Inability of CMIP5 Models to Simulate Recent Strengthening of the Walker Circulation: Implications for Projections

Article *in* Journal of Climate · January 2015

DOI: 10.1175/JCLI-D-13-00752.1

CITATIONS
85

READS
103

2 authors:

Greg Kociuba
Bureau of Meteorology
12 PUBLICATIONS **616** CITATIONS

SEE PROFILE

Scott Power
Bureau of Meteorology
129 PUBLICATIONS **9,406** CITATIONS

SEE PROFILE

Inability of CMIP5 Models to Simulate Recent Strengthening of the Walker Circulation: Implications for Projections

GREG KOCIUBA AND SCOTT B. POWER

Centre for Australian Weather and Climate Research, Bureau of Meteorology, Melbourne, Victoria, Australia

(Manuscript received 2 December 2013, in final form 9 July 2014)

ABSTRACT

This paper examines changes in the strength of the Walker circulation (WC) using the pressure difference between the western and eastern equatorial Pacific. Changes in observations and in 35 climate models from the Coupled Model Intercomparison Project (CMIP) phase 5 (CMIP5) are determined. On the one hand, 78% of the models show a weakening of the WC over the twentieth century, consistent with the observations and previous studies using CMIP phase 3 (CMIP3) models. However, the observations also exhibit a strengthening in the last three decades (i.e., from 1980 to 2012) that is statistically significant at the 95% level. The models, on the other hand, show no consensus on the sign of change, and none of the models shows a statistically significant strengthening over the same period. While the reasons for the inconsistency between models and observations is not fully understood, it is shown that the ability of the models to generate trends as large as the observed from internal variability is reduced because most models have weaker than observed levels of both multidecadal variability and persistence of interannual variability in WC strength.

In the twenty-first-century future projections, the WC weakens in 25 out of 35 models, under representative concentration pathway (RCP) 8.5, 9 out of 11 models under RCP6.0, 16 out of 18 models under RCP4.5, and 12 out of 15 models under RCP2.6. The projected decrease is also consistent with results obtained previously using models from CMIP3. However, as the reasons for the inconsistency between modeled and observed trends in the last three decades are not fully understood, confidence in the model projections is reduced.

1. Introduction

The Walker circulation (WC) is one of the world's most prominent and important atmospheric systems (e.g., Gill 1982). It extends across the entire tropical Pacific Ocean, encompassing 1) the trade winds blowing from east to west; 2) air forced to rise over the western Pacific, Southeast Asia, and northern Australia through enhanced convection; 3) winds blowing counter to the trades aloft; and 4) air descending over the eastern Pacific Ocean (Power and Kociuba 2011b).

The WC exhibited a weakening trend between periods beginning 1900–50, ending at 2005 (Tanaka et al. 2004; Vecchi et al. 2006; Meehl et al. 2007; Power and Smith 2007; Collins et al. 2010; Vecchi and Wittenberg 2010; Nicholls 2008; Power and Kociuba 2011b; Tokinaga et al. 2012a,b). Vecchi et al. (2006) found that without anthropogenic forcing, the trends in the models from the Coupled Model Intercomparison Project (CMIP) phase 3 (CMIP3) were much lower than the observations. They also examined preindustrial runs, and found the magnitude of natural, internally generated trends to be much smaller than the observational trend, concluding that internal variability alone cannot account for the observed trend.

Power and Kociuba (2011b) examined the WC strength in CMIP3 models and found that 15 out of 23 models showed a weakening of the WC during the twentieth century (1900–99). External forcing was estimated to account for 50% ± 20% of the observed weakening. A stronger weakening was found during the twenty-first century in 13 out of 21 models for Special Report on Emissions Scenarios (SRES) A1B, and 14 out of 17 models for SRES A2.

