Vol. 12: 1–14, 1999 | CLIMATE RESEARCH Clim Res | Published June 15

Validation of downscaling models for changed climate conditions: case study of southwestern Australia

Stephen P. Charles[1,*], Bryson C. Bates[1], Peter H. Whetton[2], James P. Hughes[3]

[1]CSIRO Land and Water, Private Bag, Wembley, Western Australia 6014, Australia
[2]CSIRO Atmospheric Research, Private Bag, Aspendale, Victoria 3195, Australia
[3]Department of Biostatistics, University of Washington, Seattle, Washington 98195, USA

ABSTRACT: Statistical downscaling of general circulation models (GCMs) and limited area models (LAMs) has been promoted as a method for simulating regional- to point-scale precipitation under changed climate conditions. However, several studies have shown that downscaled precipitation is either insensitive to changes in climatic forcing, or inconsistent with the broad-scale changes indicated by the host GCM(s). This has been recently attributed to the omission of the effect that changes in atmospheric moisture content have on precipitation. We describe validation of a nonhomogeneous hidden Markov model (NHMM) for changed climate conditions and apply it to a network of 30 daily precipitation stations in southwestern Australia. NHMMs fitted to $1 \times CO_2$ LAM data were validated by assessing their performance in predicting $2 \times CO_2$ LAM precipitation. The inclusion of 850 hPa dew point temperature depression, a predictor reflecting relative (rather than absolute) atmospheric moisture content, was found to be crucial to successful performance of the NHMM under $2 \times CO_2$ conditions. The NHMM validated for the LAM data was fitted to the historical 30 station network and then used to downscale the $2 \times CO_2$ LAM atmospheric data, producing plausible predictions of station precipitation under $2 \times CO_2$ conditions. Our results highlight that the validation of a statistical downscaling technique for present day conditions does not necessarily imply legitimacy for changed climate conditions. Thus statistical downscaling studies that have not attempted to determine the plausibility of their predictions for the changed climate conditions should be viewed with caution.

KEY WORDS: Climate change modelling · Statistical downscaling · Limited area models · Precipitation occurrence

1. INTRODUCTION

Demand for realistic assessments of the regional impacts of natural climate variability, and possible climate change due to the enhanced greenhouse effect, has generated increased interest in regional climate simulation. Although current 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), they perform poorly at the smaller space and time scales relevant to local and regional impact analyses. Also, inter-GCM variation in either precipitation or surface air temperature simulation is greater at the regional scale (Gates et al. 1996, Houghton et al. 1996, p. 44).

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 (Giorgi & Mearns 1991, Walsh & McGregor 1995). The host GCM provides the large-scale synoptic forcing to the LAM through the LAM's lateral boundaries. Although GCMs and LAMs perform reasonably well in simulating synoptic-scale atmospheric fields, they both tend to overestimate the frequency and

*E-mail: stephen.charles@per.clw.csiro.au

underestimate the intensity of daily precipitation and thus fail to reproduce the statistics of historical records at local scales (Mearns et al. 1995, Walsh & McGregor 1995, 1997, Bates et al. 1998). Moreover, 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 above limitations of GCMs and LAMs have led to the development of statistical downscaling techniques (Hewitson & Crane 1996). Early techniques classified large-scale atmospheric circulation patterns, and then modelled the daily precipitation process conditional on the derived patterns (Bardossy & Plate 1991, 1992, Bogardi et al. 1993, Matyasovszky et al. 1993a,b, Wilby et al. 1994, Bartholy et al. 1995, Kidson & Watterson 1995). These schemes produce weather patterns that are identified independently of the precipitation data and have had only limited success in reproducing wet and dry spell length statistics (Hay et al. 1991, Wilson et al. 1991, 1992, Zorita et al. 1995). When these schemes are used for climate change studies, it is often found that the changes in weather pattern frequencies are insignificant and thus little can be inferred about possible changes in precipitation (Matyasovszky et al. 1993a,b, Wilby et al. 1994). Matyasovszky et al. (1994) used both weather patterns and spatially averaged geopotential height (GPH) at 500 hPa on the basis that this variable exhibited a significant change between GCM results for $1 \times CO_2$ and $2 \times CO_2$ equilibrium climates. Recent research has pursued the application of regression or neural networks to continuous atmospheric indices, rather than the classification of weather patterns. However, these approaches have had limited success in simulating present day precipitation statistics and producing precipitation changes that are consistent with the broad-scale changes indicated by host GCMs (e.g. Huth 1997, Wilby & Wigley 1997, Crane & Hewitson 1998, Wilby et al. 1998).

Other challenges for downscaling have been noted. Wilby (1997) shows that the relationships between circulation and precipitation may not be time invariant over inter-decadal periods. Pittock (1993), Conway et al. (1996) and Huth (1997) warn that the relationships between circulation and local weather variables derived from historical data may not be valid for $2 \times CO_2$ conditions. In particular, Pittock (1993) notes global warming in the models causes an intensification of the hydrological cycle, and Wilby & Wigley (1997) state that changes in atmospheric circulation may not be enough to derive realistic simulations of precipitation for possible future climates. Thus the effect of changes in atmospheric moisture content on precipitation must be also taken into account (Wilby & Wigley 1997).

Here we use the nonhomogeneous hidden Markov model (NHMM) of Hughes et al. (1999) for a downscaling experiment for $2 \times CO_2$ equilibrium climate conditions in southwestern Australia (SWA). The NHMM determines the most distinct patterns in a multi-station precipitation occurrence record, rather than patterns in atmospheric circulation. These patterns can then be defined as conditionally dependent on a range of atmospheric predictor variables. A split sample validation using historical atmospheric and precipitation data has shown that the NHMM can provide credible reproductions of occurrence probabilities and spell length statistics for 30 daily precipitation stations located throughout SWA (Hughes et al. 1999). The present study has 3 major objectives: (1) to assess the potential advantages of using a set of predictor variables containing information about atmospheric moisture content, as well as atmospheric circulation, in a downscaling model; (2) to determine if agreement between the daily precipitation occurrence changes indicated by LAM simulations and those obtained from downscaled LAM atmospheric fields is achievable by objective and independent means; and (3) to develop plausible $2 \times CO_2$ scenarios for precipitation occurrence statistics at existing stations in SWA.

A feature of the study is our attempt to validate the downscaling results for $2 \times CO_2$ conditions (Step 2 above). In other downscaling studies, this is usually done informally by comparing the grid-scale changes in precipitation indicated directly by a GCM with the differences between downscaled $2 \times CO_2$ and historical precipitation (e.g. Wilby et al. 1998). Such comparisons are based on disparate spatial scales and this makes interpretation of results difficult. In contrast, we fit *and* validate the NHMM for the LAM grid data before developing $2 \times CO_2$ scenarios for precipitation occurrence statistics at the stations.

2. STUDY AREA AND DATA

SWA 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 passages of cold fronts lodged between high pressure systems at latitudes 30° to 35° S. Frontal rains spread rapidly northwards and increase in intensity throughout May. They retreat southwards and decrease in intensity during August to October. Thus the entire data set was analysed on a seasonal basis: 'winter' (May to October) and 'summer' (November to April). A full set of 'winter' results will be presented here.

2.1. Limited area model data. A 10 yr simulation of atmospheric and precipitation fields on a 125 km grid for 1

$\times CO_2$ (present day) and $2 \times CO_2$ conditions were obtained from the CSIRO Division of Atmospheric Research limited area model (DARLAM; Walsh & McGregor 1995, 1997). Synoptic forcing at the lateral boundaries of the DARLAM grid was provided by the Mk 2 version of the CSIRO 9-level atmospheric GCM (CSIRO9 GCM; Watterson et al. 1997). DARLAM currently uses the same number of vertical levels as CSIRO9. The spatial resolution of the GCM is roughly 500 km (spectral R21). We extracted the 12:00 h GMT DARLAM fields for precipitation, mean sea level pressure (MSLP), and the GPH, air temperature and humidity fields for 850 and 500 hPa surfaces. These variables correspond with the historical data held in the Commonwealth Bureau of Meteorology's upper air archive. The atmospheric fields were interpolated onto a rectangular 3.75° longitude by 2.25° latitude grid as the DARLAM grid has a finer resolution than that used for the historical archive (see Fig. 1 and Section 2.2). The precipitation data were not interpolated.

