Evaluation of simulated recent climate change in Australia

Jonas Bhend[1,2,3] and Penny Whetton[1,2]

[1] CSIRO Climate Adaptation Flagship
[2] CSIRO Marine and Atmospheric Research, Aspendale, Australia
[3] Federal Office of Meteorology and Climatology MeteoSwiss, Zurich, Switzerland

(Manuscript received September 2014; accepted March 2015)

The ability to reproduce recent observed climate change in climate models is a pertinent prerequisite for trust in climate projections. Also, information on the consistency of simulated and observed recent changes helps users to interpret near-term climate change projections. A comprehensive assessment of simulated regional trends, however, is often not available. Therefore, we evaluate daily maximum and minimum temperature trends and rainfall trends from 1956-2005 in Australia in simulations from the CMIP5 archive. For all variables and all models, we find significant (at the 10% level) differences between simulated and observed trends in some areas. Except for daily minimum temperature in spring and summer, however, the areas where we find significant differences are smaller than what we expect by chance. In a multivariate evaluation, simulated joint temperature and rainfall trends of all but one model, however, are found to be significantly (at the 10% level) different from the observed trends. Hence, multivariate evaluation provides a stricter test. We conclude that CMIP5 models share trend biases and regional projections therefore have to account for the presence of biases shared across models.

Introduction

Climate models are important tools for climate science. They help us to test our understanding of the climate system and they allow us to make predictions about its future evolution. Confidence in climate model predictions - both forecasts and hindcasts - stems from the careful evaluation of climate models with respect to their ability to reproduce observed phenomena and characteristics (Randall et al., 2007). When evaluating climate models, most often metrics describing the mean climate and climate variability such as the seasonal cycle (Gleckler, Taylor, and Doutriaux 2008; Reichler and Kim 2008; Watterson 2008; Irving et al., 2011; Smith and Chandler 2010; Smith et al., 2013) are used. Aspects of climate change have been extensively evaluated at global to continental scales and the large-scale simulated warming is found to agree well with observations (see Hegerl et al., 2007; Stott et al., 2010 for a review of recent advances). Changes in global precipitation, on the other hand, are less well reproduced in simulations (Zhang et al., 2007; Noake et al., 2012; Polson et al., 2013). Recent studies on regional climate change evaluation focus on specific regions (van Oldenborgh et al., 2009; Haren et al., 2013) or single variables (Bhend and Whetton 2013; van Oldenborgh et al., 2013).

Here we evaluate recent regional trends in seasonal mean daily maximum and minimum temperature and rainfall over Australia. We compare observed trends with trends in experiments from state-of-the-art global climate models from the World Climate Research Programme's (WCRP) Coupled Model Intercomparison Project phase 5 (CMIP5, Taylor, Stouffer, and Meehl (2011)). These models have been run with a comprehensive set of observed and reconstructed boundary conditions including the changing atmospheric concentrations of greenhouse gases, aerosols, and ozone as well as solar irradiance changes. The models thus produce a realistic - within model limitations - representation of recent climate change. It is important to note, however, that a fraction of the observed and simulated recent change is due to natural internal variability in the climate system. This internal variability can be thought of as random and therefore internal variability in simulations differs from the observed. The remainder of the change – the signal – is due to changes in external forcing mechanisms and is in principle reproducible in long-term simulations. Only this deterministic, forced component of climate change can be used for evaluation of climate models. Therefore, being able to separate signal from noise is crucial when evaluating transient behaviour in climate models. Separating signal and noise, however, is

Corresponding author address: Jonas Bhend, Federal Office of Meteorology and Climatology, MeteoSwiss, Operation Center 1, CH–8058 Zürich-Flughafen
Email: jonas.bhend@meteoswiss.ch

often non-trivial and a multitude of approaches exist (Hegerl and Zwiers 2011). For simplicity, we assume here that the regional signal in both temperature and rainfall over the period from 1956 to 2005 is approximately linear and the remainder is representative of internal variability (see later text, van Oldenborgh et al. (2009), and Bhend and Whetton (2013)).

Recent warming in Australia has been partially attributed to human influence (Karoly and Braganza 2005b; Min and Hense 2007). Successful attribution of recent rainfall trends, on the other hand, has not yet been achieved, but understanding recent observed regional precipitation changes in Australia is a focus of many recent research projects such as the South Eastern Australian Climate Initiative (CSIRO 2012) and the Indian Ocean Climate Initiative (Indian Ocean Climate Initiative 2012). Numerous studies have analysed the contribution of different external forcing mechanisms (Timbal, Arblaster, and Power 2006; Narisma and Pitman 2003; Timbal and Arblaster 2006; Rotstayn et al., 2009) to the observed rainfall changes in different parts of Australia. In addition, the influence of proximate causes of Australian rainfall variability has been intensively studied and linkages with major atmospheric and oceanic modes of variability such as the Southern Annular Mode or the El Niño-Southern Oscillation have been documented (Risbey et al., 2009; Nicholls 2010; Larsen and Nicholls 2009; Frederiksen et al., 2010; Smith and Timbal 2012; Cai and Cowan 2012). These studies into proximate drivers imply that a considerable fraction of recent rainfall variability and trends is due to natural internal variability rather than externally forced.

Variability in seasonal mean daily maximum temperature is strongly negatively correlated with co-located rainfall variability across Australia (Jones and Trewin 2000; Nicholls 2004). That is, wet years tend to result in lower average daily maximum temperatures, dry years in anomalously hot days. As a consequence of the strong cross-variable correlations, our uncertainty about climate change estimates is highly correlated across variables. For example, the concurrent recent increase in summer (DJF) rainfall and cooling in daily maximum temperature in north-western Australia is much more likely than the combination of increasing rainfall and strong warming. Consequently, it is important to take cross-variable correlation into account when evaluating climate models with respect to recent changes.

Following Nicholls (2003) and Karoly and Braganza (2005a), we also analyse temperature trends after removing all rainfall-related variability. Rainfall variability is a proxy for circulation variability and the residual temperature trends thus reflect warming independent of circulation changes. Removing rainfall-related variability proves to be an effective means to increase the signal-to-noise ratio of an externally forced warming (Karoly and Braganza 2005) and, therefore, residual temperature trends are well suited for model trend evaluation.

The evaluation of simulated recent climate change in Australia builds on previous work on the consistency of regional climate change assessed globally (Bhend and Whetton 2013). In contrast to the global study, we introduce a multivariate assessment of the ability of state-of-the-art global climate models to reproduce recent observed climate trends in Australia. By taking into account correlation between temperature and rainfall variability, the multivariate evaluation allows us to explain some of the identified inconsistencies between simulated and observed trends.

Observations and simulations used

We evaluate simulations submitted to the multi-model data archive of the WCRP's Coupled Model Intercomparison Project phase 5 (CMIP5). We use simulations for the 20th century from the historical experiment. These simulations are constrained by observed boundary conditions including changing atmospheric greenhouse gas concentrations and other anthropogenic and natural forcings up to 2005 (Taylor, Stouffer, and Meehl 2011). We average all available initial condition members per model to maximise the signal-to-noise ratio of externally forced changes. In addition, we also analyse simulations from the pre-industrial control run. These simulations do not exhibit changes in external forcings and can therefore be used to study natural internal variability. For these control runs we require that at least 200 years of data for all three variables are available. The list of models, initial condition ensemble members for the historical simulation and length of the pre-industrial control simulation is shown in Table 1.

We evaluate the simulations against high-resolution gridded monthly precipitation and daily maximum and minimum temperature data sets. Homogeneised daily maximum and minimum temperatures from the Australian Climate Observations Reference Network - Surface Air Temperature (ACORN-SAT, Trewin 2013) have been used. These stations were corrected for changes in measurement practices (e.g. changes in instrumentation and/or changes to the observing site). For rainfall, homogeneised series are unfortunately not available yet and we use rainfall compiled as part of the Australian Water Availability Project (AWAP; Jones et al., 2009) instead. These data consist of interpolated Australian station data for daily precipitation on a regular latitude/longitude grid with horizontal resolution of 0.05° (approximately 5km). In addition to the ACORN-SAT and AWAP data, we also use rainfall and daily maximum and minimum temperature from the CRU TS 3.20 data set, a global high-resolution gridded station-based data set from the Climatic Research Unit (Mitchell and Jones 2005; Harris et al., 2013). CRU TS 3.20 and AWAP / ACORN-SAT are largely based on the same station data, but the preprocessing and spatial interpolation routines differ. For example, in CRU TS 3.20 monthly temperature and precipitation are interpolated whereas the interpolation is carried out using daily values in the other data sets. Contrasting these two observation-based datasets allows us to derive a crude estimate of the contribution of interpolation uncertainty which is likely the largest contributor to observation uncertainty in spatially interpolated climate series (Haylock et al., 2008).

