Vol. 25: 191–203, 2004 | CLIMATE RESEARCH Clim Res | Published January 23

Impact of varying the complexity of the land surface energy balance on the sensitivity of the Australian climate to increasing carbon dioxide

A. J. Pitman*, B. J. McAvaney

Department of Physical Geography, Macquarie University, North Ryde, 2109 New South Wales, Australia

ABSTRACT: This study explores whether regional climate model scenarios are sensitive to uncertainty in the representation of surface energy balance (SEB) complexity. Simulations with the Bureau of Meteorology Research Centre climate model and the CHAmeleon Surface Model (CHASM) were used to explore the sensitivity of changes in air temperature (T), evaporation (E), precipitation (P) and soil moisture (W) over Australia resulting from a doubling of atmospheric carbon dioxide (ΔCO_2). The $1 \times CO_2$ and the $2 \times CO_2$ simulations of T, E, P and W were sensitive to the complexity of the SEB, even though the grand mean of these quantities was almost always insensitive to SEB complexity. Seasonal variations in T, E, P and W at $1 \times CO_2$ and $2 \times CO_2$ were sensitive in terms of the point-by-point temporal mean and temporal variance. The overall spatial and temporal variances of T and P were insensitive to SEB complexity, but E and W were sensitive during periods of drying. The simulated seasonal change in T, E, P and W was insensitive to the SEB, and uncertainty in SEB parameterisation does not limit the reliability of existing climate change scenarios for Australia. However, the temporal variance of E, P and W was sensitive to the SEB complexity during periods of drying. Use of temporal variances of these quantities in future impact assessments are therefore likely to be very limited until uncertainty in the representation of SEB in climate models is reduced. To simulate the climate over Australia at either $1 \times CO_2$ or $2 \times CO_2$, a reasonably complex representation of the SEB, including a temporally and spatially variable surface resistance and an explicit representation of canopy interception, is required. Finally, our results say nothing about the importance of the land surface in general, since our analysis is restricted to a consideration of the SEB alone.

KEY WORDS: Surface energy balance · Land surface models · Climate changes · Land surface complexity · Australian climate · Climate modelling

1. INTRODUCTION

In climate models, the land surface affects the atmosphere through the partitioning of available energy between sensible and latent heat. It is important to model this partitioning well, since more latent heat contributes water vapour to the atmosphere and tends towards increasing cloudiness and precipitation, while an increase in sensible heat tends to warm the atmosphere (see Betts et al. 1996, Sellers et al. 1997). The land surface also controls the partitioning of available water between evaporation and runoff. It is also important to model this partitioning well, since errors in the simulation of runoff lead to errors in the simulation of evaporation (Liang et al. 1998).

The modeling of the surface energy balance (SEB) is one part of modeling the land surface in general. Parameterisations of the SEB in climate models range from simple approaches based on large scale climatological observations (e.g. Manabe, 1969) to highly complex models that include radiative transfer through canopies and the incorporation of boundary layer theory to help estimate turbulent energy fluxes (e.g. Sellers et al. 1996). In the Manabe (1969) land surface model (LSM), the surface is characterized by an albedo, aerodynamic roughness, evaporative wetness and a single surface

*Email: apitman@penman.es.mq.edu.au

temperature. More complex SEB models place a greater emphasis on the explicit representation of diurnal scale processes and on the influence of vegetation on land-atmosphere interactions. The LSM community has not reached a consensus regarding the level of complexity required to represent SEB. This contributes to the uncertainty in regard to the reliability of climate change scenarios produced by climate models.

There are 2 key areas where the SEB may play an important role:

(1) The SEB may affect the partitioning of available energy, affecting the simulation of climate by a model. The role of the SEB in affecting the sensitivity of a global climate model to doubling of carbon dioxide (ΔCO_2) was explored by Pitman & McAvaney (2003), who found that a climate model was sensitive (at statistically significant levels) to the SEB complexity at $1 \times CO_2$ and $2 \times CO_2$. However, the climate model's simulation of changes caused by ΔCO_2 were insensitive to the SEB-complexity at the global scale.

(2) The SEB may play a key role in climate change scenario development. Climate change scenarios are usually developed at or below continental scales. In this study we ask 'Does the complexity of the SEB representation matter for the development of climate change scenarios over the Australian continent?' If it does, then our uncertainty in how to represent the SEB may limit projections of the continental scale impact of increasing CO_2. Conversely, if the SEB proves to be relatively unimportant within the Australian environment, this adds confidence to our ability to model greenhouse-induced climate changes and helps to concentrate attention on model development on other aspects of land surface processes (e.g. hydrology).

We chose to focus on Australia to explore the sensitivity of the results of Whetton et al. (2001), who demonstrated that different patterns of climate change, following ΔCO_2, were simulated by a regional climate model embedded in a global climate model, in comparison to the patterns simulated by the global model. They noted that different (but plausible) scenarios would be obtained if different regional models or different climate models were used. It is impossible to explore all the possible combinations of regional and climate models, or all the alternative parameterisations of processes included in these models, to obtain an overall uncertainty estimate from this ensemble of scenarios. It is, however, possible to explore the uncertainty that is derived from specific ways of representing some processes within a climate model.

This study therefore explores the sensitivity of simulations of climate change over Australia to the parameterisation of the SEB. Significant sensitivity would substantially reduce the value of any small ensemble of scenarios. Conversely, if the SEB played a minor role, then this would add confidence to the scenarios described by Whetton et al. (2001), and to any management or policy initiatives developed from these scenarios.

2. METHODOLOGY

We used the Atmospheric Model Intercomparison Project (AMIP; Gates 1992) framework, which includes prescribed sea surface temperatures (SST) and sea ice cover. For the $2 \times CO_2$ experiments, we followed Crossley et al. (2000) in using changes in SST and sea ice cover from a transient run of the Hadley Centre climate model (Mitchell et al. 1995). Average anomalies for each month were calculated from a 20 yr period around the time of CO_2 doubling and then applied to SST values from the 1979–1988 period, as defined by the AMIP project. Crossley et al. (2000) describe this experimental design in detail.

2.1. Modelling and statistical methodology

A 21 yr simulation using the most complex version of the CHAmeleon land Surface Model (CHASM; see below) was performed for $1 \times CO_2$ and $2 \times CO_2$ scenarios. Subsequent experiments using simpler modes of CHASM were run for 5 yr. Following Desborough et al. (2001), simulations with 5 modes of CHASM are compared to determine whether explicitly parameterizing temperature differentiation (for vegetated and non-vegetated surface fractions), canopy resistance, bare ground evaporation and canopy interception affect key simulated quatities. In analyzing the results, each mode is compared to the control simulation (SLAM).

Any differences between simulations may be the result of the parameterisation change or of model variability. It is therefore necessary to evaluate the statistical significance of any changes. It is not adequate to simply test a change in the mean or variance using standard statistical tests, because of the problems of multiplicity, spatial autocorrelation or other unknown sampling distributions (Santer & Wigley 1990). To address this issue, Wigley & Santer (1990) and Santer & Wigley (1990) proposed a series of 9 tests that examine the statistical significance of results using a pool-permutation procedure (Preisendorfer & Barnett 1983). We selected 5 test statistics that explore changes in the mean, and in the spatial and temporal variance. Specifically, using the same terminology as Wigley & Santer (1990) (see Appendix 1 and Table 1 for a definition of all statistical tests) we used 'T1' to test the differences in grand means. NT5 and NF5 test the differences in

temporal means and temporal variances point-by-point with a statistical significance level of 5%, and SPRET1 and SPREX1 test the overall difference in the temporal and spatial variances. In these tests, the length of the control simulation must be equivalent to that of the perturbation experiments, hence only 5 yr of the control is used. We use 1000 permutations. While 5 yr for each experiment is less than we would like (although this still required 50 yr of model integration) Santer & Wigley (1990) point out that the pool permutation method provides a means to circumvent problems of small time samples in model data. We are therefore confident that longer simulations would not alter the significance of our results.