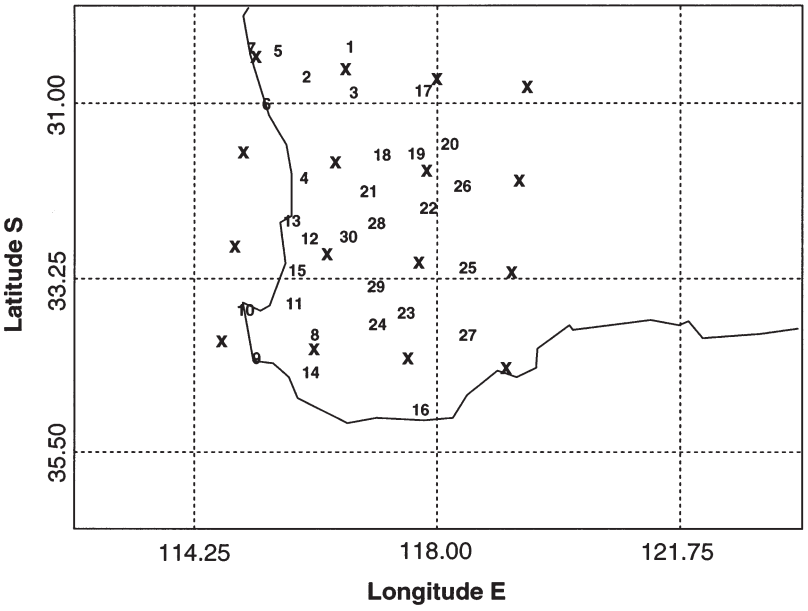


Fig. 1. Location of study area and the precipitation station network. Broken lines denote the grid used for historical and DARLAM atmospheric fields, crosses denote DARLAM grid points and numerals denote precipitation stations. In Table 2, Grid Point 1 is the northwesternmost grid point (adjacent to Stn 7), Grid Point 2 is immediately south of Stn 1, ..., and Grid Point 16 is the southeasternmost grid point (southeast of Stn 27)

2.2. Historical data. Historical atmospheric data were obtained from the Commonwealth Bureau of Meteorology, Australia, on a Lambert conformal grid of 11:00 h GMT values for the period from 1978 to 1992. The 11:00 h GMT records were chosen on the basis that they are close to the mid-point of the daily precipitation recording period for SWA (24 h total precipitation recorded at 09:00 h local time, 01:00 h GMT). A 10 yr period (1978 to 1987) is used in this study to match the length of the DARLAM simulation. The available variables included MSLP and, at 850 and 500 hPa levels, the GPH, air temperature, dew point temperature (T_d), and U and V wind components. These data were interpolated to the same grid used for the DARLAM data (Fig. 1). Bates et al. (1998, Table 1) derived 24 variables from this data set. These included the raw variables listed above, the north-south and east-west gradients of the raw variables, and lagged raw variables. Given the limited number of atmospheric variables available, we use the dew point temperature depression at 850 hPa (DT_d^{850}) herein as a moisture dependent predictor variable:

$$DT_d^{850} = T^{850} - T_d^{850} \quad (1)$$

where T^{850} and T_d^{850} denote the air temperature (K) and dew point temperature (K) at 850 hPa, respectively. DT_d^{850} is a measure of how close the atmosphere at 850 hPa is to saturation with moisture.

Daily precipitation data for 30 stations were obtained from the same source for the same 10 yr period. The locations of the stations are shown in Fig. 1. These stations were chosen as they had no missing records over this period. The range in mean annual precipitation for the 30 stations is 309 (Stn 20) to 1280 mm (Stn 12), and the range in elevation is 2 (Stn 7) to 353 m (Stn 17). These ranges are representative of conditions in SWA. There is a 300 m escarpment arising from the coastal plain, along the line of Stns 5, 4, 12, 15, 11 and 14, with rain shadow effects eastwards.

3. DOWNSCALING MODEL

We use the NHMM of Hughes et al. (1999) for downscaling. 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 & Juang 1986). In our case, the NHMM defines stochastic conditional relationships that link daily multi-site precipitation occurrence patterns (the observed outcomes) to a discrete set of weather states (the hidden process). A first-order Markov chain defines the transition probabilities from weather state to weather state; these probabilities are conditional on a set of atmospheric circula-

tion predictors. Thus the NHMM is defined by a state transition probability matrix (**A**) and a precipitation occurrence probability distribution (**B**):

$$A_{ij}(\mathbf{X}^t) = P(S^t = j \,|\, S^{t-1} = i, \mathbf{X}^t) \quad (t = 1, \ldots, T;\ i, j = 1, \ldots, N) \tag{2}$$

$$B_{jl} = P(\mathbf{R}^t = l \,|\, S^t = j) \tag{3}$$

where $S^t = j$ denotes the jth (unobserved or hidden) weather state at time t, $\mathbf{R}^t = l$ denotes the rain state (precipitation occurrence, using a wet day threshold of 0.3 mm) at the M stations at time t, and $\mathbf{X}^t$ denotes the derived atmospheric data at time t. The NHMM is termed 'nonhomogeneous' as **A** depends on a set of observed covariates (**X**), and is completely specified once N and **X** are selected. Thus the compact notation N~*i.j.k.* may be used to represent an N-state NHMM with observed covariates *i,j,k*.. The NHMM reverts to a standard HMM if no atmospheric data is specified (i.e. no $\mathbf{X}^t$ in Eq. 2).

An overview of the fitting methodology follows. A full description of the fitting of the NHMM to the historical atmospheric and precipitation data for the 30 stations in SWA can be found in Charles et al. (1996) and a split-sample test is presented in Hughes et al. (1999). Classification trees and other exploratory analyses were used to determine which derived 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 $\lambda = (\mathbf{A}, \mathbf{B})$ for a series of models with different numbers of weather states and different sets of derived atmospheric variables. The estimation problem is to maximise $P(\mathbf{R}|\mathbf{X}, \lambda)$ where $\mathbf{R} = (\mathbf{R}^1, \mathbf{R}^2, ..., \mathbf{R}^T)$ and $\mathbf{X} = (\mathbf{X}^1, \mathbf{X}^2, ..., \mathbf{X}^T)$. Estimation is computationally intensive; the model fitting times range from less than a day to more than 4 wk (on a Sun Ultra workstation) depending on the number of parameters involved. Methodological issues involved in the maximum likelihood estimation are described in detail by Hughes et al. (1999).

The resulting model fits were evaluated using the Bayes information criterion (BIC), an approximation to the Bayes factor (Kass & Raftery 1995) given by $-2L + k\log(T)$, where L is the log-likelihood, k is the number of model parameters, and T is the number of days of data. This criterion is a useful guide for comparing models with different combinations of atmospheric variables and numbers of parameters. The objective is to select an NHMM that minimises the BIC, thus identifying a relatively parsimonious model that fits the data well, whilst (1) producing distinct and realistic weather state patterns, and (2) including the greatest amount of atmospheric information. The weather state sequence for the selected model is obtained using the Viterbi algorithm to assign each day to its respective state (Forney 1978, Hughes et al. 1999). Hughes & Guttorp (1994) describe the generation of precipitation occurrence sequences from the selected model conditional on an observed or modelled sequence of atmospheric variables.

Box-plots of the $1 \times CO_2$ DARLAM atmospheric data (hereafter designated by LAM1) and the historical data indicated that their spreads (defined by the difference of the upper and lower quartiles) and their extremes were in reasonable agreement. For the experiments presented herein, the LAM1 and historical data are centred using their respective means, and the $2 \times CO_2$ DARLAM atmospheric data (hereafter designated by LAM2) are centred using the means of the LAM1 data. Thus any biases in the LAM1 means are removed before downscaling, while the $2 \times CO_2$ signal ($1 \times CO_2$ to $2 \times CO_2$ mean shift) is retained. We prefer this method to that of Wilby et al. (1998). They re-scaled simulated current-climate airflow data from the HadCM2 'SUL' experiment (a Hadley Centre coupled ocean-atmosphere GCM run with combined CO_2 and sulphate aerosol forcing) to correspond with their historical data, and applied the same re-scaling to the HadCM2 'SUL' data for projected future conditions. The re-scaling consisted of an empirical quadratic transformation of the modelled airflow data to enforce agreement between the modelled and historical airflow percentiles. Such a re-scaling may effect the internal consistency of the relationships between the atmospheric and precipitation data, as there is no *a priori* way of assessing whether the scaling function derived from the $1 \times CO_2$ data applies under $2 \times CO_2$ conditions. As the NHMM relies completely upon these relationships, we are confident in keeping the centred DARLAM data as they stand.