Table 1. Models from the CMIP5 archive used in this study along with the number of initial condition ensemble members (nens), and the number of years in the control simulation used (control).

	model	*nens*	*control*
1	ACCESS1-0	2	500
2	ACCESS1-3	3	500
3	BNU-ESM	1	559
4	CCSM4	6	501
5	CESM1-BGC	1	
6	CESM1-CAM5	3	319
7	CESM1-FASTCHEM	3	222
8	CESM1-WACCM	4	200
9	CMCC-CESM	1	
10	CMCC-CM	1	
11	CMCC-CMS	1	
12	CNRM-CM5	10	
13	CNRM-CM5-2	1	
14	CSIRO-Mk3-6-0	10	500
15	CanESM2	5	996
16	EC-EARTH	4	
17	FIO-ESM	3	
18	GFDL-CM3	3	500
19	GFDL-ESM2G	1	
20	GFDL-ESM2M	1	500
21	GISS-E2-H	17	
22	GISS-E2-H-CC	1	251

	model	*nens*	*control*
23	GISS-E2-R	1	531
24	GISS-E2-R-CC	1	251
25	HadCM3	10	
26	HadGEM2-AO	1	
27	HadGEM2-CC	1	240
28	HadGEM2-ES	5	575
29	IPSL-CM5A-LR	6	1000
30	IPSL-CM5A-MR	3	
31	IPSL-CM5B-LR	1	
32	MIROC-ESM	3	630
33	MIROC-ESM-CHEM	1	
34	MIROC4h	3	
35	MIROC5	5	
36	MPI-ESM-LR	3	
37	MPI-ESM-MR	3	
38	MPI-ESM-P	2	
39	MRI-CGCM3	5	
40	MRI-ESM1	1	
41	NorESM1-M	3	
42	bcc-csm1-1	3	
43	bcc-csm1-1-m	3	
44	inmcm4	1	500

Both the observation and simulation data have been regridded to a common $1.5° \times 1.5°$ latitude/longitude grid using conservative remapping (Jones et al., 1999) prior to further analysis. The horizontal resolution in the CMIP5 models used ranges from $0.5° \times 0.5°$ to $3.75° \times 3.75°$. The majority of the models have been run at a horizontal resolution comparable to the common grid used here.

Analysis scheme

Following Haren et al. (2013), we define the climate change signal A at each grid box as the linear trend from 1956 to 2005. The effect of variability unrelated to long-term changes is taken into account using the standard error of the regression ΔA. We correct standard errors for autocorrelation using the effective sample size as outlined in Wilks (1997).

In addition to the effect of natural variability, gridded observations are subject to various sources of uncertainty including inhomogeneities in the station records and interpolation uncertainty due to the sparsity of the observation network. The latter is of particular importance in Australia with its sparsely populated interior with a limited number of long-term observational records (Jones, Wang, and Fawcett 2009). We address interpolation uncertainty by comparing trends in two different observational datasets (the ACORN-SAT / AWAP and CRU TS3.20 datasets introduced earlier). Following Haren et al. (2013) and Yokohata et al. (2012), we estimate observation uncertainty by computing ΔA_{int}, the standard deviation of trends in the two different datasets (A_{awap} and A_{cru} respectively), as in Eq. 1.

$$A_{obs} = (A_{awap} + A_{cru})/2 \tag{1}$$

$$\Delta A_{int} = \sqrt{\left[(A_{awap} - A_{obs})^2 + (A_{cru} - A_{obs})^2\right]/2} \tag{2}$$

We then add the squared observation uncertainty $(\Delta A_{int})^2$ and the pooled squared standard error of the regression $\overline{(\Delta A_{obs})^2} = \left[(\Delta A_{awap})^2 + (\Delta A_{cru})^2\right]/2$ for an estimate of total uncertainty about the observed change. This allows us to formally incorporate observation uncertainty in the analysis. While interpolation uncertainty is generally deemed the largest contributor to observation uncertainty (Haylock et al., 2008) other sources of uncertainty that are not addressed here such as inhomogeneities in the station record may affect trends more specifically.

To assess the consistency of simulated and observed climate change, we compute the standardised difference z between observed change A_{obs} and the initial condition ensemble mean of simulated changes $\overline{A_{mod}}$ for each model. The difference is standardised using the observed and simulated squared standard errors of the regression ($(\Delta A_{obs})^2$ and $\overline{(\Delta A_{mod})^2}/N$ respectively where N is the number of models) and the estimate of observation uncertainty ΔA_{int} according to Eq. 3.

$$z = \frac{A_{obs} - \overline{A_{mod}}}{\sqrt{(\Delta A_{obs})^2 + (\Delta A_{int})^2 + \overline{(\Delta A_{mod})^2}/N}} \quad (3)$$

We extend this model to be able to evaluate joint temperature and rainfall changes. In the following, we use uppercase Z to denote standardised differences for the multivariate case. In the multivariate case, Z is still a scalar, but the observed change $\boldsymbol{A_{obs}}$ and the simulated change $\boldsymbol{A_{mod}}$ are vectors of length 3 (daily maximum and minimum temperature and rainfall). Accordingly, the squared standard errors ($\boldsymbol{\Delta A_{obs}}$ and $\boldsymbol{\Delta A_{mod}}$ respectively) are 3×3 covariance matrices with the off-diagonal elements reflecting the correlation of the temperature and rainfall residuals. The diagonal elements are the squared standard errors of the regression adjusted for autocorrelation as above. The squared standardised distance Z^2 is then

$$Z^2 = (A_{obs} - \overline{A_{mod}})^T(\Delta A_{obs} + \Delta A_{int} + \overline{\Delta A_{mod}}/N)^{-1}(A_{obs} - \overline{A_{mod}}) \quad (4)$$

Under the null hypothesis that both the observed and simulated change are random draws from the same distribution - i.e. under the assumption that we have a perfect model - the squared standardised differences approximately follow a chi-square distribution with 1 (3) degrees of freedom in the univariate (multivariate) case. We use these distributions for significance testing.

To summarise the results, we also compute root mean squared standardised differences Z_{rms}. To test the significance of spatially aggregated standardised differences, we compare these to distributions of Z_{rms} derived from resampled model results. For every model with multiple initial condition ensemble members, we use each ensemble member in turn as pseudo-observation and compute standardised differences Z and the aggregated statistics Z_{rms} with respect to the remaining ensemble members. The distribution of the 130 resampled Z (from 27 models with more than one initial condition simulation) reflect the distribution of Z (and derived statistics) given we had a perfect model.

Results

Evaluation of simulated trends in daily maximum and minimum temperature

Observed trends in seasonal mean daily maximum temperature from 1956 to 2005 according to ACORN-SAT and CRU TS 3.20 are shown in Fig. 1a-d. We find significant (at the 10% level) warming in all seasons in eastern Australia and significant cooling in north-western Australia in summer (DJF). The median simulated warming in CMIP5 exhibits a weak seasonal cycle with strongest warming in winter (JJA, Fig. 1e-f). Also, simulated warming is strongest in inland Western Australia. The reduced spatial variabiltiy of the median simulated warming is an artefact of aggregating the data of multiple models. The individual simulations exhibit spatial variability in 50-year trends similar to the observations. The ratio of the simulated to the observed spatial standard deviation of seasonal trends in daily maximum temperature ranges from 0.4 to 3 with the central 50% of ratios in the range of 0.8 to 1.1

The standardised differences z between observed and simulated trends in seasonal mean daily maximum temperature range from -5.6 to 5.2 for the different models analysed (not shown). In Fig. 1i-l we show the fraction of the 44 CMIP5 models with simulated seasonal temperature trends that are inconsistent with the observed trends. We find coherent areas where the majority of CMIP5 models fail to reproduce the cooling in north-western and northern Australia (Fig. 1i-l). On the other hand, the simulations underestimate the observed warming in seasonal mean daily maximum temperatures in limited areas in eastern Australia and along the southern coast in winter (JJA).

Observed trends in seasonal mean daily minimum temperature (Fig, 2a-d) show widespread significant warming and local patches of little change or slight cooling for the period from 1956 to 2005. As with maximum temperature, the median simulated warming (Fig, 2e-h) shows a seasonally varying pattern of warming with maximum warming in south-eastern Australia in summer (DJF) and north-western Australia in winter (JJA). Compared with trends in maximum temperature, the area for which the majority of the simulated trends in minimum temperature differs significantly (at the 10% level) from the observations is larger. Simulations generally overestimate the observed warming in south-western Australia and underestimate the observed trend in minimum temperatures in central and north-eastern Australia (Fig, 2i-l).

Figure 1. Observed trend in seasonal mean daily maximum temperature from 1956 to 2005 (a-d), median of simulated trends from 44 CMIP5 models (e-h), and percentage of models with trends that are significantly different from the observed (i-l). Stippling in a-d denotes areas where the observed trend is significantly (at the 10% level) different from zero. Red (blue) shading in i-l denotes areas where the models simulate warming that is larger (smaller) than the observed.