2.2. CHAmeleon Surface Model (CHASM)

CHASM was designed by Desborough (1999) to explore the role of SEB complexity, and it can run with several SEB configurations, or modes (see Table 2, Fig. 1). Each mode is combined with a common soil moisture and soil temperature model. These are briefly discussed below.

2.3. Basic model parameterisation

To resolve SEB, CHASM combines similar elements throughout a grid square to form tiles (a 'grouped mosaic approach'; e.g. Koster & Suarez 1992). Each tile is further divided into aerial cover fractions of vegetation, snow and ground. The vegetation fraction is further divided into wet and dry fractions if the surface configuration mode allows for canopy interception. Each tile has a prognostic bulk temperature for the storage of energy and a diagnostic skin temperature for the calculation of surface energy fluxes. There are individual parameters for albedo and roughness length for each type of cover, and the overall surface albedo and roughness length is calculated from an area-weighted approach across the snow, vegetation and bare ground fractions. Soil temperature is simulated using 4 layers with a total depth of 4 m and energy transfers within the soil are represented using a finite difference method with constant values for volumetric heat capacity and thermal conductivity and a zero-flux boundary condition at the base of the profile.

CHASM's hydrology follows Manabe (1969) in that the root zone is treated as a bucket with finite water holding capacity, and beyond this capacity runoff occurs. Water can also be stored as snow, or (depending on the mode) stored on the canopy following interception of precipitation or on the surface for bare ground evaporation. While the use of a simple hydrology model may seem outdated, Robock et al. (1995) have shown it to work well in mid-latitude regions.

Each tilecan have up to 4 evaporation sources for canopy evaporation, transpiration, bare ground evaporation and snow sublimation. Resistances may be applied to reduce evaporation and transpiration rates. As these resistances and the different ways in which they are applied are a key part of the differences between modes, they are described in detail below.

Table 1. Statistical tests used in this study

Statistic	Designed to test …
T1	Differences in the grand means
NT5	Differences in time means, grid point by grid point. Gives the fraction of grid points with a statistically significant difference at a 5% significance level
NF5	Differences in temporal variances, grid point by grid point. Gives the fraction of grid points with a statistically significant difference at a 5% significance level
SPRET1	Overall difference in temporal variance
SPREX1	Overall difference in spatial variance

Table 2. Summary of the differences between the CHASM modes used in this paper; ✓ indicates that the model includes explicit parameterisations

	SLAM	SLAM1T	RSGI	RSI	RS
Canopy interception	✓	✓	✓	✓	
Bare ground evaporation	✓	✓	✓		
Canopy resistance	✓	✓			
Temperature differentiation	✓				

2.4. CHASM modes

Table 2 and Fig. 1 list the modes of CHASM used in this study, in increasing order of complexity.

Mode RS is the simplest one used in this study. The surface is constructed from one tile, so the fraction of the surface occupied by the tile, A^t, equals 1. The calculation of aerodynamic resistance to turbulent transport for heat and moisture (r_a) accounts for atmospheric stability. Moisture available for evaporation is stored in the root zone (W_r) and on the surface as snow (W_n), so there are only 2 evaporation sources, evaporation from the snow (E_n) and evaporation from the root zone (E_{tr}), to which r_a is applied (Fig. 1a), and a temporally invariant surface resistance (r_s) is added to the resistance pathway of snow-free evaporation. a_n represents the

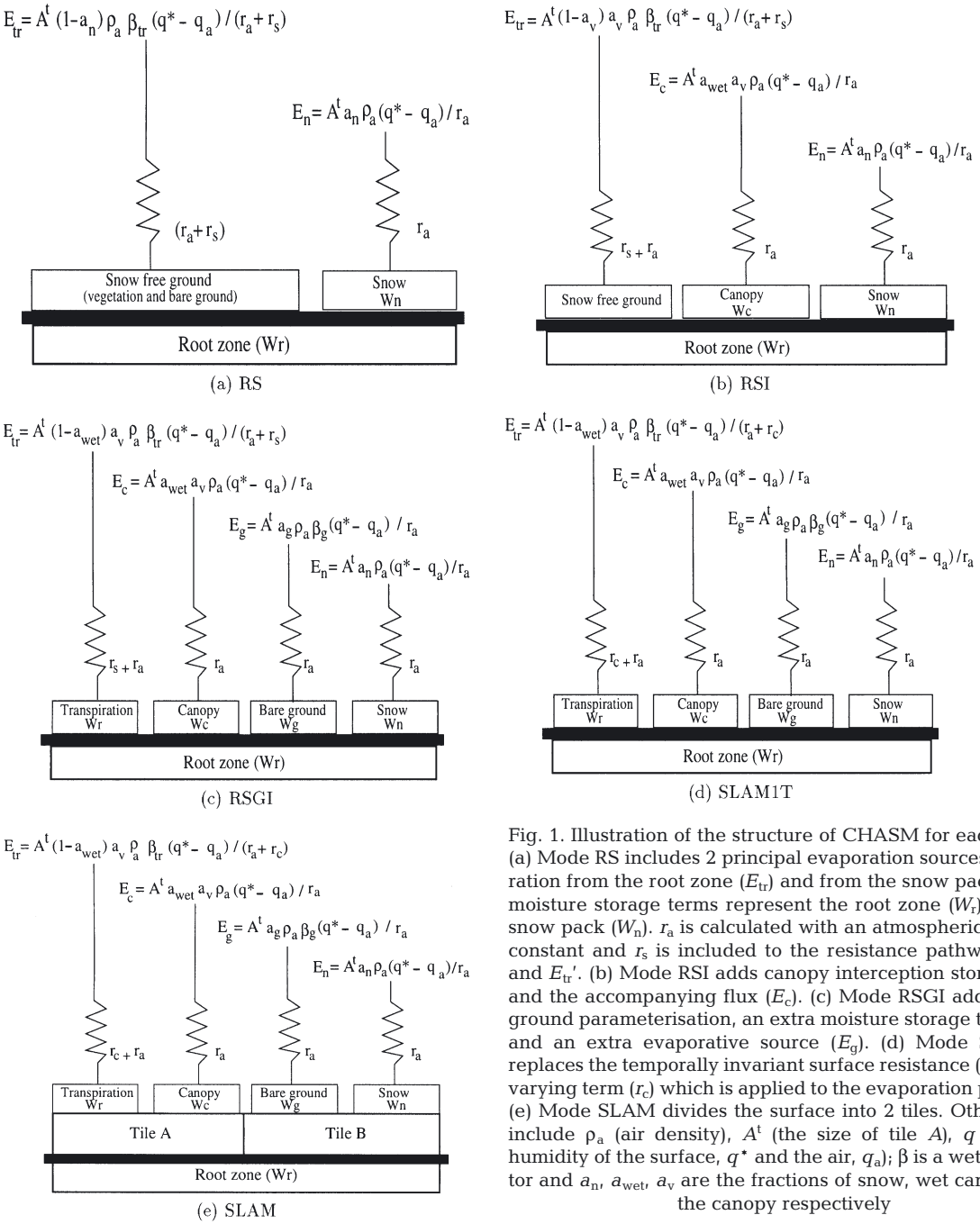

Fig. 1. Illustration of the structure of CHASM for each mode. (a) Mode RS includes 2 principal evaporation sources: evaporation from the root zone (E_{tr}) and from the snow pack (E_n); 2 moisture storage terms represent the root zone (W_r) and the snow pack (W_n). r_a is calculated with an atmospheric stability constant and r_s is included to the resistance pathway of E_{tr} and E_{tr}'. (b) Mode RSI adds canopy interception storage (W_c) and the accompanying flux (E_c). (c) Mode RSGI adds a bare ground parameterisation, an extra moisture storage term (W_g) and an extra evaporative source (E_g). (d) Mode SLAM1T replaces the temporally invariant surface resistance (r_s) with a varying term (r_c) which is applied to the evaporation pathway. (e) Mode SLAM divides the surface into 2 tiles. Other terms include ρ_a (air density), A^t (the size of tile A), q (specific humidity of the surface, q^* and the air, q_a); β is a wetness factor and a_n, a_{wet}, a_v are the fractions of snow, wet canopy and the canopy respectively

fraction of the surface covered by snow. E_{tr} is reduced below the potential rate by an additional moisture availability resistance (β_{tr}). The calculation of E_{tr} and E_n also involves the density of air (ρ_a), the surface saturated specific humidity (q^*) (calculated as a function of the skin temperature) and specific humidity of the air (q_a).