4. APPROACH

A climate change downscaling experiment, using the NHMM, was undertaken according to the following steps:

(1) Fit NHMMs to the LAM1 data, with the LAM1 atmospheric predictor variables interpolated to the 3.75° longitude by 2.25° latitude grid. The LAM1 precipitation occurrence data were kept on the finer LAM grid. Use both 'circulation-only' and 'circulation+moisture' sets of predictor variables. For each case, select the model providing distinct weather states and a reasonable compromise between the number of predictor variables used and the optimum BIC.

(2) Drive the selected NHMMs fitted to the LAM1 data ('circulation-only' and 'circulation+moisture' cases) with the corresponding LAM2 atmospheric pre-

dictor variables (also interpolated to the 3.75° longitude by 2.25° latitude grid) to generate downscaled predictions of the $2 \times CO_2$ precipitation occurrence series on the LAM grid.

(3) Compare these downscaled $2 \times CO_2$ precipitation occurrence series with the actual LAM2 precipitation occurrence series. A favourable comparison gives us confidence that the selected NHMMs capture the modelled climate change signal.

(4) Independent of Steps 1 to 3, fit NHMMs that downscale the historical atmospheric predictor variables (on the same 3.75° longitude by 2.25° latitude grid) to the historical 30 station precipitation occurrence data, again using 'circulation-only' and 'circulation+moisture' predictor variables. Compare these selected NHMMs to those selected in Steps 1 to 3, in terms of the number of weather states and the predictor variables used.

(5) If the degree of agreement in Step 4 is strong, use the corresponding LAM2 atmospheric predictor variables (again, on the same 3.75° longitude by 2.25° latitude grid) to drive the 'historically-fitted' NHMMs selected in Step 4. This provides downscaled $2 \times CO_2$ precipitation occurrence series for the 30 station network.

Steps 1 to 3 encompass the validation. We define *validity* here as the ability of a NHMM to reproduce the first-order statistics of independent data (i.e. the precipitation occurrence probabilities of the LAM2 precipitation series, see Step 3 above). This presumes that DARLAM provides reliable information about projected changes in rainfall at a length scale of 125 km. We cannot validate the results from Step 5, as there are no station data for $2 \times CO_2$ conditions. We will return to this point in our discussion.

Secondary validation measures are the fits to the distributions of wet and dry spell lengths. Let y denote wet or dry spell length in days. We compared the empirical cumulative distribution functions (cdfs) of y for the 'circulation-only' and 'circulation+moisture' NHMMs with that for LAM2 using 2 criteria:

$$E_{1,m} = \sum_{i=1}^{n} |F_Y^m(i) - F_Y(i)| / n\sigma_{F(i)} \quad (4)$$

and

$$E_{2,m} = \max_i \left[|F_Y^m(i) - F_Y(i)| / n\sigma_{F(i)} \right] \quad (5)$$

where i denotes unique wet or dry spell lengths in the LAM2 data, n denotes the maximum spell length, $F_y^m(i)$ denotes the probability of $y \le i$ for either NHMM, $F_y(i)$ denotes the probability of $y \le i$ for the LAM2 data, and $\sigma_{F(i)}$ denotes the standard error of $F_y(i)$. Thus Eqs. (4) & (5) are measures of goodness-of-fit in that they define weighted mean and weighted maximum absolute differences between the NHMM and LAM2 spell length probabilities, respectively. However, they are less important as validation criteria than the precipitation occurrence probabilities because the tails of the LAM1 distributions for present day conditions are subject to a high degree of sampling uncertainty (see Bates et al. 1998). These measures are chosen in preference to the more classical measures, such as the Kolmogorov-Smirnov test, as the standard significance levels used in such tests are not applicable to the NHMM.

5. RESULTS

5.1. DARLAM downscaling

5.1.1. NHMM selection

Table 1 reports the results of fitting 6-state NHMMs to the 10 yr of LAM1 atmospheric and precipitation occurrence data. Results for other models are not shown as 7-state NHMMs contain a redundant weather state, and NHMMs with $N \le 5$ have noticeably higher BIC values. For the 'circulation-only' case, the predictors for the selected NHMM are mean MSLP, north-south MSLP gradient and 850 hPa east-west GPH gradient (BIC = 19 546). Although this model gives the same BIC value as the model without the 850 hPa east-west GPH gradient predictor, we choose the former based on our objective of including the greatest amount of atmospheric information in a parsimonious model. For the 'circulation+moisture' case, the best predictors are mean MSLP, north-south MSLP gradient, and DT_d^{850}

Table 1. Nonhomogeneous hidden Markov model fitting results for 6-state models fitted to limited area model $1 \times CO_2$ atmospheric and precipitation data ('winter'). Atmospheric variables: 1 = mean MSLP; 3 = T_d^{850}; 4 = north-south MSLP gradient; 6 = north-south 500 hPa GPH gradient; 8 = east-west 850 hPa GPH gradient; 25 = DT_d^{850}

Atmospheric variables	Number of parameters	Bayesian information criteria (BIC)
(a) Circulation variables only		
–	126	20313
1	156	19839
4	156	19819
1, 4	186	19546
1, 4, 6	216	19696
1, 4, 8	216	19546
(b) Circulation+moisture variables		
3	156	19967
1, 3	186	19736
3, 4	186	19500
1, 3, 4	216	19518
1, 4, 25	216	19078
3, 4, 25	216	19246

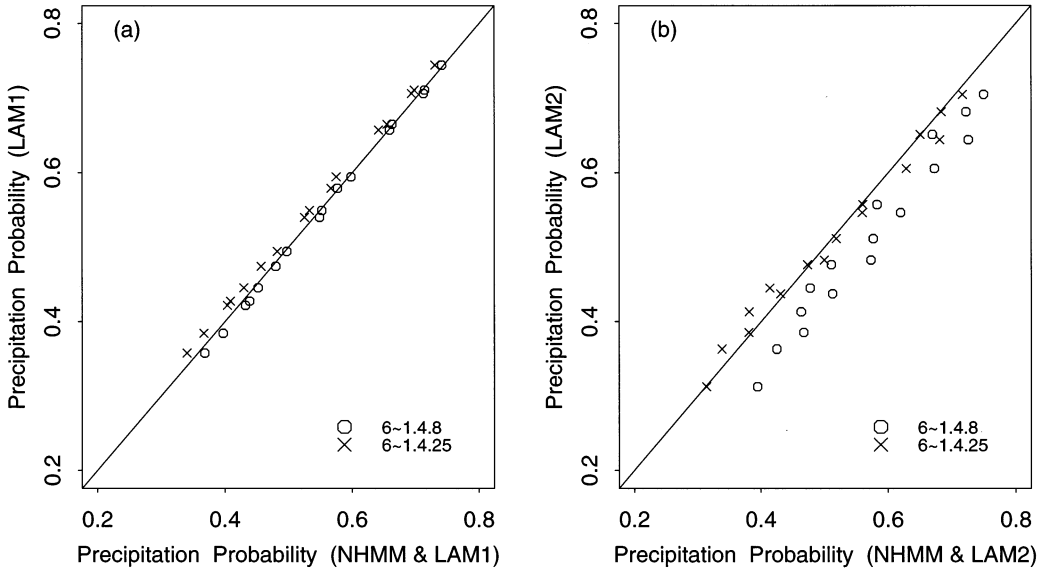


Fig. 2. Comparison of DARLAM and downscaled DARLAM 'winter' precipitation probabilities for DARLAM grid points over SWA for (a) $1 \times CO_2$ conditions and (b) $2 \times CO_2$ conditions

(BIC = 19 078). Thus the selected NHMM for the 'circulation+moisture' case provides a better fit to the DARLAM data.