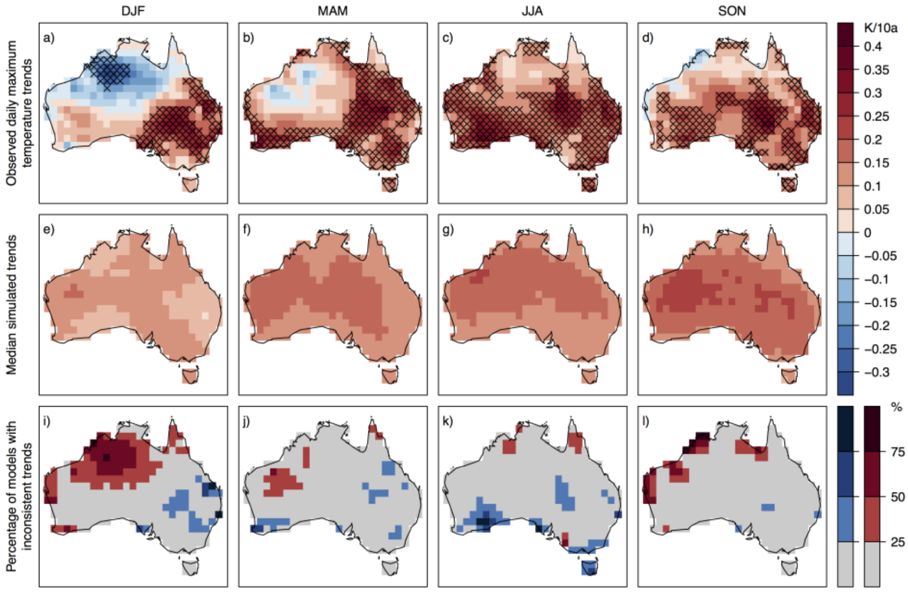
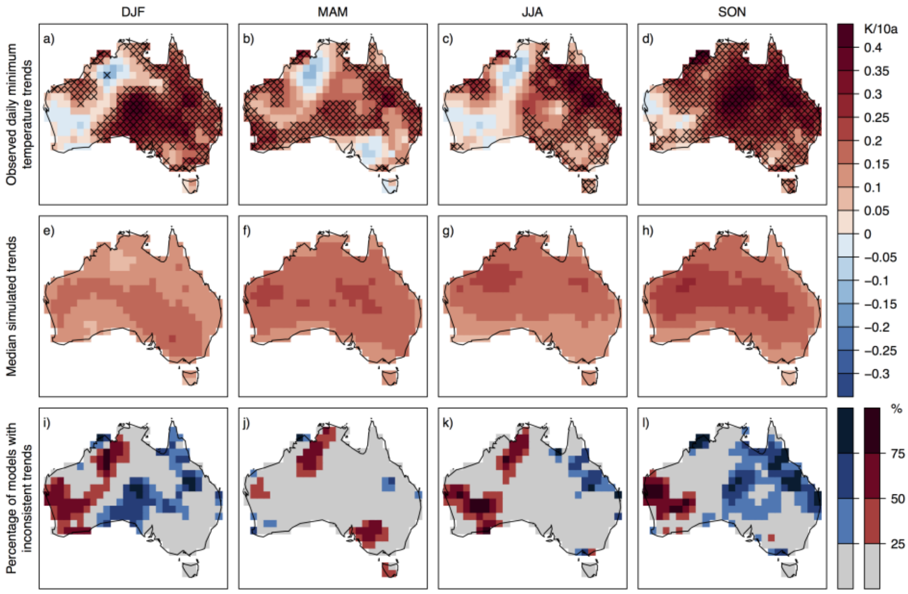


Figure 2. Observed trends in seasonal mean daily minimum temperature from 1956 to 2005 (a-d), median of simulated trends from 44 CMIP5 models (e-h), and percentage of models with trends that are significantly different from the observed (i-l). Stippling in a-d denotes areas where the observed trend is significantly (at the 10% level) different from zero. Red (blue) shading in i-l denotes areas where the models simulate warming that is larger (smaller) than observed.

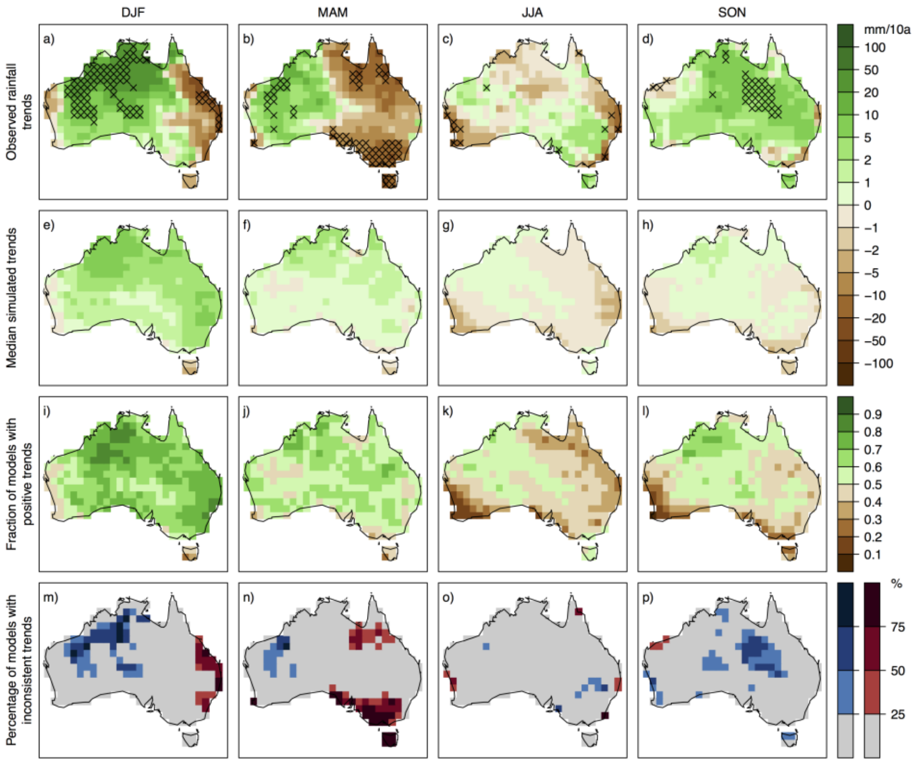


In contrast to seasonal mean daily maximum temperature, simulations underestimate the spatial variability of trends in daily minimum temperature compared to observations. Only 8 out of 44 models exhibit a mean spatial standard deviation of seasonal trends in minimum temperature that is larger than the observed (not shown).

Evaluation of simulated rainfall trends

Trends in observed seasonal precipitation exhibit considerable spatio-temporal variability (Fig. 3a-d). Significant (at the 10% level) increases in rainfall are found in north-western Australia in summer (DJF) and autumn (MAM) and in north-eastern Australia in spring (SON), whereas large areas with significant decreases in rainfall are found in south-eastern Australia in autumn (MAM).

Figure 3. Observed seasonal rainfall trends from 1956 to 2005 (a-d), median of simulated trends from 44 CMIP5 models (e-h), fraction of models with positive trends (i-l), and percentage of models with trends that are significantly different from the observed (m-p). Stippling in a-d denotes areas where the observed trend is significantly (at the 10% level) different from zero. Red (blue) shading in m-p denotes areas where the simulated trends are significantly larger (smaller) than the observed.



The corresponding median simulated trends for the same time period (Fig. 3e-h) show less spatial variability and much less intense changes compared to the observations. This is not just an artefact of the aggregation across multiple models. The individual simulations show less intense rainfall trends than the observations and none out of the 44 CMIP5 models exhibit rainfall trends with a spatial standard deviation as large as the one observed.

Furthermore, there is strong variability between models concerning the sign of the recent rainfall change (Fig. 3i-l). Models agree well on the increases across mainland Australia in summer (DJF, Fig. 3i) and drying in south-western, southern, and eastern coastal Australia in winter (JJA, Fig. 3k) and spring (SON, Fig. 3l). Elsewhere, there is less agreement among models concerning the sign of recent rainfall change.

Coherent areas where the simulated trends in the majority of models differ from the observed seasonal rainfall trends from 1956 to 2005 are found in all seasons but winter (Fig. 3m-p). The majority of the simulations significantly underestimate the observed wetting in north-western Australia in summer (DJF, Fig. 3m) and western Queensland in spring (SON, Fig. 3p). Also the models fail to reproduce the observed drying in south-eastern Australia in autumn (MAM, Fig. 3n) and along the east coast in summer (DJF, Fig. 3m). Unlike for near-surface warming (Figs. 1 and 2), significant differences between observed and simulated rainfall are restricted to regions where the observed changes are significantly different from zero (stippling in Fig. 3a-d).