Mode RSI builds on the RS mode by adding an explicit parameterisation for canopy interception of precipitation (Fig. 1b). There are therefore 3 evaporation sources including evaporation of intercepted water (E_c), following the addition of storage of water within the canopy (W_c). The canopy is divided into fractions of wet (a_{wet}) and dry (a_{dry}) areas, which depend on the amount of precipitation and on evaporation rates.

Mode RSGI adds bare ground evaporation to the RSI mode. Moisture can be stored at the surface for evaporation up to a maximum of 40 kg m^{-2}. Bare ground

evaporation is affected by moisture availability where a resistance, β_g, is included in the evaporation pathway. β_g reduces bare ground evaporation linearly to zero as moisture availability declines to zero. The RSGI mode is illustrated in Fig. 1c and includes 4 evaporation fluxes (E_{tr}, E_c, E_n and E_g) and their corresponding moisture storage terms (W_r, W_c, W_n and W_g).

SLAM1T (Fig. 1d) builds on RSGI by including a time variable in canopy resistance, thus replacing the temporally invariant surface resistance used in RS, RSI and RSGI.

Finally, the most complex mode, SLAM (Fig. 1e), is the same as SLAM1T except that the land-atmosphere interface is divided into 2 tiles, the first representing a combination of bare ground and exposed snow and the other reserved for vegetation. The tiles are area-weighted depending on the individual fractions of the surface type. A separate SEB is calculated for each tile, and this allows for temperature variations across the land-atmosphere interface.

CHASM is a useful tool for explaining some of the large simulation differences obtained by PILPS (e.g. Desborough 1999), identifying the role of the SEB in influencing climate (Desborough et al. 2001) and identifying the role of the SEB in affecting the sensitivity of global climate to ΔCO_2 (Pitman & McAvaney 2003). The maintenance of a common modeling environment as the SEB changes ensures that parameters retain the same effective value. This allows the modeller to explore the effect of increasing the complexity of the SEB configuration by sequentially adding explicit parameterisations.

3. RESULTS

We are concerned with whether the SEB complexity affects the simulation of the $1 \times CO_2$, $2 \times CO_2$ and the changes in temperature (ΔT), precipitation (ΔP), evaporation (ΔE) and soil moisture (ΔW) over Australia, resulting from ΔCO_2 at annual and seasonal timescales.

3.1. Continental scale statistics at the annual timescale

The 'T1' statistic assesses the statistical significance of changes in the grand mean (Appendix 1) between a mode of CHASM and the control (SLAM). Table 3 shows the impact of varying SEB complexity on air temperature (T) over all Australian land points. In general, changing the SEB complexity does not cause statistically significant changes at the annual timescale, although the results from the simplest mode, RS, are significantly different (at a confidence level of 95%) from SLAM at both $1 \times CO_2$ and $2 \times CO_2$. The simulated ΔT is not affected by SEB complexity. Thus, with the exception of RS, simulations of T over Australia are unlikely to be affected by the complexity of the SEB, provided SEB complexity exceeds that of mode RS. This suggests that explicitly representing canopy interception in LSMs over Australia is necessary to simulate T at $1 \times CO_2$ and $2 \times CO_2$ (since mode RSI which includes this process is not significantly different from SLAM, in contrast to model RS). The simulated ΔT and associated climate impact analyses are unlikely to be limited by uncertainty in how SEB is parameterised, even if SEB is represented very simply.

In the case of precipitation (P), at $1 \times CO_2$ any change in SEB complexity leads to a statistically significant impact. In contrast, at $2 \times CO_2$ the impact of changing SEB complexity is never significant. ΔP is only statistically significant between SLAM and SLAM1T. Thus, SEB complexity may influence the simulation of actual precipitation over Australia, but ΔP is unlikely to be affected by uncertainty in the modelling of the SEB.

In the case of evaporation (E), the change from SLAM to SLAM1T and from SLAM to RSGI leads to statistically significant changes at $1 \times CO_2$ and $2 \times CO_2$ (Table 3). In addition, at $2 \times CO_2$ the change from

Table 3. Australia-wide averaged p-values calculated following Wigley & Santer (1990). The 'T1' statistic is shown, which assesses the statistical significance of changes in the grand mean between a mode of CHASM and the control (SLAM). **Bold**: significantly different from SLAM at 95% confidence level

	SLAM1T-SLAM	RSGI-SLAM	RSI-SLAM	RS-SLAM
Air temperature (T)				
$1 \times CO_2$	0.59	0.88	0.83	**0.96**
$2 \times CO_2$	0.59	0.79	0.22	**0.98**
ΔT	0.44	0.35	0.1	0.54
Precipitation (P)				
$1 \times CO_2$	**0**	**0.01**	**0.04**	**0.04**
$2 \times CO_2$	0.43	0.17	0.22	0.14
ΔP	**0.97**	0.84	0.85	0.57
Evaporation (E)				
$1 \times CO_2$	**0.01**	**0**	0.31	0.06
$2 \times CO_2$	**0.03**	**0.01**	0.75	**0.01**
ΔE	0.72	0.76	0.75	0.44
Root zone soil moisture (W)				
$1 \times CO_2$	0.87	**0.96**	**0.54**	**1**
$2 \times CO_2$	**0.98**	**0.99**	**0.91**	**1**
ΔW	**0.97**	0.81	0.88	0.41

SLAM to RS also leads to statistically significant changes. However ΔE is never statistically significant and is insensitive to SEB complexity.

Finally, in the case of soil moisture (W), varying the complexity of the SEB almost always leads to statistically significant changes at both $1 \times CO_2$ and $2 \times CO_2$. However, the only impact on ΔW follows the change from SLAM to SLAM1T, which is probably related to the change in precipitation.

Table 3 suggests that the complexity of SEB has a statistically significant impact on the annually averaged Australian-wide surface quantities (W and E) at $1 \times CO_2$ and $2 \times CO_2$. The impact of the SEB complexity on mean precipitation is CO_2-level dependent, and only the largest change in SEB complexity influences mean temperature.

Table 3 shows little evidence that SEB complexity influences the simulation of the change in the grand mean of any of these 4 quantities resulting from CO_2. This adds confidence to existing assessments of the impacts of ΔCO_2 on the Australia-wide mean climate (e.g. Whetton et al. 2001), since uncertainty in how we model the SEB does not affect the sensitivity of these quantities to ΔCO_2.