Fig. 2a compares the downscaled precipitation probabilities for the LAM1 'circulation-only' and 'circulation+moisture' cases with those from the LAM1 simulation. Both NHMMs provide credible reproductions of the LAM1 probabilities, with the 'circulation-only' NHMM providing a marginally better fit and the 'circulation+moisture' NHMM a small negative bias.

5.1.2. Downscaling for $2 \times CO_2$ conditions

The LAM2 precipitation probabilities are slightly smaller than those for the LAM1 simulation. The range of the percent relative differences between LAM2 and LAM1 probabilities is −12.6 to −0.12 and the median percent relative difference is −6.22. Generally, the reductions in precipitation probability increase with increasing longitude.

Fig. 2b compares the LAM2 downscaled precipitation probabilities (obtained by driving the NHMM fitted to the LAM1 data with atmospheric data from the LAM2 run) for the 'circulation-only' and 'circulation+moisture' cases with those from the LAM2 simulation. The downscaled precipitation probabilities for the 'circulation-only' model consistently overestimate the LAM2 probabilities, while the 'circulation+moisture' model gives a credible reproduction of the precipitation probabilities at all 16 grid points. Moreover, the mean error (over the 16 grid points) in the precipitation probabilities predicted by the 'circulation-only' NHMM is over 4 times that predicted by the 'circulation+moisture' NHMM. Thus the 'circulation+moisture' NHMM is *validated* (in terms of the criteria set out in Section 4).

Table 2a compares the weighted mean differences (Eq. 4) and the weighted maximum absolute differences (Eq. 5), for the wet spell length distributions produced from the 'circulation-only' and 'circulation+moisture' downscaling, with the distributions obtained from the LAM2 simulation directly. The 'circulation+moisture' model gives better fits for 9 out of the 16 grid points (5, 7, 8, 10, 11, 12, 14, 15, and 16) for Eq. (4), and 19 out of the 30 precipitation stations fall within these LAM grid points (Fig. 1). For Eq. (5) the 2 models perform equally, each producing better fits at 8 of the 16 grid points. For dry spell lengths, the 'circulation-only' model gives a better fit to the weighted mean differences (Eq. 4) for 11 out of the 16 grid points whereas the 'circulation+moisture' model gives a better fit to the weighted maximum absolute differences (Eq. 5) for 11 out of the 16 grid points (Table 2b). Thus, overall, the 'circulation+moisture' model is only marginally better at reproducing the spell length distributions.

5.2. Historical downscaling

5.2.1. NHMM selection

Table 3 summarises the results of fitting the NHMM to the historical gridded atmospheric data (interpo-

Table 2. Discrepancies between downscaled and LAM2 cumulative distribution functions for (a) wet and (b) dry spell lengths. DARLAM grid is shown in Fig. 1. CO: NHMM with 'circulation-only' predictors; CM: NHMM with 'circulation+moisture' predictors

DARLAM Grid Point	Goodness-of-fit measures (Eqs. 4 & 5) $E_{1,CO}$	$E_{1,CM}$	$E_{2,CO}$	$E_{2,CM}$
(a) Wet spell lengths				
1	1.288	1.649	4.238	3.891
2	0.949	1.121	2.370	1.876
3	1.283	1.315	3.509	2.446
4	2.388	1.706	5.879	3.487
5	0.904	1.059	4.142	4.661
6	1.741	2.160	4.245	4.406
7	1.327	1.586	3.362	3.121
8	1.238	0.966	5.182	3.112
9	1.091	1.481	4.758	5.485
10	1.361	1.634	4.072	4.361
11	1.353	1.548	4.351	4.083
12	1.599	1.295	5.073	3.971
13	2.913	2.898	5.280	5.192
14	1.403	1.496	2.817	2.766
15	0.990	0.889	2.889	3.264
16	1.820	1.864	5.703	4.954
(b) Dry spell lengths				
1	0.483	2.761	2.003	5.316
2	1.469	2.565	2.728	5.300
3	1.806	2.258	5.022	6.312
4	1.136	1.986	2.067	3.243
5	1.211	1.189	2.381	4.388
6	1.102	1.805	3.003	2.821
7	2.679	1.928	4.647	4.535
8	3.919	1.753	9.301	3.807
9	0.742	1.091	3.430	4.782
10	1.594	0.727	2.935	2.634
11	2.097	0.799	4.270	4.302
12	3.944	0.756	6.770	2.506
13	1.747	2.149	7.320	7.290
14	1.666	0.698	3.340	2.950
15	2.242	0.778	3.960	4.469
16	3.476	1.398	7.530	3.819

Table 3. Nonhomogeneous hidden Markov model fitting results for 6-state models fitted to the historical 30 station southwestern Australia network. Atmospheric variables: 1 = mean MSLP; 2 = mean GPH at 500 hPa; 3 = T_d^{850}; 4 = north-south MSLP gradient; 8 = east-west 850 hPa GPH gradient; 25 = DT_d^{850}

Atmospheric variables	Number of parameters	Bayesian information criteria (BIC)
(a) Circulation variables only		
–	210	37659
1	240	36988
1, 4	270	36458
1, 2, 4	300	36521
1, 4, 8	300	36475
(b) Circulation+moisture variables		
1, 3, 4	300	36579
1, 3, 25	300	36933
1, 4, 25	300	36501

lated to the grid as described in Section 2.1) and station precipitation occurrence data. Overall, the 6~1.4 model with 'circulation-only' predictors produces the most parsimonious fit (BIC = 36 458). However, we selected the 6~1.4.8 model for the 'circulation-only' case as it has greater atmospheric information content and only a marginally higher BIC value (BIC = 36 475), and it is compatible with the NHMM selected in Section 5.1.1. The selected NHMM for the 'circulation+moisture' case (6~1.4.25) has a similar BIC value (BIC = 36 501). As the BIC criterion is itself an approximation, we conclude that 6~1.4.25 is a reasonable candidate model for the historical data. Thus there is a reasonable degree of consistency between the sets of predictor variables obtained from historical data and the LAM1 simulation (Step 4 of Section 4).

Fig. 3 shows the precipitation occurrence probability patterns and composited MSLP plots associated with