Vecchi et al. (2006) argued that the weakening of the WC arises as follows: Warming from increases in greenhouse gases increases global water vapor amount by approximately 7% $°C^{-1}$, in agreement with the Clausius–Clapeyron relationship (Soden et al. 2005). However, the precipitation response increases more slowly and ranges

Corresponding author address: Dr. Greg Kociuba, Bureau of Meteorology, GPO Box 1289, Melbourne VIC 3001, Australia. E-mail: g.kociuba@bom.gov.au

DOI: 10.1175/JCLI-D-13-00752.1

from 2% to 3% $°C^{-1}$ (Vecchi and Soden 2007; Held and Soden 2006). To balance the latent heat of precipitation and the radiative cooling in the global tropics, the circulation must slow down in these models. Strictly speaking, this is for the global mean circulation and does not necessarily apply specifically to the WC. Also, the model convective mass flux does not necessarily align with the WC (Sandeep et al. 2014). Another pioneering study is that of Knutson and Manabe (1995): the dry static stability increases more than the radiative cooling so that the circulation has to slow down in response to greenhouse gases.

Possible mechanisms for why the precipitation responds more slowly than the water vapor based on the Clausius–Clapeyron relationship were studied in some Fourth Assessment Report (AR4) models (Stephens and Ellis 2008). Globally, it was found that the atmosphere cannot emit radiation at a large enough rate to support precipitation due to the increase in water vapor. The efficiency of precipitation from water vapor is negatively affected by cloud radiative heating due to a reduction in the amount of cloud in the middle troposphere, and also a global reduction in sensible heating.

While weakening is evident over the longer time scales described above, some observational studies show that the WC has strengthened over recent decades (Sohn et al. 2013; Sohn and Park 2010; Meng et al. 2012; L'Heureux et al. 2013; Solomon and Newman 2012; Luo et al. 2012). For example, Sohn et al. (2013) found statistically significant trends in mean sea level pressure (MSLP), sea surface temperature (SST), and wind and water vapor transport linked to the WC. Sohn et al. (2013) concluded that decadal variations in El Niño in recent decades strengthened the WC. They also examined changes in the WC in 21 CMIP3 models for the historical period (1979–99), and they found no agreement on the sign of the change.

Sohn et al. (2013) concluded that the failure to capture the recent strengthening in CMIP3 models is mostly due to problems in simulating the distinct eastern, and central Pacific El Niños (Ashok and Yamagata 2009; Ham and Kug 2012; Power et al. 2013). One driving force of WC strength is the equatorial Pacific sea surface temperature gradient, which also increased in recent decades. This was tested in an atmospheric general circulation model (AGCM) forced by the historical sea surface temperature (Meng et al. 2012). Meng et al. (2012) found that an increased SST gradient led to an enhancement of both the Walker and Hadley circulations.

SST-forced AGCM experiments were also conducted by Tokinaga et al. (2012a,b). However, their experiments were conducted over the longer period 1950–2009, and a weakening of the WC resulted. L'Heureux et al. (2013) examined multidecadal trends of MSLP starting from 1900 in 10 different datasets, and they identified a strengthening of the WC in all 10 in the more recent period. Solomon and Newman (2012) used a statistical method to reduce the impact of ENSO on various SST and SLP reconstructions, and they then found no weakening of the WC over the period 1900–2010 in the residual data.

The WC index was also examined in 37 models from phase 5 of CMIP (CMIP5; 101 runs) where trends were calculated over the period 1870–2004 (DiNezio et al. 2013). The observations exhibited a weakening over this period. A weakening was found in 25% of the runs that lie within the 95% confidence interval of the observed value. A positive trend ($1.7\,Pa\,yr^{-1}$) was also identified in the observations during 1980–2004, and it was suggest that the positive trend is likely due to multidecadal internal variability.

The recent warming hiatus (2001–12) is also related to the strengthening of the WC (England et al. 2014). England et al. (2014) suggest that there has been an acceleration of the jets and an increase in the wind-driven Ekman divergence away from the equator. Half of the wind stress trend during 1992–2011 is associated with the interdecadal Pacific oscillation (IPO; Power et al. 1998; Folland et al. 2002; Meehl and Arblaster 2012; Meehl et al. 2013). The SST trend pattern is IPO-like and consistent with the strengthening of the trade winds (England et al. 2014). There has also been a shift in the western tropical Pacific sea level trend during the 1990s (Merrifield 2011).