3.2. Seasonal impacts

The results at annual timescales may hide important differences in the role of the SEB at seasonal timescales. At $1 \times CO_2$ many of the modes simpler than SLAM show high p-values in the point-by-point differences in the temporal mean of T (NT5) (Fig. 2a). These are statistically significant at a 95% confidence level when $p > 0.95$ or $p < 0.05$. These modes tend to overestimate the point-by-point temporal mean T compared to SLAM. Assuming that SLAM is the most realistic model (Xia et al. 2002, Leplastrier et al. 2003), this indicates that the simpler models perform poorer than SLAM. The point-by-point temporal variance (NF5, Fig. 2b) shows less of a bias (i.e. more significant values). The overall temporal (SPRET1, Fig. 2c) and spatial (SPREX1, Fig. 2d) variance show few statistically significant differences in the results. This suggests that SEB complexity does not contribute to significant uncertainty in the simulation of the overall spatial or temporal variance at $1 \times CO_2$. A similar result is obtained for the simulation of T at $2 \times CO_2$ (Fig. 3) suggesting that the main impact of SEB complexity is on the point-by-point temporal mean rather than the point-by-point

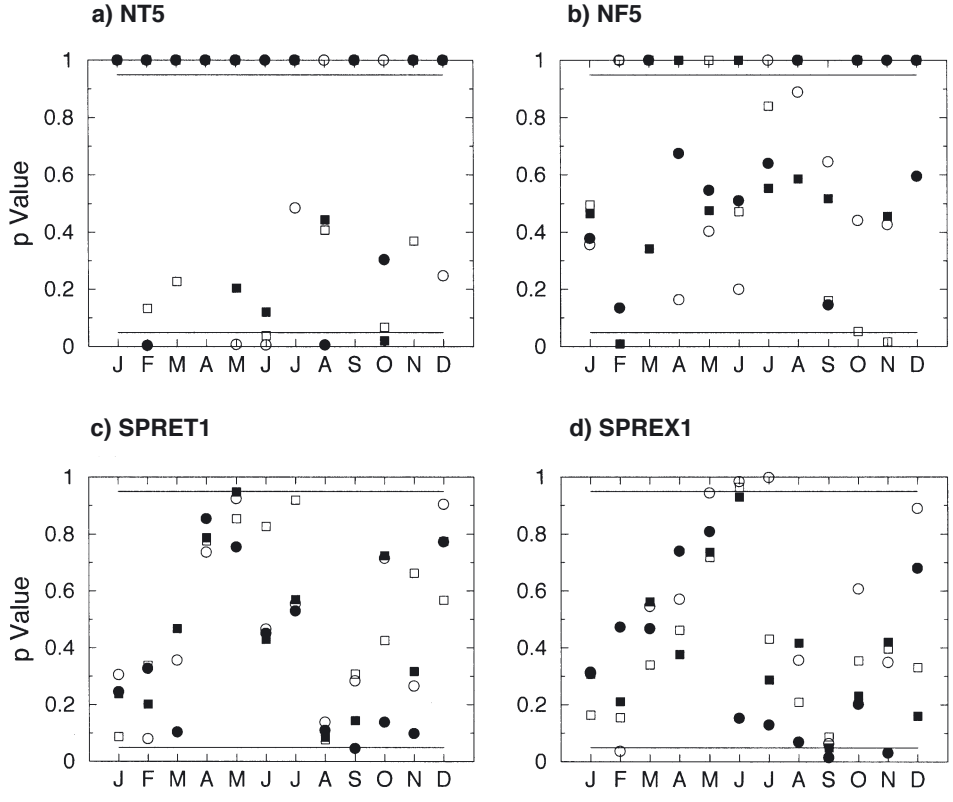

Fig. 2. Results for the 4 statistics (NT5, NF5, SPRET1 and SPREX1) for air temperature for each mode of CHASM differenced from SLAM at $1 \times CO_2$. Values >0.95 or <0.05 are statistically significant at a 95% confidence level. SLAM1T-SLAM (●); RSGI-SLAM (■); RSI-SLAM (□); RS-SLAM (○)

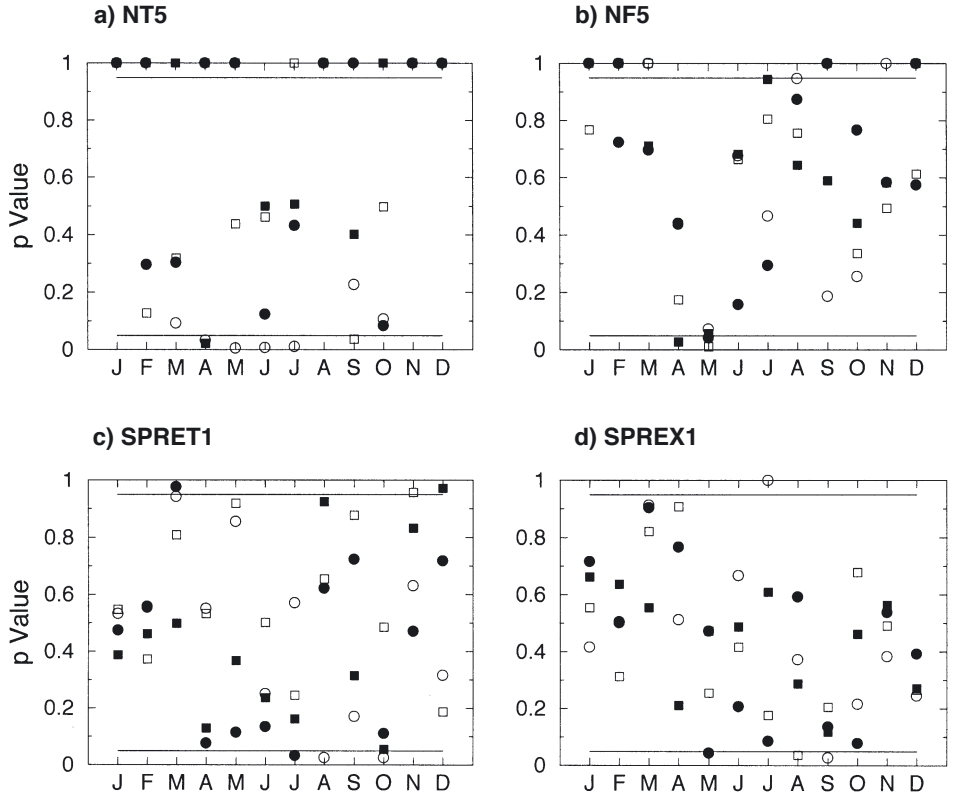

Fig. 3. As Fig. 2, but at $2 \times CO_2$

temporal variance, or on the overall spatial or temporal variance.

The impact of CO_2 on ΔT is shown in Fig. 4a as the actual test statistic value (NT5) for the point-by-point temporal mean (cf. Fig. 11 in Santer & Wigley 1990). The results from the 4 modes of CHASM fall within a single distribution in Fig. 4a which implies that the differences in ΔT between the 4 modes from SLAM are not systematic. A similar result is obtained for the point-by-point temporal variance, NF5 (Fig. 5a). In terms of assessing climate scenario uncertainty to SEB, these results show that the simulation of the point-by-point temporal mean and variance of T is insensitive to SEB complexity. The results for P are very similar for the point-by-point temporal mean (Fig. 4b), but different for the point-by-point temporal variance. Fig. 5b shows a strong seasonality in the test statistic, and this suggests that the temporal variance simulated by the 4 simpler models (RS ... SLAM1T) are similar to each other, but different from SLAM in winter (JJA). This result implies a need for a more complex LSM (e.g. SLAM) to simulate the temporal variance of wintertime P over Australia.

The impact of SEB complexity on E at $1 \times CO_2$ (Fig. 6) is quite different from T or P. In the case of the point-by-point temporal mean (Fig. 6a), mode RS shows a tendency to underestimate temporal mean E during most of the year. There is also some evidence of a seasonal response to SEB complexity in the point-by-point temporal variance for mode RS. Mode RS also underestimates the overall temporal variance (Fig. 6c) and almost all modes commonly underestimate the overall spatial variance compared to SLAM (Fig. 6d) except during winter ($0.05 < p < 0.95$). A similar result is obtained for $2 \times CO_2$ (Fig. 7). The SEB complexity does not affect the simulation of the point-by-point time temporal mean of ΔE (Fig. 4c). In the case of the point-by-point temporal variance (Fig. 5c), the results form 2 distinct groups between July and October demonstrating a statistically significant difference in the point-by-point temporal variance depending on the choice of SEB complexity.