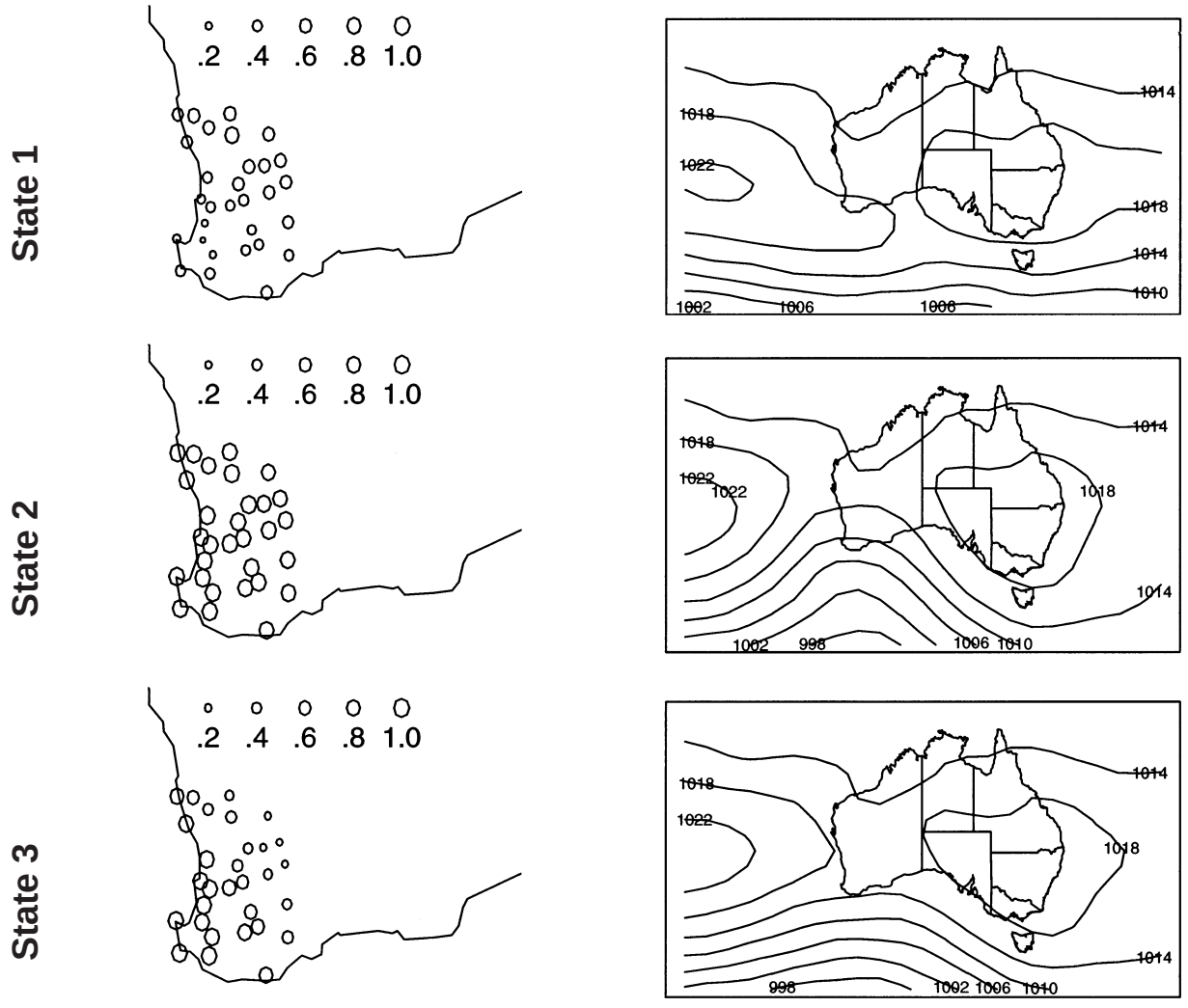


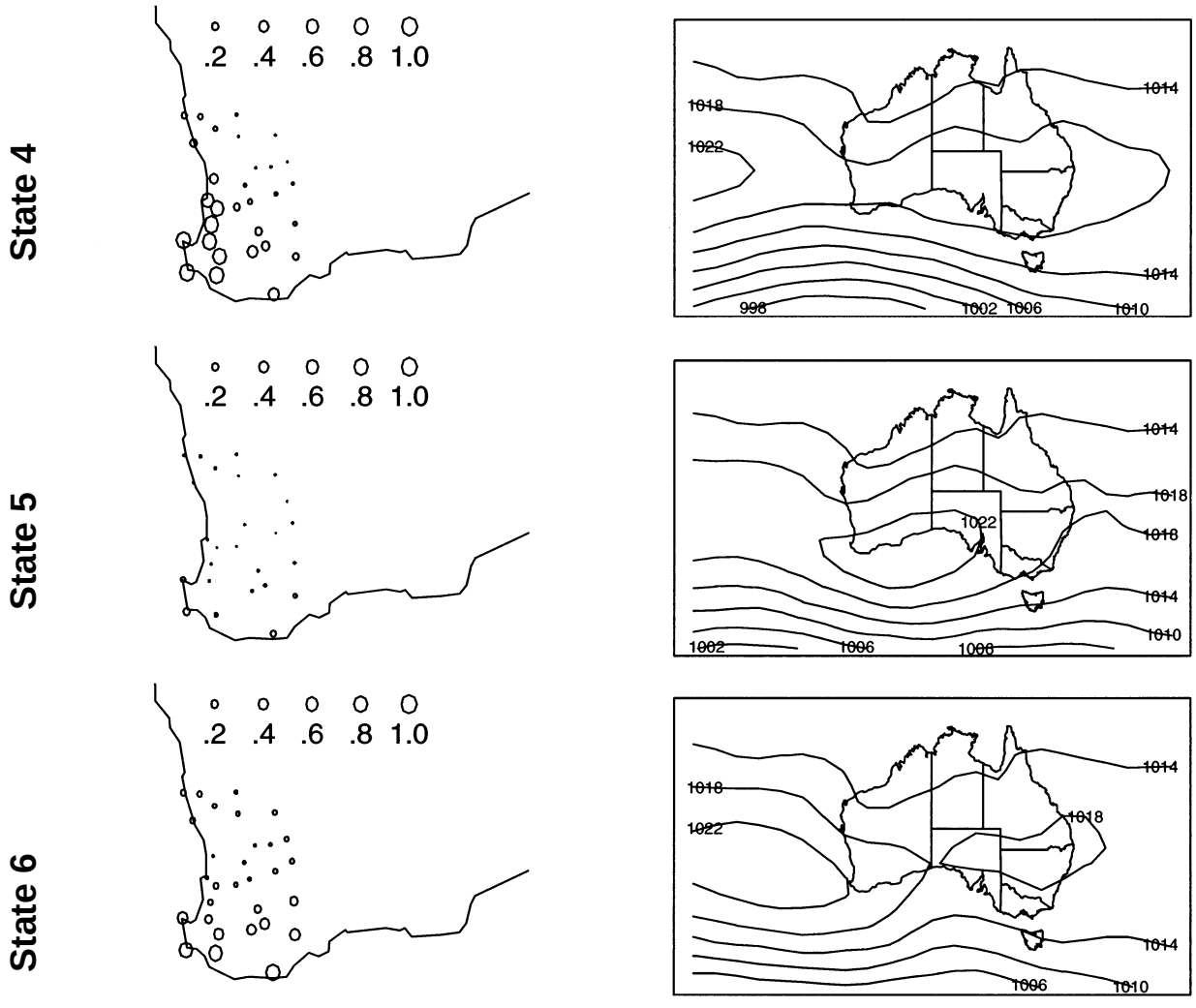
Fig. 3. (Above and facing page.) Precipitation occurrence probability patterns and composited MSLP plots associated with the 6 weather states in the 'circulation+moisture' NHMM (6~1.4.25) fitted to the historical data

the 6 weather states in the 'circulation+moisture' NHMM (6~1.4.25) fitted to the historical data. States 1 to 6 may be broadly classified as 'wet hinterland', 'wet everywhere', 'wet west coast + central', 'wet southwest corner', 'dry everywhere', and 'wet south coast', respectively. State 1 is indicative of a cut-off low with a mid-level moisture source or thunderstorms in mid- to late spring. State 2 is indicative of strong frontal systems associated with low pressure systems that ridge into SWA from the Southern Ocean. State 3 is indicative of weaker frontal systems associated with low pressure systems located in the Southern Ocean. State 4 is indicative of the presence of moist westerly winds generated by low pressure systems located well to the south of SWA. State 5 is indicative of high pressure systems located in the Great Australian Bight (southeast of SWA) that bring dry air into SWA. Finally, State 6 is indicative of the presence of moist southwest winds generated by high pressure systems centred adjacent to and to the west of the region.

Fig. 4a compares the downscaled station precipitation probabilities for the 30 precipitation stations shown in Fig. 1 with their historical values. Both the 'circulation-only' and 'circulation+moisture' NHMMs provide credible fits to the historical probabilities, with the 'circulation-only' case providing marginally better estimates for the stations with low to medium precipitation probabilities.

5.2.2. Downscaled DARLAM $2 \times CO_2$ (LAM2)

Fig. 4b shows the precipitation probabilities obtained when the fitted NHMMs are used to downscale the LAM2 atmospheric data to the 30 precipitation stations. The 'circulation-only' NHMM predicts slightly



higher precipitation probabilities across the region, whereas the validated 'circulation+moisture' NHMM predicts a slight decrease. Thus the 'circulation+moisture' NHMM produces a climate change signal that is consistent with the trend indicated by the DARLAM simulations (Section 5.1.2). While this result adds to our confidence in the 'circulation+moisture' NHMM, it is based on the underlying assumption that the LAM correctly simulates the climate change process.

Fig. 5 shows the possible changes to the persistence of wet and dry spell lengths under $2 \times CO_2$ conditions indicated by the 'circulation+moisture' NHMM at 4 of the 30 precipitation stations. The elevations and rainfall at the selected stations are representative of conditions in SWA (see Table 4). The plots suggest little change in the short-duration spells that encompass 90 % of wet and dry events. There is an overall but very slight increase in the occurrence of long-duration wet and dry spells. However, the uncertainty in the frequencies of the long-duration spells is largely due to their relatively small sample sizes within both the historical and downscaled DARLAM data.

