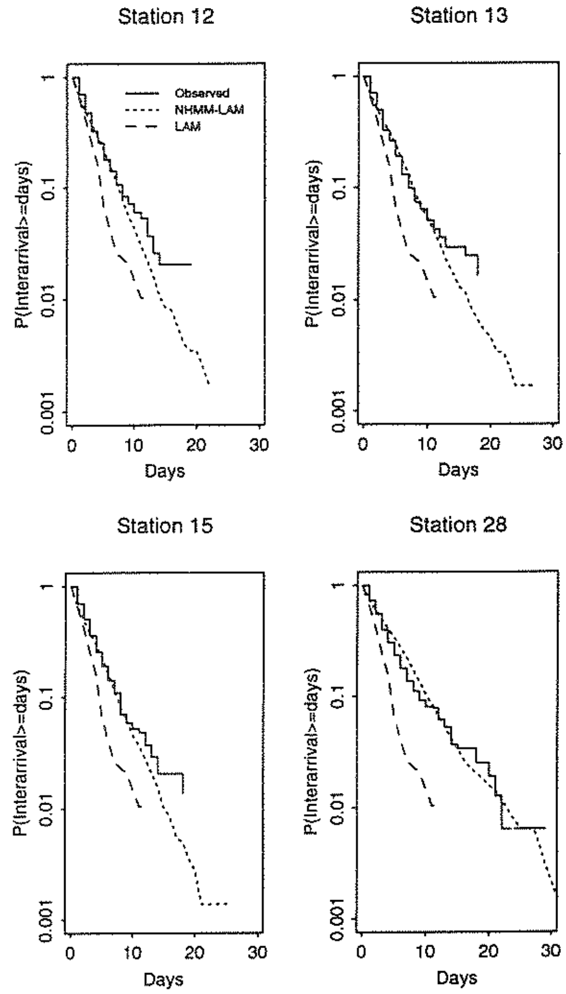

Figure 4: Comparison of observed, DARLAM and downscaled DARLAM dry spell frequencies for four precipitation stations near Perth, Western Australia.

DARLAM over-estimates the frequencies of wet spells across the range of observed spell lengths, and that the ratio of simulated to observed frequency increases with spell length. In contrast, the cumulative frequency distribution for the downscaled DARLAM simulation provides a good approximation to the cumulative histogram.

Figure 4 compares the cumulative histogram of observed dry spell lengths with the cumulative distribution functions derived from DARLAM and the downscaled DARLAM simulation. Observe that DARLAM under-estimates the frequencies of dry spells across the range of observed spell lengths, and that the ratio of observed to simulated frequency increases with spell length. In contrast, the cumulative frequency distribution for the downscaled DARLAM simulation provides a good approximation to the cumulative histogram over most of the domain.

6. CONCLUSIONS AND FURTHER WORK

This paper has investigated the utility of the non-homogeneous hidden Markov model (NHMM) of Hughes [1993] and Hughes and Guttorp [1994a,b] as a tool for downscaling numerical climate model simulations. A NHMM fitted to historical atmospheric and precipitation data for South-West Western Australia was used to downscale a simulation from the CSIRO Division of Atmospheric Research Limited Area Model (CSIRO DARLAM).

The results show that the downscaled simulations reproduce observed precipitation probabilities and wet and dry spell frequencies at five stations located near Perth, Western Australia. The DARLAM simulation tends to under-estimate the frequency of dry spells and over-estimate the probability of precipitation and the frequency of wet spells. Nevertheless, the above results indicate that DARLAM produces reasonable simulations of historical atmospheric circulation patterns. If this were not the case, the downscaled DARLAM simulation would not have been able to reproduce the historical precipitation statistics.

Future work will involve:

- the development of a within-state precipitation amount model for the NHMM that preserves the spatial correlation structure between stations;
- an assessment of the credibility of downscaled GCM simulations;
- the determination of the extent to which the NHMM is applicable across southern Australia where precipitation is generated largely by synoptic systems; and
- an intercomparison of downscaling techniques across a variety of climate regimes and geographical settings.

7. ACKNOWLEDGMENTS

Insights into the synoptic climatology of the study region were provided by P.M. (Mick) Fleming (CSIRO Land and Water, Canberra, Australia) and T.J. Lyons (Murdoch University, Perth, Australia). Our thanks go to R. Srikanthan (Bureau of Meteorology, Australia) for providing the historical atmospheric and precipitation data used in this study, and to J.J. Katzfey (CSIRO Division of Atmospheric Research) for providing the CSIRO DARLAM data. This work contributes to the CSIRO Climate Change Research Program and is part funded through Australia's National Greenhouse Research Program.

8. REFERENCES

Bardossy, A. and E.J. Plate, Modelling daily rainfall using a semi-Markov representation of circulation pattern occurrence, *J. Hydrol., 122,* 33-47, 1991.

Bardossy, A. and E.J. Plate, Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resour. Res., 28*(5), 1247-1259, 1992.

Bartholy, J., I. Bogardi, and I. Matyasovszky, Effect of climate change on regional precipitation in Lake Balaton watershed, *Theor. Appl. Climatol., 51,* 237-250, 1995.