The case of W is similar to E, and the results from $1 \times CO_2$ and $2 \times CO_2$ are therefore omitted. The impact of ΔCO_2 on ΔW includes a systematic tendency for the simpler modes to overestimate the point-by-point temporal mean W (Fig. 8a). The point-by-point temporal variance is simulated in ways more similar to SLAM (Fig. 8b). However, all modes except SLAM1T underestimate the point-by-point temporal variance in winter (JJA). This suggests that there is a need to include tiling in the simulation of ΔW.

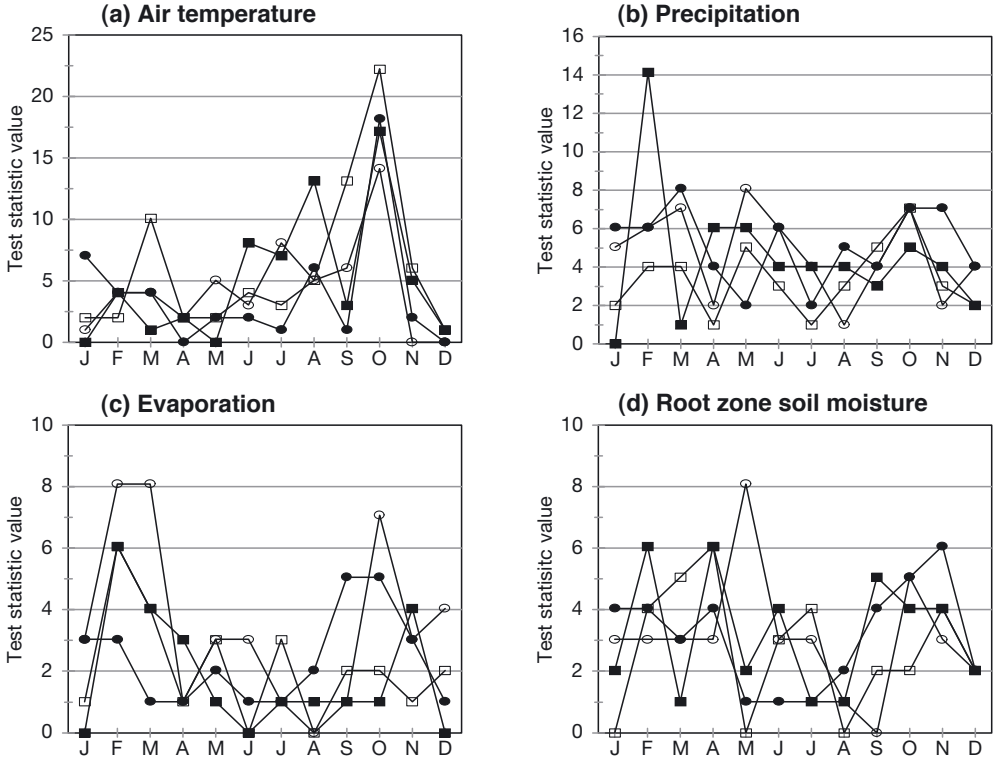

Fig. 4. Actual test statistics for the point-by-point temporal mean (NT5) for the seasonal cycle for SLAM1T-SLAM (●); RSGI-SLAM (■); RSI-SLAM (□) and RS-SLAM (○). (a) Air temperature; (b) Precipitation; (c) Evaporation; (d) Root zone soil moisture. The *y*-axis shows the number of 'successful' local 2-tailed 5% *t*-tests, where success is defined as a locally significant result

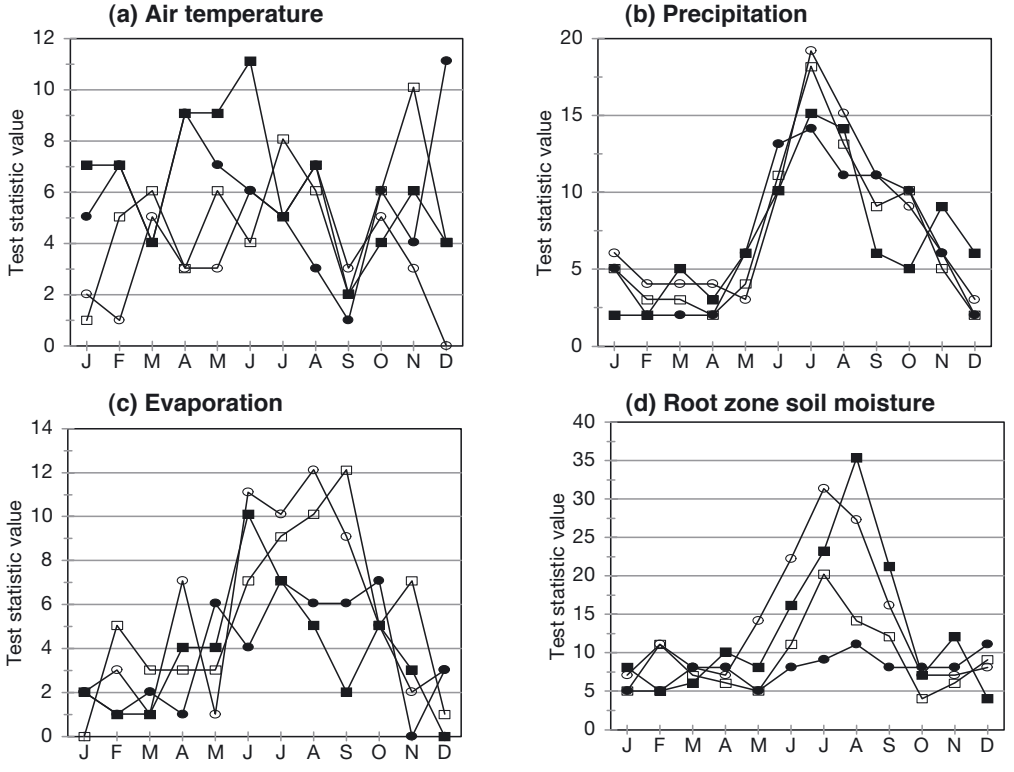

Fig. 5. As Fig. 4 but for the point-by-point temporal variance (NF5)