Table 5 reports the weather state transition and steady-state probabilities for the historical and downscaled LAM2 data, and indicates changes that are judged to be significant (greater than 2 SE difference) or moderate (greater than 1 SE difference). The standard errors were obtained using a jackknife procedure (Efron & Tibshirani 1993). For each data set (historical and $2 \times CO_2$), each of the 10 years were successively removed from the data and the transition and steady-state probabilities were recalculated. These 10 sets of probabilities were then used to compute the jackknife variance. Note that this procedure does not involve refitting the NHMM for each deleted year, which would have been prohibitively expensive.

In Table 5, the steady-state probabilities for States 1, 3, 5, and 6 are essentially unchanged under $2 \times CO_2$ conditions. There is a marginal increase in the probability of the southwest corner of SWA being wet

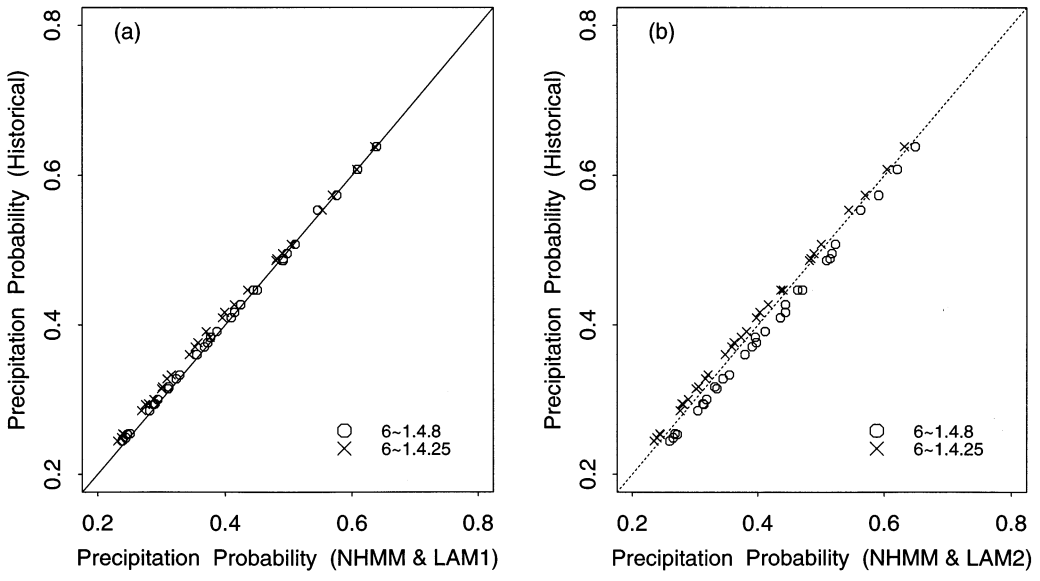


Fig. 4. Comparison of historical and downscaled winter precipitation probabilities for 30 stations in SWA for (a) historical and (b) $2 \times CO_2$ conditions

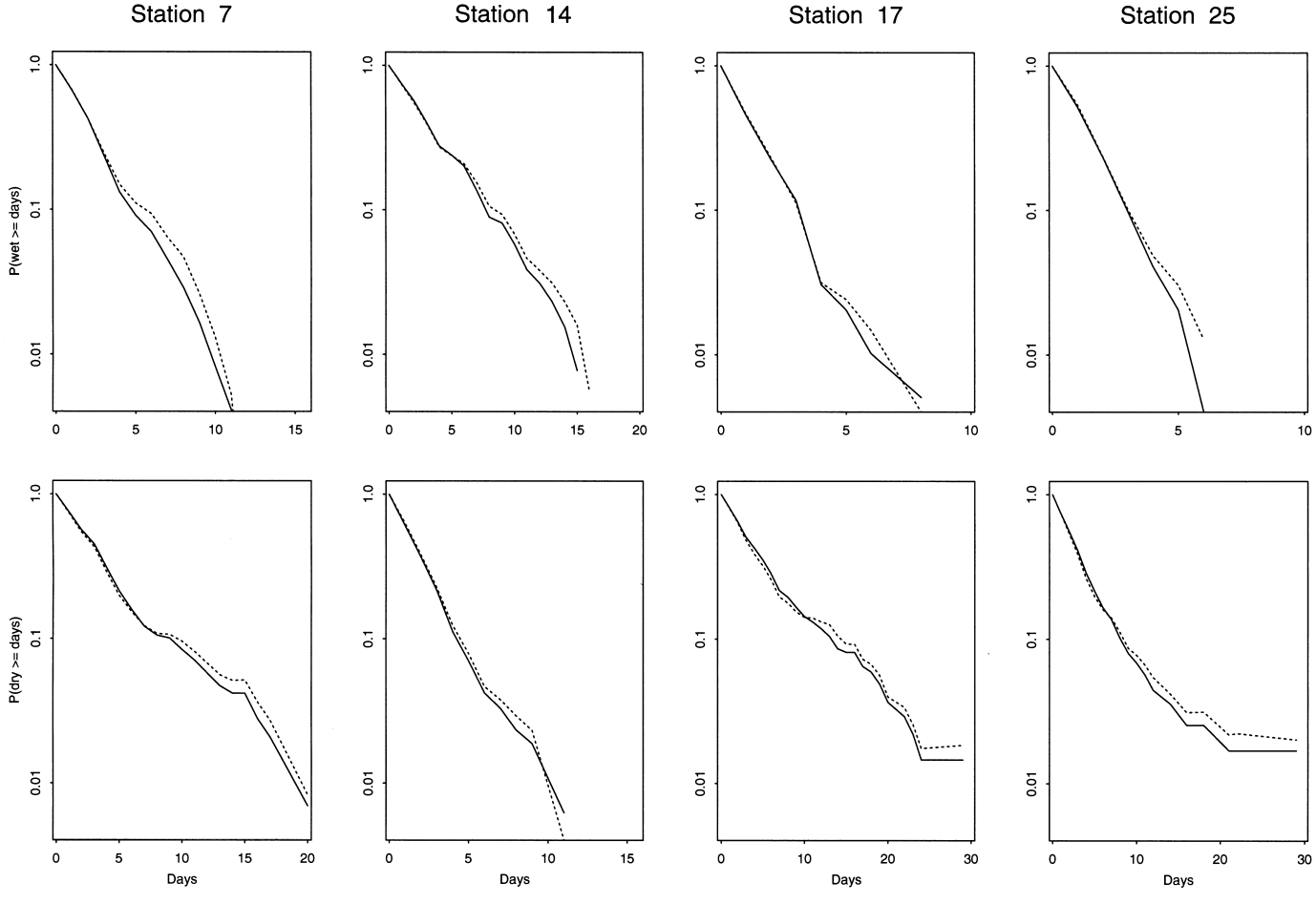


Fig. 5. Frequency characteristics of wet and dry spell lengths under historical (solid lines) and projected future [historical + (LAM2–LAM1)] conditions (dashed lines) for 4 selected stations

Table 4. Climatological summaries for the 4 selected stations shown in Fig. 5 that are representative of conditions in southwestern Australia. Winter is defined here as the May–October half-year

Stn	Latitude (°S), longitude (°E)	Elevation (m)	Winter rainfall (mm)	% of total annual rainfall	Winter number of raindays	% of total annual raindays
7	30.30, 115.04	2	466	83	80	77
14	34.45, 116.04	174	956	79	115	68
17	30.81, 117.86	353	211	66	54	74
25	33.10, 118.46	286	244	69	65	74

Table 5. Weather state (a) transition probabilities and (b) steady-state probabilities for 'circulation+moisture' NHMM (6~1.4.25) Figures in parentheses are downscaled probabilities for $2 \times CO_2$ conditions. *Changes greater than 1 SE; **Changes greater than 2 SE

State	1	2	3	4	5	6
(a) Transition probabilities						
1	0.202	0.139	0.144	0.076	0.226	0.213
	(0.187)	(0.113)	(0.157)	(0.059)	(0.269)*	(0.215)
2	0.069	0.427	0.223	0.090	0.030	0.161
	(0.082)	(0.476)*	(0.234)	(0.072)*	(0.011)**	(0.125)*
3	0.052	0.338	0.208	0.109	0.126	0.167
	(0.065)*	(0.361)	(0.189)	(0.133)*	(0.082)**	(0.170)
4	0.023	0.210	0.214	0.247	0.205	0.101
	(0.021)	(0.158)*	(0.230)	(0.265)	(0.194)	(0.132)*
5	0.052	0.081	0.066	0.119	0.622	0.060
	(0.051)	(0.056)**	(0.051)*	(0.144)*	(0.623)	(0.075)*
6	0.031	0.045	0.101	0.209	0.432	0.182
	(0.024)	(0.030)*	(0.112)	(0.204)	(0.456)	(0.174)
(b) Steady-state probabilities						
	0.058	0.209	0.150	0.138	0.318	0.127
	(0.059)	(0.199)	(0.148)	(0.149)	(0.317)	(0.128)

(State 4) and a corresponding decrease in the probability of the entire region being wet (State 2). This suggests a small decrease in the northward spread of frontal rain.