Evaluation of warming independent of rainfall variability

We also evaluate trends in seasonal mean daily maximum and minimum temperature after removing all rainfall-related variability from the observed and simulated time series. Temperature variability associated with rainfall variability is estimated using linear regression of detrended temperature time series and rainfall series as in Karoly and Braganza (2005a). The thus derived regression coefficient is then used to compute residual temperature time series. It is important to note that we estimate the relationship between temperature and rainfall using detrended series to not bias our estimate in case there is a strong forced signal. But we remove the full rescaled rainfall series to get temperature residuals based on the assumption that most of the observed changes in rainfall are due to natural variability rather than externally forced.

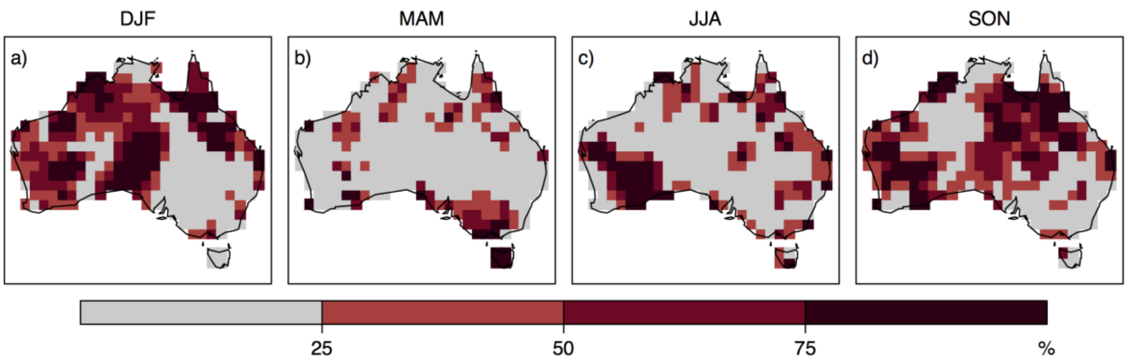
We find more uniform warming in residual maximum and minimum temperature (not shown) and the area with significant warming is larger than in the original time series (Figs. 1 and 2). Simulated median residual warming is very similar to median warming without removing precipitation related variability in most cases. Consequently, we find less inconsistency between simulated and observed residual warming after removing all rainfall-related variability than with simulated and observed warming (not shown).

Evaluation of joint rainfall and temperature trends

Interannual variability of detrended rainfall and maximum and minimum temperature in Australia is correlated (Nicholls 2003). Observed correlation between daily maximum and minimum temperature is positive except for elevated inland areas in southern Australia in autumn and winter (not shown). Maximum temperature and rainfall are generally negatively correlated with correlations ranging from -0.92 to 0.34. That is, we expect cooler daily maximum temperatures in years with above-average rainfall and vice versa. The pattern is more mixed for rainfall and minimum temperature with mainly negative correlation in summer ranging from -0.81 to 0.44 and positive correlations in winter ranging from -0.23 to 0.77.

Taking into account observed and simulated correlations, we compute the standardised differences Z between simulated and observed joint rainfall and temperature changes (see earlier section). The fraction of models with joint trends significantly different from the observed are shown in Fig. 4. As expected, the areas where we find inconsistencies in joint trends across a majority of the 44 CMIP5 models correspond well with the areas with shared inconsistencies in the univariate assessments (Figs. 1i-l, 2i-l and 3m-p). The joint assessment of trends in multiple variables, however, is a more stringent test than the evaluation of trends in individual variables. The area of joint inconsistency is generally larger than the area of inconsistency of trends in any individual variable. Also, we find that correlation between co-located variables matters. The area of inconsistency is larger if correlation is taken into account in the computation of the multivariate standardised differences compared to if variables are assumed uncorrelated (not shown).

Figure 4. Percentage of models with significantly (at the 10% level) inconsistent joint temperature and precipitation changes.



Estimating the effect of internal variability

Estimates of the significance of differences in trends are contingent on the assumption that our estimate of the variability of trends is correct. We compare the standard deviation of trends derived directly from the forced runs (ΔA_{mod} in Eq. 3) to the standard deviation of 50-year trends in the control run of the 19 models for which at least 200 years of control run data are available (Table 1).

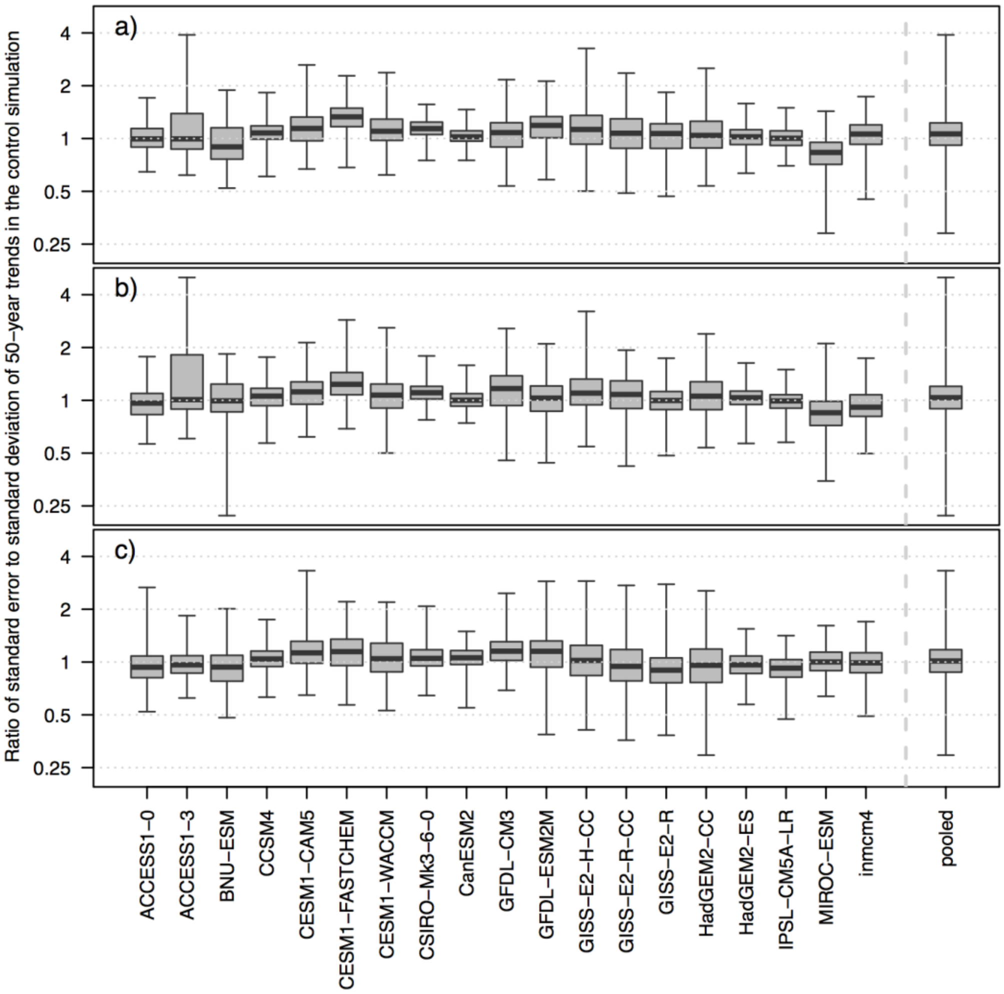
Figure 5 shows the ratio of the standard error of the regression (ΔA_{mod}) to the standard deviation of 50-year trends in the pre-industrial control run. The box plots summarise the ratios of variability in seasonal trends at individual grid boxes. For most models and all three variables, the ratios are centred about one and we find no indication of a consistent tendency towards under- or overestimation of the effect of internal variability on 50-year trends. The spread of ratios is considerable, however, with ratios ranging from 0.2 to 5. The central 50% (boxes in Figure 5) of the ratios of standard deviation of trends range from 0.9 to 1.2 for all three variables combined. In addition, further analysis shows that there is no consistent spatial nor seasonal pattern of under- or overestimation of uncertainty in 50-year trends among models (not shown).

Area-aggregate evaluation of simulated trends

Finally we present root mean squared standardised differences Z_{rms} as a robust, aggregate metric integrating differences in trends spatially and across seasons (Fig. 6). We use Z_{rms} as a summary metric of CMIP5 model performance in reproducing observed temperature and rainfall trends in Australia. The average standardised differences in joint temperature and rainfall trends range from 2 to 3 for the IPSL-CM5B-LR and

CESM1-WACCM models respectively with the remaining models almost equally distributed in between (Fig. 6a). When comparing Z_{rms} to the resampled Z_{rms} (reflecting random variability in Z_{rms} if we had a perfect model), we find that all but one (the IPSL-CM5B-LR model) of the 44 CMIP5 models analysed exhibit Z_{rms} that are significantly (at the 10% level) larger than resampled Z_{rms}. When analysing joint maximum and minimum temperature changes alone, we also find that most of the CMIP5 models are significantly different from a perfect model in how they represent regional trends from 1956 to 2005 (Fig. 6b). Residual joint regional temperature trends after removing rainfall-related variability, however, are much better reproduced. Z_{rms} for residual temperature trends (Fig. 6c) range from 1.5 to 2.0 and none of the 44 CMIP5 models is significantly different from a perfect model.