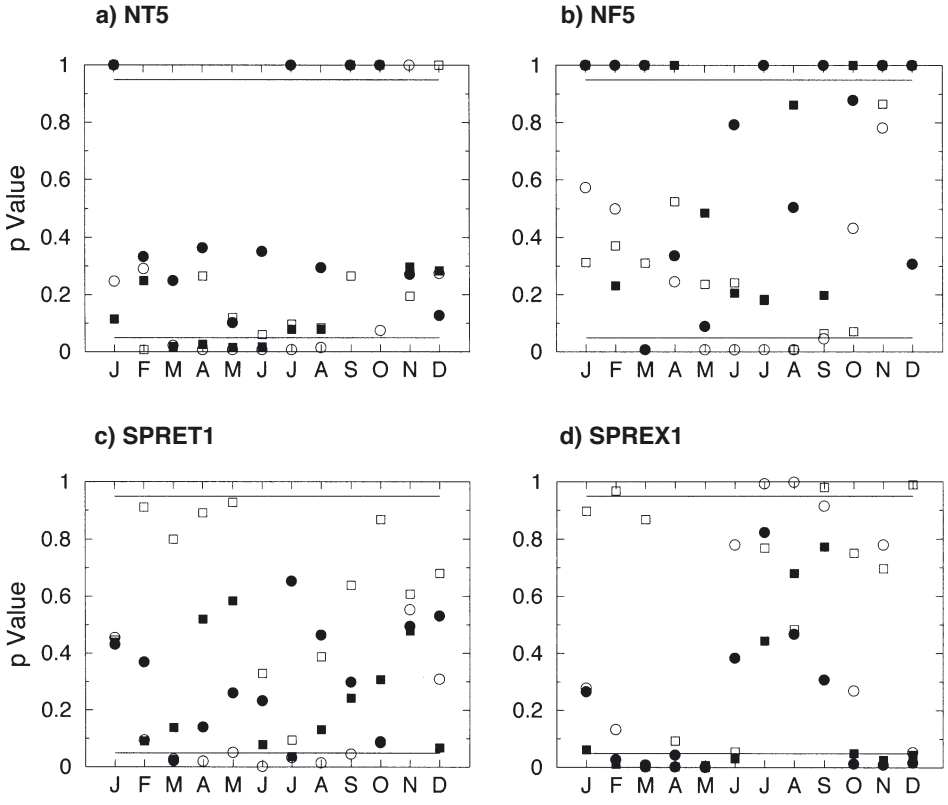

Fig. 6. As Fig. 2, but for latent heat flux

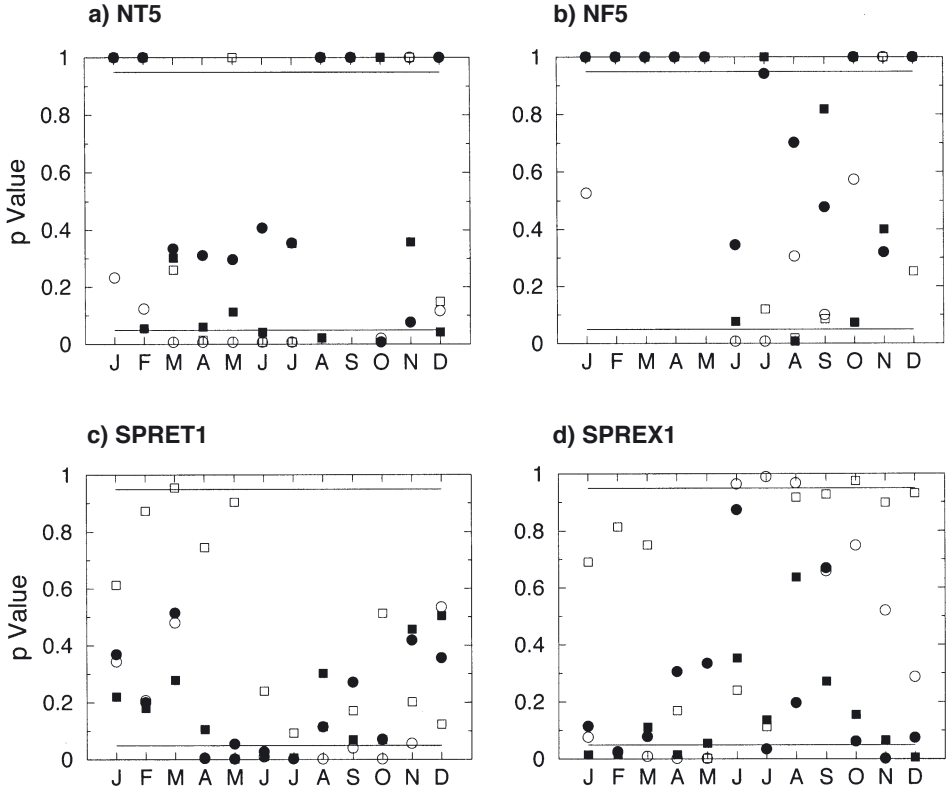

Fig. 7. As Fig. 3, but for latent heat flux

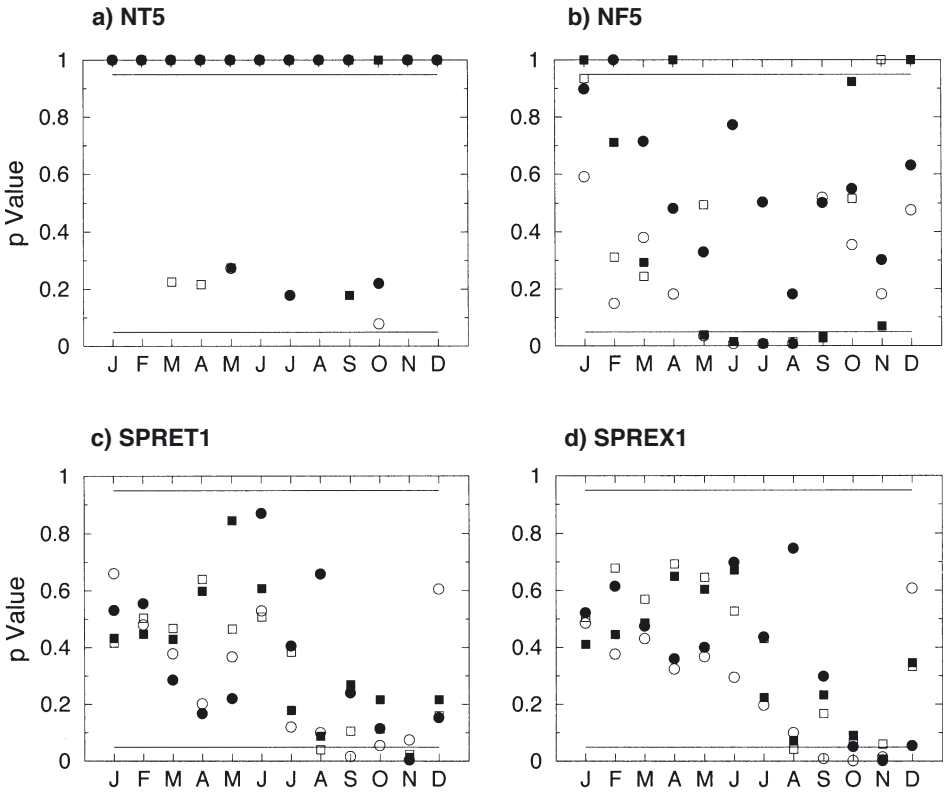

Fig. 8. As Fig. 2, but for ΔW

A careful assessment of Fig. 8d shows that during the first half of the year, the simpler modes capture the behaviour of SLAM in terms of the overall temporal and spatial variance, but fail to replicate SLAM as the continent dries. Thus, many levels of SEB-parameterisation are satisfactory when simulating W under wet conditions, but as that the continent dries a more physically-based LSM is required. This result is supported by Fig. 5d where the test statistics show 2 distributions in the results from CHASM, suggesting that SEB complexity leads to statistically significant differences in the simulation of the point-by-point spatial and temporal variances in ΔW following ΔCO_2.

Overall, our results demonstrate that at seasonal timescales, the complexity of SEB affects the simulation of the point-by-point temporal mean and temporal variance for T, E, P and W through much of the year at both $1 \times CO_2$ and $2 \times CO_2$. There is little evidence that the overall temporal or spatial variances are affected by SEB complexity in either T or P at $1 \times CO_2$ and $2 \times CO_2$. However, these statistics are statistically significant for E and W during dry seasons.

In terms of assessing the uncertainty in climate change scenarios resulting from SEB uncertainty, the changes in quantities following ΔCO_2 show a different behaviour. There is no evidence that the simulation of the change in the *temporal mean* (Fig. 4) for T, E, P and W is sensitive to the SEB complexity. There is also no evidence that the simulation of the *temporal variance* in ΔT is sensitive to the SEB complexity. However, Fig. 5 shows a strong seasonal variation in the *temporal variance* associated with SEB complexity for ΔP, ΔE and ΔW. This demonstrates that the point-by-point temporal variance is sensitive to the complexity of SEB for these variables and that the use of temporal variance in climate scenario development is limited by SEB uncertainty.