Bogardi, I., I. Matyasovszky, A. Bardossy, and L. Duckstein, Application of a space-time stochastic model for daily precipitation using atmospheric circulation patterns, *J. Geoph. Res., 98*(D9), 16, 653-16,667, 1993.

Charles, S.P., J.P. Hughes, B.C. Bates, and T.J. Lyons, Assessing downscaling models for atmospheric circulation - local precipitation linkage, in *Proc. Int. Conf. Water Resour. & Environ. Res.: Towards the 21st Century*, pp. 269-276, Water Resources Research Center, Kyoto University, Kyoto, Japan, 1996.

Forney, Jr., G.D., The Viterbi algorithm, *Proc. IEEE, 61,* 268-278, 1978.

Gates, W.L., A. Henderson-Sellers, G.J. Boer, C.K. Folland, A. Kitoh, B.J. McAvaney, F. Semazzi, N. Smith, A.J. Weaver, and Q.-C. Zeng, Climate Models - Evaluation, in *Climate Change 1995: The Science of Climate Change*, Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change, edited by J.T. Houghton, L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg, and K. Maskell, pp. 229-284, Cambridge Univ. Press, Cambridge, 1996.

Grotch, S.L. and M.C. MacCracken, The use of general circulation models to predict regional climatic change, *J. Climate, 4,* 286-303, 1991.

Hay, L.E., G.J. McCabe, D.M. Wolock, and M.A. Ayers, Simulation of precipitation by weather type analysis, *Water Resour. Res., 27*(4), 493-501, 1991.

Houghton, J.T., L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg, and K. Maskell, K., Technical Summary, in *Climate Change 1995: The Science of Climate Change*, Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change, edited by J.T. Houghton, L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg, and K. Maskell, pp. 9-50, Cambridge Univ. Press, Cambridge, 1996.

Hughes, J.P., A class of stochastic models for relating synoptic atmospheric patterns to local hydrologic phenomena, PhD thesis, Univ. Washington, Seattle, 1993.

Hughes, J.P. and P. Guttorp, A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena, *Water Resour. Res., 30*(5), 1535-1546, 1994a.

Hughes, J.P. and P. Guttorp, Incorporating spatial dependence and atmospheric data in a model of precipitation, *J. Appl. Meteorol., 33,* 1503-1515, 1994b.

Kalkstein, L.S., P.C. Dunne, and R.S. Vose, Detection of climatic change in the western North American Arctic using a synoptic climatological approach, *J. Climate, 3,* 1153-1167, 1990.

Kass, R.E. and A.E. Raftery, Bayes factors, *J. Am. Statist. Ass., 90,* 773-795, 1995.

Kidson, J.W. and I.G. Watterson, A synoptic climatological evaluation of the changes in the CSIRO nine-level model with doubled CO2 in the New Zealand region, *Int. J. Climatol., 15,* 1179-1194, 1995.

Matyasovszky, I., I. Bogardi, A. Bardossy, and L. Duckstein, Estimation of local precipitation statistics reflecting climate change, *Water Resour. Res., 29*(12), 3955-3968, 1993a.

Matyasovszky, I., I. Bogardi, A. Bardossy, and L. Duckstein, Space-time precipitation reflecting climate change, *Hydrol. Sci. J., 38*(6), 539-558, 1993b.

Mearns, L.O., F. Giorgi, L. McDaniel, and C. Shields, Analysis of daily variability of precipitation in a nested regional climate model: comparison with observations and doubled CO_2 results, *Glob. Planet. Change, 10,* 55-78, 1995.

Rabiner, L.R. and Juang, B.H., An introduction to hidden Markov models, *IEEE Acoustics Speech Signal Processing Mag.*, 4-16, Jan., 1986.

Robock, A., R.P. Turco, M.A. Harwell, T.P. Ackerman, R. Andressen, H.-S. Chang, and M.V.K. Sivakumar, Use of general circulation model output in the creation of climate change scenarios for impact analysis, *Clim. Change, 23,* 293-335, 1993.

Von Storch, H., E. Zorita, and U. Cubasch, Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime, *J. Climate, 6,* 1161-1171, 1993.

Walsh, K.J.E. and J.L. McGregor, January and July climate simulations over the Australian region using a limited-area model, *J. Climate, 8*(10), 2387-2403, 1995.

Wilby, R.L., B. Greenfield, and C. Glenny, A coupled synoptic-hydrological model for climate change impact assessment, *J. Hydrol., 153,* 265-290, 1994.

Wilson, L.L., D.P. Lettenmaier, and E.F. Wood, Simulations of daily precipitation in the Pacific Northwest using a weather classification scheme, *Surveys Geophys., 12,* 127-142, 1991.

Wilson, L.L., D.P. Lettenmaier, and E. Skyllingstad, E. A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation, *J. Geoph. Res., 97*(D3), 2791-2809, 1992.

Zorita, E., J.P. Hughes, D.P. Lettenmaier, and H. Von Storch, Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation, *J. Climate, 8,* 1023-1042, 1995.