Figure 5. Root mean squared standardised differences Z_{rms} between joint temperature and rainfall trends (a), joint temperature changes (b), and joint residual temperature trends (c) over Australia averaged across the four seasons. The horizontal dashed lines indicate the respective 90th percentile of Z_{rms} from resampled model results reflecting the distribution of Z_{rms} given we had a perfect model (see earlier section).

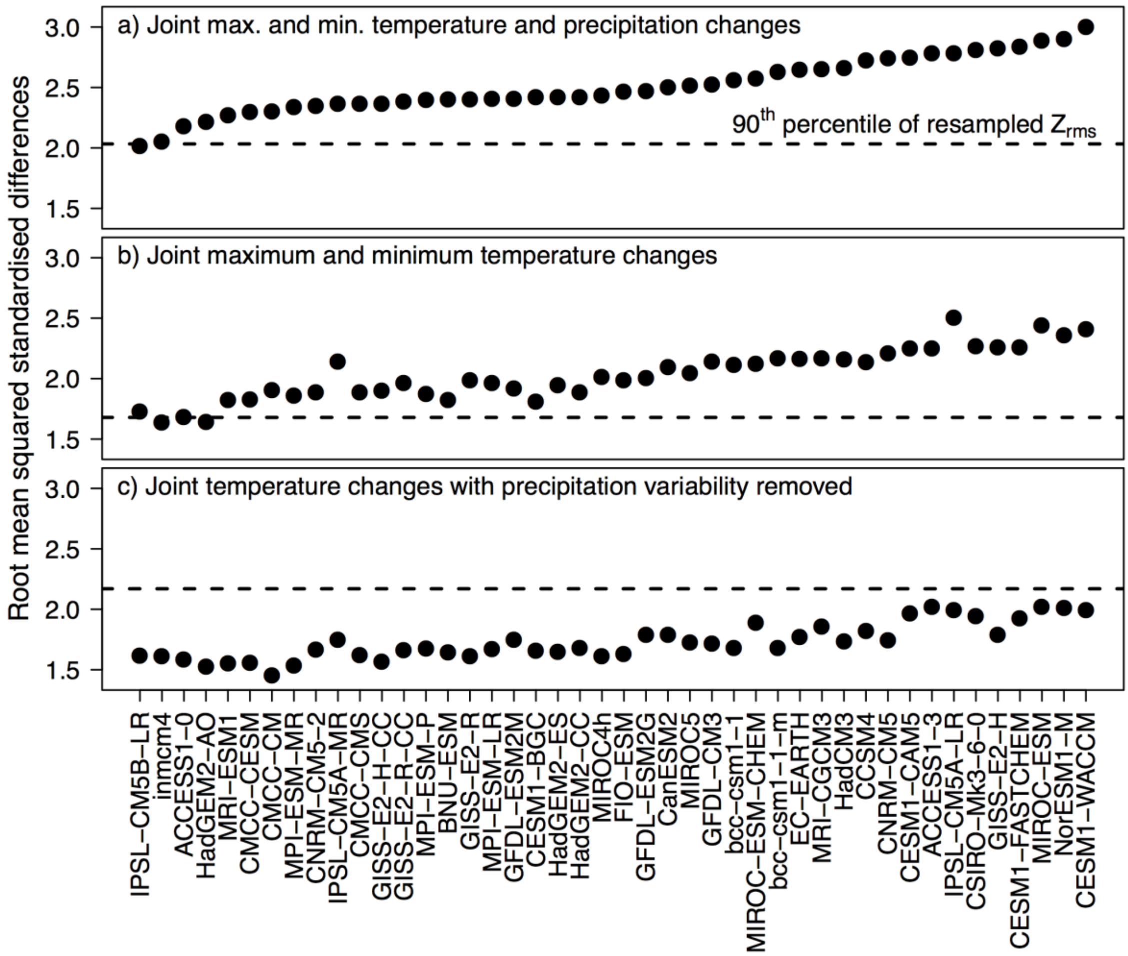


Discussion

We find significant differences between observed and simulated trends in seasonal mean daily maximum and minimum temperature (Figs. 1 and 2) and rainfall (Fig. 3) in some areas of Australia.

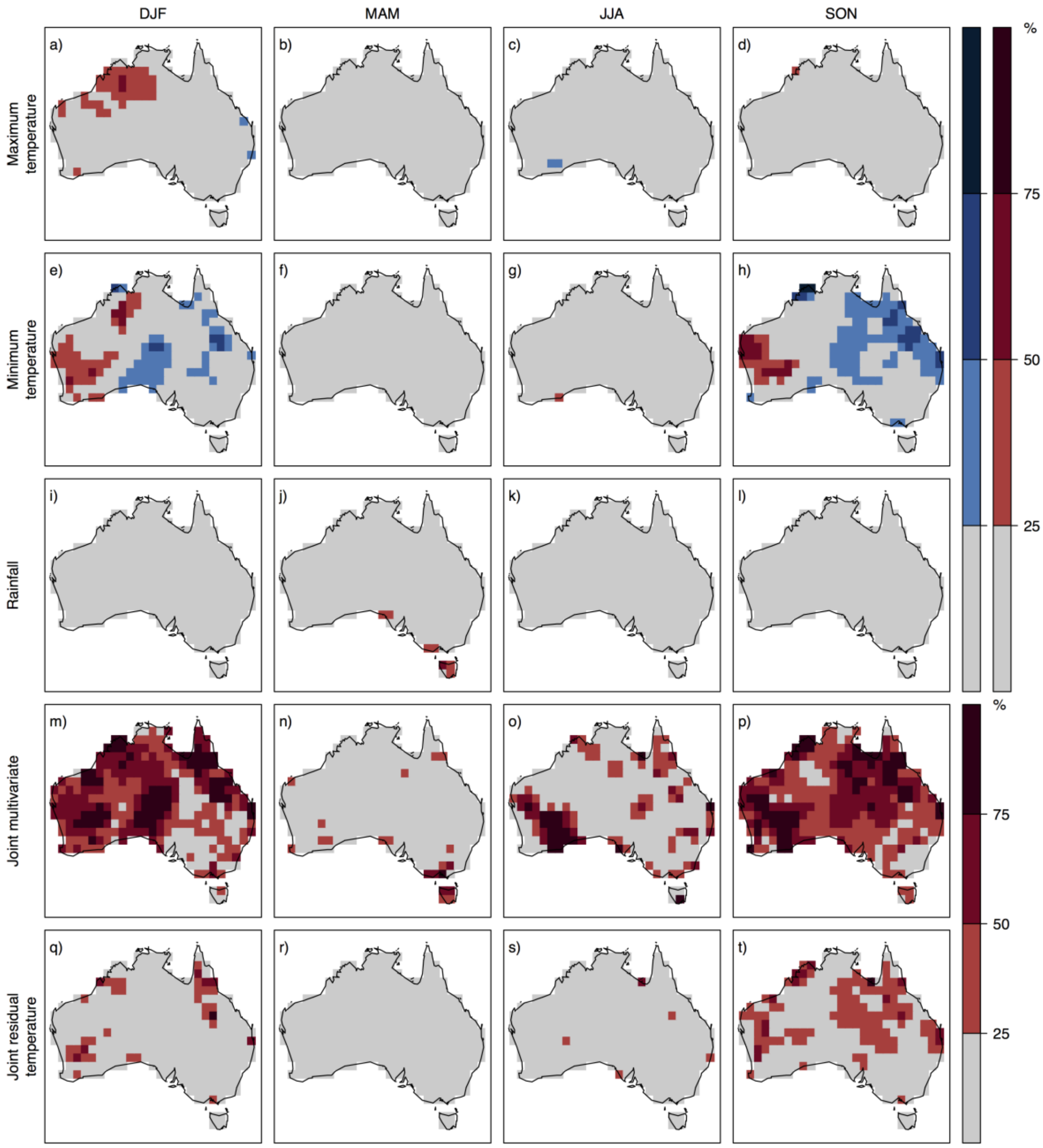
While previous studies have attributed area-mean warming in Australia (Karoly and Braganza 2005b; Min, Simonis, and Hense 2007) to human influences, this is to our knowledge the first spatially explicit evaluation of simulated trends in maximum and minimum temperatures for Australia. Karoly and Braganza (2005b) have found consistency between area-average maximum and minimum temperature trends in global climate models and observations. We, on the other hand, find evidence of significant differences between simulated and observed regional warming in Australia (Figs. 1 and 2), although only the areas of significant differences in trends in minimum temperature in spring and summer are larger than what we expect by chance (see Fig. 7e,h). Global climate models seem to under- and overestimate recent regional warming to roughly equal parts in these cases and area-average simulated trends are thus expected to be consistent with the observations. In contrast, if we summarise the consistency of simulated and observed temperature trends using the root mean squared standardised differences Z as shown in Figure 6, we find that simulated joint maximum and minimum temperature trends are significantly different from the observations in all models. Compared to analysing area-average warming, our approach incorporates spatial details in the regional warming pattern and we thus conclude that global climate models are able to reproduce the large-scale warming (Hegerl et al., 2007), but fail to reproduce the full spatial detail of the regional warming pattern as of yet.