4. DISCUSSION AND CONCLUSIONS

We used the AMIP framework with CHASM to explore the impact of SEB complexity on model behaviour over Australia. The representation of the SEB in LSMs can vary widely, representing a lack of consensus in how these processes should be modelled. This study has investigated whether variations in the complexity of the SEB affect the simulation of air temperature, rainfall, evaporation and soil moisture at $1 \times CO_2$, $2 \times CO_2$ and in the simulation of change in these quantities over Australia following ΔCO_2. Our rationale is to determine whether impact scenarios produced for a

continent (e.g. Whetton et al. 2001) are limited by uncertainty in how represent SEB.

At the continental scale, annually averaged T was insensitive to the complexity of SEB, unless the simplest mode (RS) was used (Table 3). This suggests that, provided interception is modelled explicitly (as in mode RSI), simulations of annually averaged T over Australia are insensitive to the complexity of SEB. In the case of P, simulations are sensitive to the complexity of the SEB at $1 \times CO_2$, but the $2 \times CO_2$ simulation or ΔP are largely insensitive. Similarly, in the case of E and W, the SEB may be important at $1 \times CO_2$ or $2 \times CO_2$ (Table 3), but at the continental scale, the annually averaged ΔE is insensitive to the SEB complexity. This means that at the continental and annual scales, simulations of the impact of ΔCO_2 on the mean climate of Australia are unlikely to be affected by uncertainty in SEB. Thus, scenarios produced by, for example, Whetton et al. (2001) are unlikely to be limited by uncertainty in the representation of the SEB.

At the seasonal timescale, results show that ΔT is insensitive to the complexity of SEB in both the point-by-point temporal mean and the temporal variance. Further, at the seasonal timescale, there is no evidence that the complexity of the SEB affects the simulation of the temporal mean of ΔT, ΔP, ΔE and ΔW (Fig. 4). SEB complexity also does not affect the simulation of the point-by-point temporal variance for ΔT. However, the point-by-point temporal variance for ΔP, ΔE and ΔW are sensitive to SEB complexity during dry seasons. This strongly implies a link between the continental hydrology and the atmosphere via changes in evaporation.

In terms of the design of LSMs for the Australian continent, these results suggest the following: A surface resistance and explicit interception should be included in the SEB and the resistance should probably be allowed to vary spatially and temporally, since this appears to matter in the simulation of quantities under $1 \times CO_2$ and $2 \times CO_2$. If grand or temporal mean changes in continental scale results are all that are required, this will probably suffice and will be as good as a far more complex SEB parameterisations in capturing the impact of CO_2. If seasonal changes resulting from ΔCO_2 are required at a sub-continental scale, then SEB parameterisation of the complexity of SLAM appears to perform differently to simpler schemes in the simulation of point-by-point temporal variance in P, and during dry periods, in E and W. The point-by-point temporal variance in T appears insensitive to the complexity of SEB (see Section 3.2).

This result should not be overinterpreted nor confused with the significance of the land surface (as distinct from SEB). (1) The results discussed here do not include carbon and associated feedbacks which, if included, could enhance the importance of the SEB. (2) This study only considers the role of the SEB and does not include changes in soil temperature or soil hydrology models, a key component of land surface models (e.g. Gedney et al. 2000). Indeed, in our results we found a common signal in the point-by-point spatial variance in ΔP, ΔE and ΔW, suggesting that SEB complexity affects the role of surface hydrology. We therefore suspect that the re-focussing of efforts from the SEB towards the inclusion of catchment-based models (e.g. Koster et al. 2000) are likely to improve model predictions. (3) We were only able to use a single climate model, and therefore some caution needs to be exercised in generalizing our results. (4) Generalizing our results to other continents is difficult, but we suspect that the key observation that the SEB appears to increase in importance as the surface dries applies widely.

In conclusion, our results are encouraging for scenario developments of the type discussed by Whetton et al. (2001), as CO_2-induced changes in the *means* of T, P, E and W are insensitive to the complexity of SEB. Since these quantities are commonly used in impact analyses, this suggests that uncertainty in SEB models does not limit climate change scenarios over Australia. However, if the impacts of CO_2 on the *temporal variance* of ΔP, ΔE and ΔW are required, then uncertainly in how to model the SEB may be a major problem that limits the value of individual scenarios. While variances are not used in current impact analyses these statistics may be needed in the future and it is worth noting that they appear much less reliable than the equivalent means.

Acknowledgements. This work was partly funded via an Australian Research Grant. Support for B.M. was provided by the Australian Greenhouse Office. We thank M. Zhao for help with the statistical analyses.

LITERATURE CITED

Betts AK, Ball JH, Beljaars ACM, Miller MJ, Viterbo PA (1996) The land surface-atmosphere interaction: a review based on observational and global modeling perspectives. J Geophys Res 101:7209–7225

Crossley JF, Polcher J, Cox PM, Gedney N, Planton S, (2000) Uncertainties linked to land surface processes in climate change simulations. Clim Dynam 16:949–961

Desborough CE, (1999) Surface energy balance complexity in GCM land surface models. Clim Dynam 15:389–403

Desborough CE, Pitman AJ, McAvaney B, (2001) Surface energy balance complexity in GCM land surface models, Part II: coupled simulations. Clim Dynam 17:615–62

Gates WL (1992) AMIP: The atmospheric model intercomparison project. Bull Am Meteorol Soc 73:1962–1970

Gedney N, Cox PM, Douville H, Polcher J, Valdes PJ (2000) Characterising GCM land surface schemes to understand their responses to climate change. J Clim 13:3066–3079

Koster RD, Suarez MJ, (1992) Modelling the land surface boundary in climate models as a composite of independent vegetation stands. J Geophys Res 97:2697–2715

Koster RD, Suarez MJ, Ducharne A, Stieglitz M, Kumar P (2000) A catchment-based approach to modeling land surface processes in a general circulation model 1. Model structure. J Geophys Res 105:24809–24822

Leplastrier M, Pitman AJ, Gupta H, Xia Y, (2003) Exploring the relationship between complexity and performance in a land surface model using the multi-criteria method. J Geophys Res 107:4443, doi: 10.1029/2001JD000931

Liang X, Wood EF, Lettenmaier DP, Lohmann D and 25 others (1998) The project for intercomparison of land-surface parameterisation schemes (PILPS) Phase 2(c) Red-Arkansas River basin experiment: 2. Spatial and temporal analysis of energy fluxes. Glob Plan Change 19:137–159

Manabe S, (1969) Climate and the ocean circulation: 1. The atmospheric circulation and the hydrology of the earth's surface. Mon Weath Rev 97:739–805

Mitchell JFB, Johns TC, Gregory JM, Tett SFB (1995) Climate response to increasing levels of greenhouse gases and sulphate aerosols. Nature 376:501–504

Pitman AJ, McAvaney BJ (2003) The role of surface energy balance complexity in land surface models' sensitivity to increasing carbon dioxide. Clim Dynam 19:609–618

Preisendorfer RW, Barnett TP (1983) Numerical model-reality intercomparison tests using small-sample statistics. J Atmos Sci 40:1884–1896

Robock A, Vinnikov K Ya, Schlosser CA, Speranskaya NA, Xue Y (1995) Use of midlatitude soil moisture and meteorological observations to validate soil moisture simulations with biosphere and bucket models. J Clim 8:15–35

Santer B, Wigley TML (1990) Regional validation of means, variances, and spatial patterns in general circulation model control runs. J Geophys Res 95:829–850