There are only a few significant changes in the transition probabilities. Consider the transitions from State 2 ('wet everywhere'). There is a significant decrease in the probability of entering State 5 ('dry everywhere') and moderate decreases in the probabilities of entering State 4 ('wet southwest corner') and State 6 ('wet south coast'). Correspondingly, there is a moderate increase in the persistence of State 2. This may indicate a decrease in the frequency of post-frontal showers. Upon leaving State 1 there is a moderate increase in the probability of entering State 5 ('dry everywhere') suggesting a reduction in the persistence of rainfall over the hinterland. State 3 ('wet west coast + central') is moderately more likely to preceed State 1 ('wet hinterland') and State 4 ('wet southwest corner') with a significant decrease in the probability of entering State 5 ('dry everywhere'). The significant decrease in the probability of leaving State 5 and entering State 2, and the moderate decrease in the probability of entering State 3 ('wet west coast + central'), suggests a decrease in frontal strength. This is supported by the corresponding moderate increases in the probabilities of leaving State 5 and entering States 4 and 6. The moderate decrease in the probability of leaving State 4 ('wet southwest corner') and entering State 2, and the increase in the probability of entering State 6 ('wet south coast'), indicates an increase in the tendency for rainfall in the southwest corner of SWA to contract to the south coast. This is also consistent with the suggestion of weaker frontal processes. Nevertheless, it must be emphasised that these projected changes are often for small probabilities that are based on small sample sizes (10 yr LAM simulation) and thus they may confound climate change and interdecadal climate variability.

6. DISCUSSION

A fundamental caveat of downscaling methods is the presumption that the relationships between atmospheric circulation and precipitation are maintained for changed climate conditions. Many studies (e.g. Matyasovszky et al. 1993a,b, Kidson & Watterson 1995, Wilby & Wigley 1997) have shown that downscaling models with 'circulation-only' predictors are largely insensitive to changed climate conditions or do not reproduce the broad-scale changes indicated by GCM or LAM precipitation fields. Recently, Busuioc et al. (1999) attempted to verify a downscaling scheme for Romania, based on a canonical correlation analysis (CCA) applied to monthly station precipitation and European-scale sea-level pressure. Their verification determined that (1) the CCA for data from a control GCM run produced similar results to the CCA for the historical data (i.e. verifying the performance of the GCM), and (2) that the downscaling relationship derived from the historical data, when driven by the 2 $\times CO_2$ GCM data, produced results consistent with the $2 \times CO_2$ GCM precipitation fields. Busuioc et al. (1999) conclude that their results are not proof that the downscaling model can be used under $2 \times CO_2$ conditions, but they add confidence that the empirical relationships derived from the CCA remain valid under the changed climate. They also note that most previous downscaling studies have not considered the veracity of the assumption that their downscaling relationships hold for climate change conditions.

In our study, the incorporation of a measure of relative moisture (DT_d^{850}) as a predictor led to a downscaling model that could better reproduce the daily precipitation occurrence probabilities of the LAM2 simulation. Thus the use of a measure of atmospheric moisture content, as suggested by Pittock (1993) and Wilby & Wigley (1997), has produced results that can be validated against LAM data directly. In assuming that the LAM simulation of circulation and precipitation is consistent with a possible future climate, we have been able to use a dynamical model to validate a downscaling technique. We believe this is preferable to methods that have not considered criteria to assess the validity of their changed climate predictions. As the selected predictors are consistent with the physically plausible LAM2 simulation, we have confidence in the results obtained when an NHMM using these predictors is used to downscale the LAM2 simulation to the 30 station network. We obviously cannot independently validate the predictions for the 30 station network under the $2 \times CO_2$ conditions.

Crane & Hewitson (1998) predicted large precipitation increases (32% for spring and summer in the United States) for $2 \times CO_2$ conditions when their artificial neural network was driven by $2 \times CO_2$ GCM GPH and specific humidity data. In our study an NHMM including T_d^{850} rather than DT_d^{850}, fitted to the LAM1 data, overestimates the LAM2 grid precipitation probabilities. The range of the percent relative differences between this NHMM and the LAM2 probabilities is 6.30 to 59.2 and the median percent relative difference is 27.6. Thus, based on the criteria set out in Section 4, a model using dew point temperature rather than dew point temperature depression could not be validated.

Specific humidity and dew point temperature are a measure of and a proxy for the absolute moisture content of the atmosphere, respectively. It can be argued that the probability of precipitation occurrence would be better related to a measure of saturation (such as DT_d^{850}), and so probability of cloud formation, rather than absolute moisture content. Comparison of the LAM1 and LAM2 values for T_d^{850} and DT_d^{850} reveals that the 25, 50 and 75% quantiles for T_d^{850} increase by 4.1, 3.5 and 3.2 K whereas the 25, 50 and 75% quantiles for DT_d^{850} increase by 0.7, 0.8 and 1.0 K under doubled CO_2. Thus we suggest that the use of T_d^{850} produced an implausibly large climate change signal in terms of precipitation occurrence. The use of a measure of absolute moisture does not account for the compensatory effect caused by the increase in temperature under a changed climate (a rise in temperature would increase the moisture holding capacity of the atmosphere). We therefore caution against the use of predictors based on absolute rather than relative moisture content, and more generally we caution against the use of downscaling methods that do not attempt to ascertain the legitimacy of their results for the changed climate conditions.

Finally, the NHMM can be placed against the criteria of Wilby et al. (1998) for a useful downscaling model:

(1) *As far as possible the scheme should be globally applicable*. To date the NHMM has only been applied to regions in which the dominant precipitation producing systems are synoptic (northwestern USA and SWA). Further work is required to assess its applicability to regions in which precipitation is predominantly convective. If sufficient atmospheric data are available to capture the 'signal' of the convective atmospheric processes, there is no theoretical reason why the NHMM should not perform well. Conversely, if the historical synoptic variables do not provide enough information to allow satisfactory downscaling then any GCM-based method is likely to fail.

(2) *The predictor variables used to downscale should be both physically and conceptually sensible*. The predictor variables used herein are interpretable in terms of the known circulation patterns and synoptic-scale precipitation generation mechanisms of the region (see Hughes et al. 1999).

(3) *The predictors should ideally be continuous variables in order to model extreme events.* All of the predictors used herein are continuous and the downscaled data match low probability dry and wet spells reasonably well. Nevertheless, the NHMM can work with either continuous or discrete data or both simultaneously. However, less confidence can be placed in findings regarding extremes as they are model dependent.

(4) *The predictor variables should be accurately modelled and readily available from GCM output.* Our choice of predictors was based on the availability of historical data. We verify that the GCM or LAM adequately simulates the distributions of the historical data using box-plots.

(5) *The predictor variables should be strongly correlated with a suite of relevant hydrometeorological variables.* The atmospheric predictors in the NHMM are selected on the basis of the strength of their relationship with multi-station precipitation occurrence.

(6) *The predictors should preserve observable correlations between downscaled parameters.* The NHMM preserves precipitation occurrence correlations among a network of stations (Hughes et al. 1999). Extension of the NHMM to include rainfall amounts (Charles et al. 1999) and surface air temperature is on-going.

(7) *Time series of the predictor variables should be responsive to greenhouse gas forcing.* We have demonstrated that the addition of a predictor that reflects relative atmospheric moisture content improves the sensitivity of downscaling models to greenhouse gas forcing. Moreover, the inclusion of the predictor was justified in terms of a statistical model selection criterion (BIC) rather than professional judgement or its observed sensitivity to greenhouse gas forcing alone.