Figure 6. Ratio of standard error of 50-year trends to standard deviation of 50-year trends in overlapping segments of the control simulation for seasonal daily maximum (a) and minimum (b) temperature and seasonal rainfall (c). In addition to the results for individual models, the pooled ratio from all models available is shown in the right-most column. The boxes and solid vertical line show the interquartile range and median of the distribution of ratios at individual grid boxes, the whiskers indicate the full range of the ratios.



Some of the spatial variability in recent warming in Australia may be related to natural variability. Although we account for the effect of natural variability in the analysis, our simple estimate of natural variability may be biased low as it is based on only 50 years of data and centennial-scale variability is not taken into account. Therefore, we also analyse residual temperature trends after removing all rainfall-related variability. By removing rainfall-related variability we intend to increase the signal-to-noise ratio of externally forced warming (Karoly and Braganza 2005a) based on the understanding that most of the recent rainfall variability is natural. We find that simulated residual warming is consistent with the observations (Fig. 6c). That is, global climate models reproduce the first-order effect of global climate change - the thermodynamic response to changes in the radiation balance - well. The first-order warming effect, however, is presumably a large-scale effect and therefore such evaluation provides little insight into regional specifics and regional model biases.

Figure 7. Percentage of models with trends from 1956 to 2005 that are significantly different from the observed trends when controlling the False Discovery Rate and thus taking into account multiple testing. Shown are results for seasonal mean daily maximum temperature (a-d), minimum temperature (e-h), rainfall (i-l), joint temperature and rainfall trends (m-p), and joint trends in temperature residuals after removing all rainfall-related variability (q-t).



Previous studies have highlighted deficiencies in the global climate models' ability to reproduce aspects of observed Australian rainfall variability and change (e.g. Shi et al., 2008; Brown, Jakob, and Haynes 2010; Kirono and Kent 2011; Cai et al., 2011; Colman, Moise, and Hanson 2011; Jourdain et al., 2013). Differences between simulated and observed trends in rainfall have been noted in Räisänen (2007) and Kirono and Kent (2011). Our assessment extends these findings by testing the significance of the difference between observed and simulated rainfall trends for individual seasons. Our findings are in line with previous studies in that the increase in summer (and annual) precipitation in north-western Australia is not reproduced in the simulations. In addition, our results also suggest that the significant observed drying in south-eastern Australia in autumn (MAM) is not reproduced in most of the CMIP5 simulations. However, while we find locally significant differences between simulated and observed rainfall trends in Australia in some regions (Fig. 3m-p), the area fraction of significant differences is not larger than what we expect by chance (Fig. 7i-l). We may thus conclude that although there appears to be indication of a misrepresentation of recent rainfall trends in simulations, this may be due to natural variability alone. It is important to note, however, that our assessment is based purely on statistical reasoning and additional insight into potential misrepresentations of processes in the climate models may lead us to revise the above conclusion (see for example Cai et al., 2011).

Robustness of the significance statements

The significance of differences in trends is contingent on our estimate of variance in trends due to sources not related to long-term climate change such as natural internal variability. Here we assume that we can estimate internal variability from residuals of 50-year trends (corrected for autocorrelation).

To analyse the validity of this assumption, we compare the estimate of the standard error of 50-year trends to the standard deviation of trends in the pre-industrial control simulations (Figure 5). Although the standard error differs considerably from the standard deviation of control run trends at individual grid boxes, there is no indication that we underestimate the variance in trends using the standard error of the regression in general. In contrast, the pooled estimate (right-most box in Figures 5a-c) is slightly biased positive indicating that the standard error is larger than the estimate of variability from the control runs on average. Therefore, we expect our estimates of the significance of differences in simulated and observed trends to be conservative.

Comparison with the distribution of control run trends reveals that our approach to estimate the effect of internal variability on 50-year trends is unbiased. The advantage of the estimate from regression residuals is that we can also estimate the same quantity from observations. While we focus on long-term changes here and while we do not formally compare simulated and observed standard errors and correlations between trends in different variables, we note that such an analysis would be very valuable to understand simulation biases with respect to interannual to decadal variability.

The spatial variability of trends is generally smaller in simulations than in the observations for daily minimum temperature and rainfall. The discrepancy in spatial variability of trends could be due to two main factors: i) The simulations produce variability (both or either forced and random variability) that does not exhibit enough small-scale structure. ii) Part of the spatial variability in the observations is due to the interpolation of station information (that is subject to local effects) to larger areas. We note that a careful evaluation of the spatial variability in simulations is beyond the scope of this paper, but we conclude that the discrepancy in the representation of variability in simulations and observations is an additional source of uncertainty not quantified in the current analysis.

Simulated cross-variable correlations in CMIP5 are similar to the ones observed (not shown). Furthermore, we find that the correlation identified using the residuals of detrended temperature and rainfall time series closely corresponds with correlation of 50-year trends computed from 50-year segments of the control run of the respective models (not shown). Trends in control runs only reflect internal variability as the boundary conditions are held constant in these experiments.

As we repeatedly test for significance at individual grid boxes, uncertainty estimates are affected by the problem of multiple testing. That is, we expect to reject the statistical test at a fraction of the grid boxes even if the null-hypothesis holds everywhere. We do not address multiple testing in the spatially explicit results in Figs. 1-4. Additional analysis, however, shows that when we correct for multiple testing by controlling the False Discovery Rate (Ventura et al., 2004) as in Bhend and Whetton (2013), most of the locally significant differences across Australia in daily maximum temperature trends and precipitation are not field or 'globally' significant (Fig. 7a-d,i-l). That is, we expect - by chance - to find areas with significant differences between simulated and observed trends in Australia that are as large as the ones identified. This is in line with Bhend and Whetton (2013) who find that differences in simulated and observed regional precipitation trends are not field significant based on a global analysis with a different set of observation data and for a different time period than in this study. In contrast to precipitation and daily maximum temperature, we find significant differences between observed and simulated trends in daily minimum temperature shared across models in summer (DJF, Fig. 7e) and spring (SON, Fig. 7h) even after correcting for the effect of multiple testing.

Multi-variate evaluation of trends

Evaluation of joint trends, finally, allows us to account for the effect of correlation between variables in the significance assessment. A positive bias in simulated precipitation trends in conjunction with a negative bias in simulated maximum temperature trends, for example, is less likely to be significant than positive trend biases in both precipitation and maximum temperature. As a consequence of the negative correlation between interannual variability in these variables, we expect trend biases of opposite sign due to natural variability alone. In other words, trend biases are less likely to be significant if they align with the major modes of natural variability.

We find that joint trend evaluation (Fig. 4) is a stricter test than evaluating individual variables one at a time. Also, when we control for multiple testing in the joint trend evaluation (Fig. 7m-p) remaining large areas with significant inconsistencies shared across models in winter (DJF) and spring (SON) are found.

Consequently, Z_{rms} of joint rainfall and temperature trends are significantly (at the 10% level) larger than the resampled statistics for all but one out of the 44 CMIP5 models analysed (Fig. 6). In the resampled statistics based on differences in trends between different initial condition ensemble members of the same model, spatial dependence and multiple testing is dealt with to the extent that simulated and observed spatial correlation and unforced variability is comparable. Hence, we conclude that global climate models are not able to fully reproduce the spatial details of the observed recent climate change in Australia.

It is important to note, however, that Z_{rms} for models with only one initial condition simulation exhibit much larger spread than Z_{rms} for models with multiple ensemble members. That is, a model is more likely to do well in reproducing observed trends by chance. Without access to additional initial condition members it is difficult to assess, unfortunately, what part of the inconsistency between simulated and observed trends is due to chance and what part is due to deficiencies in the model. For models with multiple initial condition runs it is much more unlikely that the correspondence of simulated and observed trends is by chance as internal variability in the different initial condition simulations is not in phase and thus tends to cancel out on average.

This internal variability uncertainty also affects our ability to distinguish between models that have some ability to reproduce recent observed change and those that do not. The distribution of Z_{rms} in Fig. 6 does not provide strong evidence that some models are markedly superior in reproducing observed trends. The spread of Z_{rms} is comparable to the spread of resampled Z_{rms} and thus the differences between models could be entirely due to internal variability rather than due to the varying ability of models to reproduce the recent climate change signal. Internal variability uncertainty, however, does not affect our conclusion that models are not able to reproduce the spatial details of the observed recent climate change in Australia.