Sellers PJ, Dickinson RE, Randall DA, Betts AK, Hall FG, Berry JA, Collatz GJ, Denning AS, Mooney HA, Nobre CA, Sato N, Field CB, Henderson-Sellers A (1997) Modelling the exchanges of energy, water and carbon between continents and the atmosphere. Science 275:502–509

Sellers PJ, Randall DA, Collatz CJ, Berry JA, Field CB, Dazlich DA, Zhang C, Collelo G, Bounoua L (1996) A revised land-surface parameterization (SiB2) for atmospheric GCMs. Part 1: Model formulation. J Clim 9: 676–705

Whetton PH, Katzfey JJ, Hennessy KJ, Wu X, McGregor JL, Nguyen K (2001) Developing scenarios of climate change for Southeastern Australia: an example using regional climate model output. Clim Res 16:181–201

Wigley TML, Santer B (1990) Statistical comparison of spatial fields in model validation, permutation and predictability experiments. J Geophys Res 95:851–865

Xia Y, Pitman AJ, Gupta HV, Leplastrier M, Henderson-Sellers A, Bastidas LA (2002) Calibrating a land surface model of varying complexity using multi-criteria methods and the Cabauw data set, 2002. J Hydrometeorol 2:181–194

Appendix 1. Outline of the statistical methods

The derivation of the statistical methods used in this study is provided in full by Wigley & Santer (1990) (hereafter WS90), and we use the same terminology as WS90. A key feature of the WS90 approach is the use of the re-sampling procedure, *pooled permutation* (Preisendorfer & Barnett 1983); provided a sufficient number of randomised re-samplings are performed (WS90 recommend 1000) a null sampling distribution can be generated against which the test statistics can be compared. It is then a straightforward task to compute the statistical significance of the test statistic value.

A1. Comparison of overall means, point-by-point means and point-by-point variances

T1 is designed to test the statistical significance of differences between the grand means of 2 space-time fields (d and m). The grand mean for d ($<d>$) is (WS90, Table 1):

$$<d> = \left(\sum_x \sum_t d_{xt}\right) / n_x n_t \quad (1)$$

where x is space, t is time and n_x and n_t are the number of data points in space and time, respectively. (The grand mean for $<m>$ is defined similarly). The statistical significance of the differences in the grand means, $<d> - <m>$ can be estimated using an appropriate t-test. The statistic, T1 (WS90, Eq. 1) is:

$$t = (<d> - <m>) / S_1 \quad (2)$$

where S_1 is a measure of the variance of the sampling distribution of $<d> - <m>$ and contains contributions from d and m. In our paper we use statistic S_1 (WS90 offer alternative ways of testing the statistical significance of a change in the grand mean). S_1 ignores the distinction between x (space) and t (time) and considers the overall d and m variances (WS90 Eq. 2):

$$S_1^2 = (\mathrm{GSSD} + \mathrm{GSSM}) / [n_x n_t (n_x n_t - 1)] \quad (3)$$

and from WS90 (Table 1) the grand mean of the sum of the squares for d (GSSD) is:

$$\mathrm{GSSD} = \sum_x \sum_t (d_{xt} - <d>)^2 = \left(\sum_x \sum_t d_{xt}^2\right) - n_x n_t <d> \quad (4)$$

and where GSSM is defined similarly with respect to m; d_{xt} is the value of d at point x and time t ($x = 1, n_x$; $t = 1, n_t$).

NT5 is designed to test the differences in time means, *grid point by grid point*. In our study NT5 is used to compare the time means at each grid point (defined by Eq. 5) using a 2-tailed local test for the difference in means. Assuming that the d and m variances do not differ significantly, WS90 (Eq. 7) state that:

$$t = (\bar{d}_{x.} - \bar{m}_{x.}) / S_x \quad (5)$$

where (WS90, Table 1) $\bar{d}_{x.}$ is the time average of d at point x where $\bar{d}_{x.} = \sum_t d_{xt} / n_t$ and $m_{x.}$ and $\bar{m}_{x.}$ are defined similarly (WS90, Eq. 8) and:

$$S^2{}_x = (s^2{}_{d,x} + s^2{}_{m,x}) / (n_t - 1) \quad (6)$$

where (WS90, Table 1) $s^2{}_{d,x}$ is the time variance of d at point $x = \sum_t (d_{xt} - \bar{d}_{x.})^2 / n_t$ and $s^2{}_{m,x}$ is defined similarly.

To calculate NT5, the number (n_s) of 'successful' points

Appendix 1 (continued)

where t is locally significant at the 5% level is determined from Student's t distribution with $2n_t - 2$ degrees of freedom. NT5 then accesses field significance and is defined as the fractional number of successes, so NT5 = (n_s/n_x).

NF5 is designed to test the differences in temporal variances, grid-point by grid-point. We used a 2-tailed local test for significance. WS90 (equation 18) state:

$$F = s^2_{d,x} / s^2_{m,x} \quad (7)$$

To calculate NF5, the number of locally significant points is used, taking into account multiplicity and spatial autocorrelation.

A2. Comparison of temporal variances

SPRET1 is the ratio of the spatial means of the time variances (WS90 equation 15) and tests the overall difference in temporal variances: SPRETI = $\overline{s^2_{d,x}} / \overline{s^2_{m,x}}$ where the overbar denotes a spatial average of $s_{d,x}$ and $s_{m,x}$ such that (WS90, Table 1): $\overline{s^2_{d,x}}$ = GSSD − SSXD/$n_x n_t$; and SSXD is the between-x sum of squares:

$$\mathrm{SSXD} = n_t \sum_x (\bar{d}_x - \langle d \rangle)^2 = n_t \left(\sum_x \bar{d}^2_x \right) - n_x n_t \langle d \rangle^2 \quad (8)$$

SPRET1 can also be defined in terms of sums of squares (WS90, Eq. 16):

$$\mathrm{SPRET1} = (\mathrm{GSSD} - \mathrm{SSXD}) / (\mathrm{GSSM} - \mathrm{SSXM}) \quad (9)$$

where SSXM is defined similarly to SSXD (WS90, Table 1). Either large or small values of SPRET1 are significant.

The statistical test for the differences in temporal variance ($s^2_{d,x}$ and $s^2_{m,x}$), compared grid point by grid point is (WS90, Eq. 18): $F = s^2_{d,x} / s^2_{m,x}$

Local significance can be judged using an F-distribution with $n_t - 1$, $n_t - 1$ degrees of freedom.

A3. Comparison of spatial variances

SPREX1 is the ratio of the time-mean spatial variances, i.e. the overall difference in the spatial variances. It is (WS90, Eq. 19):

$$\mathrm{SPREX1} = \overline{s^2_{d,t}} / \overline{s^2_{m,t}} = (\mathrm{GSSD} - \mathrm{SSTD}) / (\mathrm{GSSM} - \mathrm{SSTM}) \quad (10)$$

where the time average of $s^2_{d,t}$ is given by (WS90, Table 1):

$\overline{s^2_{d,t}}$ = (GSSD − SSTD)/$n_x n_t$ and $s^2_{m,t}$ is defined similarly with respect to m. SSTD is the within-x sum of squares (WS90, Table 1):

$$\mathrm{SSTD} = n_x \sum_t (\bar{d}_t - \langle d \rangle)^2 = n_x \left(\sum_t \bar{d}_t^{\,2} \right) - n_x n_t \langle d \rangle^2 \quad (11)$$

and the spatial average of d at time t is WS90, Table 1):

$$d_t = \sum_x d_{xt} / n_x \quad (12)$$

WS90 discuss many features of these statistics and provide guidance regarding their use. We highly recommend that the primary source for these tests is used in determining the appropriateness of statistical tests for model examination.

Editorial responsibility: Penny Whetton, Aspendale, Australia

Submitted: May 17, 2002; Accepted: December 22, 2003
Proofs received from author(s): January 14, 2004