7. CONCLUSIONS

We followed the suggestion by Pittock (1993) and Wilby & Wigley (1997) that downscaling models should use predictor variables that characterise atmospheric moisture content as well as circulation. We investigated the application and validation of the nonhomogeneous hidden Markov model (NHMM) of Hughes et al. (1999) using LAM and historical data for southwestern Australia. Our results show that the NHMM provides a credible downscaling technique for assessing climate change in this region provided a variable characterising the closeness of the atmosphere to saturation rather than absolute moisture content is included in the predictor set. The inclusion of 850 hPa dew point temperature depression was found to be defensible when the NHMM was fit to either LAM1 or historical data. When driven by the LAM2 atmospheric fields, the NHMM fitted to the LAM1 data could reproduce the gridded precipitation occurrence statistics of the LAM2 simulation. This indicates that the selected NHMM is valid. Driving the NHMM fitted to historical data with the LAM2 atmospheric fields results in a small decrease in the probability of precipitation across the 30 stations in southwestern Australia. This result is based on only 1 LAM simulation and is only relevant to the case study region. It should not be extrapolated to other mid-latitude regions of the world.

Our results highlight that the validation of a statistical downscaling technique using historical data does not necessarily imply the model will produce legitimate results for changed climate conditions. This implies that many published findings about downscaling derived climate change impacts should be viewed with caution. Finally, it is important to stress that our results apply only to precipitation occurrence and not necessarily to precipitation amounts.

Acknowledgements. We thank John McGregor and Jack Katzfey (CSIRO Atmospheric Research) for providing the DARLAM data series used in this investigation. Jack Katzfey kindly helped with code development and prompt advice during DARLAM data extraction. Historical atmospheric and precipitation data was provided by the National Climate Centre, Australian Bureau of Meteorology. This work contributes to the CSIRO Climate Change Research Program and is part funded through the Australian Government's National Greenhouse Research Program. S.P.C. and B.C.B. are grateful for their Visiting Scholarships held at the National Research Center for Statistics and the Environment, University of Washington, during which some of the work described herein was carried out. J.P.H. was supported by National Science Foundation grant number DMS-9524770. We also appreciate the constructive comments made by the anonymous reviewers.

LITERATURE CITED

Bardossy A, Plate EJ (1991) Modelling daily rainfall using a semi-Markov representation of circulation pattern occurrence. J Hydrol 122:33–47

Bardossy A, Plate EJ (1992) Space-time model for daily rainfall using atmospheric circulation patterns. Water Resour Res 28(5):1247–1259

Bartholy J, Bogardi I, Matyasovszky I (1995) Effect of climate change on regional precipitation in Lake Balaton watershed. Theor Appl Climatol 51:237–250

Bates BC, Charles SP, Hughes JP (1998) Stochastic downscaling of numerical climate model simulations. Environ Model Soft 13(3-4):325–331

Bogardi I, Matyasovszky I, Bardossy A, Duckstein L (1993) Application of a space-time stochastic model for daily precipitation using atmospheric circulation patterns. J Geophys Res 98(D9):16653–16667

Busuioc A, von Storch H, Schnur R (1999) Verification of GCM-generated regional seasonal precipitation for current climate and of statistical downscaling estimates under changing climate conditions. J Clim 12(1):258–272

Charles SP, Hughes JP, Bates BC, Lyons TJ (1996) Assessing

downscaling models for atmospheric circulation—local precipitation linkage. In: Preprints International Conference on Water Resources and Environmental Research: towards the 21st century. Water Resources Research Center, Kyoto University, p 269–276

Charles SP, Bates BC, Hughes JP (1999) A spatio-temporal model for downscaling precipitation occurrence and amounts. J Geophys Res (in press)

Conway D, Wilby RL, Jones PD (1996) Precipitation and airflow indices over the British Isles. Clim Res 7:169–183

Crane RG, Hewitson BC (1998) Doubled CO_2 precipitation changes for the Susquehanna basin: down-scaling from the GENESIS general circulation model. Int J Climatol 18: 65–76

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman and Hall, New York

Forney GD Jr (1978) The Viterbi algorithm. Proc IEEE 61: 268–278

Gates WL, Henderson-Sellers A, Boer GJ, Folland CK, Kitoh A, McAvaney BJ, Semazzi F, Smith N, Weaver AJ, Zeng QC (1996) Climate models—evaluation. In: Houghton JT, Meira Filho LG, Callander BA, Harris N, Kattenberg A, Maskell K (eds) Climate change 1995, the science of climate change. Cambridge University Press, Cambridge, p 229–284

Giorgi F, Mearns LO (1991) Approaches to the simulation of regional climate change: a review. Rev Geophys 29: 191–216

Hay LE, McCabe GJ, Wolock DM, Ayers MA (1991) Simulation of precipitation by weather type analysis. Water Resour Res 27(4):493–501

Hewitson BC, Crane RG (1996) Climate downscaling: techniques and application. Clim Res 7:85–95

Houghton JT, Meira Filho LG, Callander BA, Harris N, Kattenberg A, Maskell K (1996) Technical summary. In: Houghton JT, Meira Filho LG, Callander BA, Harris N, Kattenberg A, Maskell K (eds) Climate change 1995, the science of climate change. Cambridge University Press, Cambridge, p 9–50

Hughes JP, Guttorp P (1994) Incorporating spatial dependence and atmospheric data in a model of precipitation. J Appl Meteorol 33(12):1503–1515

Hughes JP, Guttorp P, Charles SP (1999) A nonhomogeneous hidden Markov model for precipitation occurrence. Appl Stat 48:15–30

Huth R (1997) Potential of continental-scale circulation for the determination of local daily surface variables. Theor Appl Climatol 56:165–186

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90: 773–795

Kidson JW, Watterson IG (1995) A synoptic climatological evaluation of the changes in the CSIRO nine-level model with doubled CO_2 in the New Zealand region. Int J Climatol 15:1179–1194

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

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

Matyasovszky I, Bogardi I, Duckstein L (1994) Comparison of two general circulation models to downscale temperature and precipitation under climate change. Water Resour Res 30(12):3437–3448

Mearns LO, Giorgi F, McDaniel L, Shields C (1995) Analysis of daily variability of precipitation in a nested regional climate model: comparison with observations and doubled CO_2 results. Global Planet Change 10:55–78

Pittock AB (1993) Climate scenario development. In: Jakeman AJ, Beck MB, McAleer MJ (eds) Modelling change in environmental systems. John Wiley and Sons, New York, p 481–503

Rabiner LR, Juang BH (1986) An introduction to hidden Markov models. IEEE Acoustics Speech Signal Process Mag, Jan 1986, p 4–16

Walsh KJE, McGregor JL (1995) January and July climate simulations over the Australian region using a limited-area model. J Clim 8(10):2387–2403

Walsh KJE, McGregor JL (1997) An assessment of simulations of climate variability over Australia with a limited area model. Int J Climatol 17:201–223

Watterson IG, O'Farrell SP, Dix MR (1997) Energy and water transport in climates simulated by a general circulation model that includes dynamic sea ice. J Geophys Res 102(D10):11027–11037

Wilby RL (1997) Non-stationarity in daily precipitation series: implications for GCM downscaling using atmospheric circulation indices. Int J Climatol 17:439–454

Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. Prog Phys Geogr 21(4):530–548

Wilby RL, Greenfield B, Glenny C (1994) A coupled synoptic-hydrological model for climate change impact assessment. J Hydrol 153:265–290

Wilby RL, Hassan H, Hanaki K (1998) Statistical downscaling of hydrometeorological variables using general circulation model output. J Hydrol 205:1–19

Wilson LL, Lettenmaier DP, Wood EF (1991) Simulations of daily precipitation in the Pacific Northwest using a weather classification scheme. Surv Geophys 12:127–142

Wilson LL, Lettenmaier DP, Skyllingstad E (1992) A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation. J Geophys Res 97(D3):2791–2809

Zorita E, Hughes JP, Lettenmaier DP, von Storch H (1995) Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. J Clim 8:1023–1042

Editorial responsibility: Hans von Storch, Geesthacht, Germany

Submitted: June 30, 1998; Accepted: March 19, 1999
Proofs received from author(s): May 21, 1999