Trend biases shared among models

The trend evaluation for Australia indicates that the CMIP5 models share trend biases. Summary statements on the performance of the ensemble as a whole, however, are affected by the codependence of individual models (Jun, Knutti, and Nychka 2008; Masson and Knutti 2011). Consequently, we report the results on a *per model* basis and refrain from evaluating the multi-model ensemble mean performance or to assess the reliability of the ensemble as in Annan and Hargreaves (2010), van Oldenborgh et al. (2013), and Yokohata et al. (2012).

The shading in Figs. 1i-l, 2i-l, 3m-p, and 4 starts at 25% (at least 11 of the 44 models). Assuming that the models are independent random draws from the population of models consistent with the real climate system, the probability of getting at least 22 models with significant (at the 10% level) inconsistencies out of 44 models (corresponding to 50% or more) is < 0.001. Even when considering that the models are likely dependent (see Masson and Knutti 2011 for a discussion of dependence in CMIP3 models) and when assuming that there are only 4 independent models, the probability of getting 3 or more significant results (corresponding to at least 50%) is 0.0037. This illustrates that no matter what the assumption on the independence of models, the areas with more than 50% of the models showing significant inconsistencies are likely not due to sampling in model space alone. Thus, we conclude that the CMIP5 simulations share trend biases for regional recent changes in rainfall and maximum and minimum temperature in Australia. This is in line with previous studies (Räisänen 2007; Haren et al., 2013; van Oldenborgh et al., 2013) indicating that the CMIP5 and other multi-model ensembles show regional trends that are outside the range of simulated trends in more regions than would be expected due to natural variability alone.

Conclusions

We present a comprehensive assessment of recent regional climate change in Australia. We compare externally forced change and take into account uncertainty due to natural internal variability in both the simulations and the observations. In addition, our analysis also includes an estimate of observation error. We evaluate simulated regional trends from 1956 to 2005 in seasonal mean daily maximum and minimum temperature and seasonal rainfall totals and find coherent areas with inconsistencies between simulated and observed trends in the majority of the models (Figs. 1-3).

Our assessment of simulated recent climate trends allows us to address two different questions: First, we are able to identify regions where and variables for which most of the CMIP5 models fail to reproduce the observed recent trends. Projections of future change based on the CMIP5 global models for these regions and variables should be flagged as less reliable if we understand what processes - i.e. the lack or misrepresentation thereof in simulations - causes the inconsistency between simulated and observed changes and if these processes are relevant for future projected change.

Second, the comprehensive nature of our analysis allows us to address the field significance of our results and also allows us to develop a more holistic metric of climate change evaluation by jointly evaluating changes across variables. We find that differences between simulated and observed changes in seasonal mean daily maximum temperature and rainfall are not field significant, whereas simulated trends in seasonal mean daily minimum temperatures in Australia are significantly different from the observed trends in spring and summer. Simulated joint rainfall and temperature changes across Australia from 1956 to 2005 are also significantly different from the observed trends. The analysis presented here illustrates the importance of moving from univariate model evaluation to more holistic metrics of model performance, as these allow for stricter tests.

The model evaluation results presented here could be used to constrain projections of future climate change. Additional research, however, is needed to identify what aspects of recent climate change provide useful constraints. Our summary evaluation for all of Australia does not provide evidence for strong constraints on projections as the difference between models in the ability to simulate recent trends is rather limited. In specific regions and seasons, on the other hand, evaluation of simulated trends may add to our understanding of projected future climate change.

Finally, we find trend biases shared across the majority of the CMIP5 models. This provides further evidence that we cannot assume independence in the CMIP5 multi-model ensemble (see also Jun et al., 2008; Masson and Knutti 2011; Pennell and Reichler 2011). Methods to produce regional climate change projections based on multi-model output thus have to account for dependence across the individual models in the ensemble at the regional scale to be able to produce reliable probabilistic projections.

Acknowledgements

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison (PCMDI) provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. The observation data have been obtained from the Bureau of Meteorology, the Climatic Research Unit at the University of East Anglia, and from the Australian Water Availability Project - a collaboration of the Bureau of Rural Sciences, CSIRO and the Bureau of Meteorology. We thank Seung-Ki Min, Ian Smith and an anonymous reviewer for very valuable comments to the manuscript.

References

Annan, J. D., and J. C. Hargreaves, 2010. Reliability of the CMIP3 Ensemble. *Geophys Res Lett*, 37, L02703, doi: 10.1029/2009GL041994.

Bhend, J., and P. Whetton, 2013. Consistency of Simulated and Observed Regional Changes in Temperature, Sea Level Pressure and Precipitation. *Climatic Change*, 118, 799–810. doi: 10.1007/s10584-012-0691-2.

Brown, J. R., C. Jakob, and J. M. Haynes, 2010. An Evaluation of Rainfall Frequency and Intensity over the Australian Region in a Global Climate Model. *J. Climate,* 23, 6504–6525. doi: 10.1175/2010JCLI3571.1.

Cai, W., A. Sullivan, T. Cowan, J. Ribbe, and G. Shi, 2011. Simulation of the Indian Ocean Dipole: A Relevant Criterion for Selecting Models for Climate Projections. *Geophys. Res. Lett., 38,* L03704. doi: 10.1029/2010GL046242.

Cai, W. and T. Cowan, 2012. Southeast Australia Autumn Rainfall Reduction: A Climate-Change-Induced Poleward Shift of Ocean-Atmosphere Circulation. *J. Climate,* 26, 189–205. doi: 10.1175/JCLI-D-12-00035.1.

Colman, R. A., A. F. Moise, and L. I. Hanson, 2011. Tropical Australian Climate and the Australian Monsoon as Simulated by 23 CMIP3 Models. *J Geophys Res,* 116, D10116. doi: 10.1029/2010JD015149.

CSIRO, 2012. *Climate and Water Availability in South-eastern Australia: A Synthesis of Findings from Phase 2 of the South Eastern Australian Climate Initiative (SEACI)*. Technical Report, CSIRO, Australia.

Frederiksen, C. S., J. S. Frederiksen, J. M. Sisson, and S. L. Osbrough, 2010. Changes and Projections in Australian Winter Rainfall and Circulation: Anthropogenic Forcing and Internal Variability. *International Journal of Climate Change: Impacts and Responses,* 2, 143–162.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008. Performance Metrics for Climate Models. *J. Geophys. Res.,* 113, D06104. doi: 10.1029/2007JD008972.

Haren, R., G. J. van Oldenborgh, G. Lenderink, M. Collins, and W. Hazeleger, 2013. SST and Circulation Trend Biases Cause an Underestimation of European Precipitation Trends. *Clim Dynam,* 40, 1–20. doi: 10.1007/s00382-012-1401-5.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2013. Updated High-resolution Grids of Monthly Climatic Observations – the CRU TS3.10 Dataset. *Int. J. Climatol.,* 34, 623–642.

Haylock, M. R., N. Hofstra, A. M. G. Klein-Tank, E. J. Klok, P. D. Jones, and M. New, 2008. A European Daily High-resolution Gridded Data Set of Surface Temperature and Precipitation for 1950-2006. *J. Geophys. Res.-Atmos.,* 113, D20119.

Hegerl, G. C., F. W. Zwiers, P. Braconnot, N. P. Gillett, Y. Luo, J.A. Marengo Orsini, J.E. Penner N. Nicholls, and P. A. Stott, 2007. *Understanding and Attributing Climate Change.* In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, 663–745. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Hegerl, G., and F. Zwiers, 2011. Use of Models in Detection and Attribution of Climate Change. *Wiley Interdisciplinary Reviews: Climate Change,* 2, 570–591. http://dx.doi.org/10.1002/wcc.121.

Indian Ocean Climate Initiative, 2012. *Western Australia's Weather and Climate: A Synthesis of Indian Ocean Climate Initiative Stage 3 Research.* Technical Report, CSIRO and Bureau of Meteorology, Australia.

Irving, D. B., S. E. Perkins, J. R. Brown, A. Sen Gupta, A. F. Moise, B. F. Murphy, L. C. Muir, R. Colman, S. Power, F. Delage, and J. Brown, 2011. Evaluating Global Climate Models for the Pacific Island Region. *Clim. Res.,* 49, 169–187.

Jones, D. A., W. Wang, and R. Fawcett, 2009. High-quality Spatial Climate Datasets for Australia. *Aust. Met. Oceaogr. J.,* 58, 233–248.

Jones, D. A., and B. C. Trewin, 2000. On the Relationships Between the El Niño–Southern Oscillation and Australian Land Surface Temperature. *Int. J. Climatol.,* 20, 697–719. doi: 10.1002/1097-0088(20000615)20:7<697::AID-JOC499>3.0.CO;2-A

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999. Surface Air Temperature and Its Changes over the Past 150 Years. *Rev. Geophys.,* 37, 173–199.

Jourdain, N. C., A. Sen Gupta, A. S. Taschetto, C. C. Ummenhofer, A. F. Moise, and K. Ashok, 2013. The Indo-Australian Monsoon and Its Relationship to ENSO and IOD in Reanalysis Data and the CMIP3/CMIP5 Simulations. *Clim. Dyn., 41,* 3073–3102. http://dx.doi.org/10.1007/s00382-013-1676-1.

Jun, M., R. Knutti, and D. W. Nychka, 2008. Spatial Analysis to Quantify Numerical Model Bias and Dependence. *J. Am. Stat. Assoc.,* 103, 934–947. doi: 10.1198/016214507000001265.

Karoly, D. J., and K. Braganza, 2005a. A New Approach to Detection of Anthropogenic Temperature Changes in the Australian Region. *Meteorol. Atmos. Phys.,* 89, 57–67.

Karoly, D. J., and K. Braganza, 2005b. Attribution of Recent Temperature Changes in the Australian Region. *J. Climate, 18,* 457–464. doi: 10.1175/JCLI-3265.1.

Kirono, D. G. C. and D. M. Kent, 2011. Assessment of Rainfall and Potential Evaporation from Global Climate Models and Its Implications for Australian Regional Drought Projections. *Int. J. Climatol.,* 31, 1295–1308. doi: 10.1002/joc.2165.

Larsen, S. H., and N. Nicholls, 2009. Southern Australian Rainfall and the Subtropical Ridge: Variations, Interrelationships, and Trends. *Geophys. Res. Lett.,* 36, L08708. doi: 10.1029/2009GL037786.

Masson, D., and R. Knutti, 2011. Climate Model Genealogy. *Geophys. Res. Lett.,* 38, L08703. doi: 10.1029/2011GL046864.

Min, S. K., and A. Hense, 2007. A Bayesian Assessment of Climate Change Using Multimodel Ensembles. Part II: Regional and Seasonal Mean Surface Temperatures. *J. Climate,* 20, 2769–2790.

Min, S. K., D. Simonis, and A. Hense, 2007. Probabilistic Climate Change Predictions Applying Bayesian Model Averaging. *Phil. T. Roy. Soc. A*, 365, 2103–2116.

Mitchell, T. D., and P. D. Jones, 2005. An Improved Method of Constructing a Database of Monthly Climate Observations and Associated High-resolution Grids. *Int. J. Climatol.,* 25, 693–712.

Narisma, G. T., and A. J. Pitman, 2003. The Impact of 200 Years of Land Cover Change on the Australian Near-Surface Climate. *J Hydrometeorol,* 4, 424–436. doi: 10.1175/1525-7541(2003)4<424:TIOYOL>2.0.CO;2.

Nicholls, N., 2003. Continued Anomalous Warming in Australia. *Geophys Res Lett,* 30, 1370. doi: 10.1029/2003GL017037.

Nicholls, N., 2004. The Changing Nature of Australian Droughts. *Climatic Change, 63,* 323–336. doi: 10.1023/B%3ACLIM.0000018515.46344.6d.

Nicholls, N., 2010. Local and Remote Causes of the Southern Australian Autumn-winter Rainfall Decline, 1958–2007. *Clim. Dyn., 34,* 835–845. http://dx.doi.org/10.1007/s00382-009-0527-6.

Noake, K., D. Polson, G. Hegerl, and X. Zhang, 2012. Changes in Seasonal Land Precipitation During the Latter Twentieth-century. *Geophys. Res. Lett.,* 39, L03706. doi: 10.1029/2011GL050405.

Pennell, C., and T. Reichler, 2011. On the Effective Number of Climate Models. *J. Climate,* 24, 2358–2367. doi: 10.1175/2010JCLI3814.1.

Polson, D., G. C. Hegerl, X. Zhang, and T. J. Osborn, 2013. Causes of Robust Seasonal Land Precipitation Changes. *J. Climate,* 26, 6679–6697. doi: 10.1175/JCLI-D-12-00474.1.

Randall, D. A., R. A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R. Stouffer, A. Sumi, and K. Taylor, 2007. *Climate Models and their Evaluation.* In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M.Tignor, and H. L. Miller, 589–662. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Räisänen, J., 2007. How Reliable Are Climate Models? *Tellus A,* 59, 2–29. doi: 10.1111/j.1600-0870.2006.00211.x.

Reichler, T., and J. Kim, 2008. How Well Do Coupled Models Simulate Today's Climate? *Bull. Amer. Met. Soc.,* 89, 303–311. http://dx.doi.org/10.1175/BAMS-89-3-303.

Risbey, J. S., M. J. Pook, P. C. McIntosh, M. C. Wheeler, and H. H. Hendon, 2009. On the Remote Drivers of Rainfall Variability in Australia. *Mon. Wea. Rev.,* 137, 3233–3253. doi: 10.1175/2009MWR2861.1.

Rotstayn, L. D., M. D. Keywood, B. W. Forgan, A. J. Gabric, I. E. Galbally, J. L. Gras, A. K. Luhar, G. H. McTainsh, R. M. Mitchell, and S. A. Young, 2009. Possible Impacts of Anthropogenic and Natural Aerosols on Australian Climate: a Review. *Int. J. Climatol.,* 29, 461–479. doi: 10.1002/joc.1729.

Shi, G., W. Cai, T. Cowan, J. Ribbe, L. Rotstayn, and M. Dix, 2008. Variability and Trend of North West Australia Rainfall: Observations and Coupled Climate Modeling. *J. Climate,* 21, 2938–2959. http://dx.doi.org/10.1175/2007JCLI1908.1.

Smith, I. N., and B. Timbal, 2012. Links Between Tropical Indices and Southern Australian Rainfall. *Int. J. Climatol.,* 32, 33–40. doi: 10.1002/joc.2251.

Smith, I., and E. Chandler, 2010. Refining Rainfall Projections for the Murray Darling Basin of South-east Australia—the Effect of Sampling Model Results Based on Performance. *Climatic Change,* 102, 377–393. doi: 10.1007/s10584-009-9757-1.

Smith, I., J. Syktus, C. McAlpine, and K. Wong, 2013. Squeezing Information from Regional Climate Change Projections-results from a Synthesis of CMIP5 Results for South-east Queensland, Australia. *Climatic Change,* 121, 609–619.

Stott, P. A., N. P. Gillett, G. C. Hegerl, D. J. Karoly, D. A. Stone, X. Zhang, and F. Zwiers, 2010. Detection and Attribution of Climate Change: a Regional Perspective. *Wiley Interdisciplinary Reviews Climate Change,* 1, 192–211.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2011. An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Met. Soc., 93,* 485–498. doi: 10.1175/BAMS-D-11-00094.1.

Timbal, B., and J. M. Arblaster, 2006. Land Cover Change as an Additional Forcing to Explain the Rainfall Decline in the South West of Australia. *Geophys. Res. Lett.,* 33, L07717. doi: 10.1029/2005GL025361.

Timbal, B., J. M. Arblaster, and S. Power, 2006. Attribution of the Late-Twentieth-Century Rainfall Decline in Southwest Australia. *J Climate, 19,* 2046–2062. doi: 10.1175/JCLI3817.1.

Trewin, B, 2013. A Daily Homogenized Temperature Data Set for Australia. *Int. J. Climatol.,* 33, 1510–1529.

van Oldenborgh, G. J., S. Drijfhout, A. van Ulden, R. Haarsma, A. Sterl, C. Severijns, W. Hazeleger, and H. Dijkstra, 2009. Western Europe Is Warming Much Faster Than Expected. *Clim. Past.,* 5, 1–12.

van Oldenborgh, G. J., F. J. Doblas Reyes, S. S. Drijfhout, and E. Hawkins, 2013. Reliability of Regional Climate Model Trends. *Environmental Research Letters*, 8, 014055.

Ventura, V., C. J. Paciorek, and J. S. Risbey, 2004. Controlling the Proportion of Falsely Rejected Hypotheses When Conducting Multiple Tests with Climatological Data. *J. Climate,* 17, 4343–4356. doi: 10.1175/3199.1.

Watterson, I. G, 2008. Calculation of Probability Density Functions for Temperature and Precipitation Change Under Global Warming. *J. Geophys. Res.-Atmos.,* 113, D12106.

Wilks, D. S. 1997. Resampling Hypothesis Tests for Autocorrelated Fields. *J. Climate*, 10, 65–82.

Yokohata, T., J. Annan, M. Collins, C. Jackson, M. Tobis, M. Webb, and J. Hargreaves, 2012. Reliability of Multi-model and Structurally Different Single-model Ensembles. *Clim. Dyn.,* 39, 599–616. doi: 10.1007/s00382-011-1203-1.

Zhang, X., F. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and Nozawa. T, 2007. Detection of Human Influence on Twentieth-century Precipitation Trends. *Nature,* 448, 461–465.