In this study, we update and extend this earlier work by examining trends in the strength of the WC in version 2 of the Hadley Centre sea level pressure dataset (HadSLP2) up to 2012, and in the recently released CMIP5 models. We examine the ability of the models to reproduce the observed changes over 1980–2012 and the extent to which the strengthening might have been driven by external forcing and internally generated natural variability. We also examine the projected changes in strength of the WC in the twenty-first century, and compare these projected changes with those obtained previously using CMIP3 models (Power and Kociuba 2011b). Implications of the ability of models to simulate observed trends are considered.

The metric used to calculate the WC is described in section 2. The climate models and analysis methods are described in section 3. We examine trends in the observations and model simulations for the period 1900–2012 in section 4. Trends for the recent 33-yr period 1980–2012 are examined in section 5. This includes discussion on the magnitude of model variability in 33-yr trends from historical and preindustrial runs. Factors influencing the ability of models to simulate internally driven trends in the historical and preindustrial runs are examined in section 6. The future projected trends for the period 2013–2100 for the representative concentration pathway (RCP) 8.5 scenario are discussed and

compared with the CMIP3 model results in section 7. Results are summarized and discussed in section 8.

2. Index for the Walker circulation

We follow previous studies (Vecchi et al. 2006; Power and Kociuba 2011b) and use an index for the WC based on the difference between equatorial mean sea level pressure in a western box (5°S–5°N, 80°–160°E) and an eastern box (5°S–5°N, 160°–80°W). We will refer to these areal averages of MSLP as BoxW and BoxE respectively. The arithmetic difference BoxΔP = BoxE − BoxW is used as a proxy for the strength of the WC. The data for these indices are derived from the HadSLP2r (single) dataset from the Met Office (Allan and Ansell 2006). This was extended up to 2012 where the variance of the period 2006–12 was reduced by the Met Office, in order to be consistent with HadSLP2 (http://www.metoffice.gov.uk/hadobs/hadslp2/data/HadSLP2r_lowvar_description.doc). A summary of annual trends will be presented.

3. Climate models and analysis

We analyze the twentieth- and twenty-first-century integrations from numerous coupled general circulation models available from the World Climate Research Programme (WCRP)–Climate Variability and Predictability (CLIVAR)–Working Group on Coupled Modelling (WGCM) Coupled Model Intercomparison Project. In the preindustrial experiment, the greenhouse gases (GHGs), aerosols, ozone, and solar irradiance are fixed at the year 1850. The historical simulations have time-dependent external forcing, which includes GHGs, volcanic activity, the solar constant, ozone, and aerosols. The forcing data over the period 1850–2005 are taken from observations. Further details about the experiments are described by Taylor et al. (2012).

We begin by investigating the twentieth-century results and compare them with the observations. We then examine twenty-first-century changes in runs forced using the representative concentration pathways. Here we focus on RCP8.5 (which has as a radiative forcing of approximately 8.5 $W m^{-2}$ by 2100). Thirty-five models are analyzed (the first run per model, i.e., r1i1p1). All the historical runs begin at 1900 and end at 2005. We extended this to 2012 by using RCP8.5 over the period 2006–12 so that direct comparisons can be made with the observational period. We found that our results were not sensitive to the choice of the 7-yr patching, as we obtained the same conclusions when extended by RCP2.6, RCP4.5, RCP6.0, or RCP8.5. In this paper, we define the historical period to be 1900–2012. All the RCP scenarios cover 2006–2100. The twenty-first-century calculations here are based on the period 2013–2100 so that there is no overlap between the historical period and the twenty-first-century period analyzed. We define the twenty-first-century period to be 2013–2100. All our analysis of historical and preindustrial models is performed using annual data (January–December). In section 5b(3) we use 32 climate models for the preindustrial runs that have lengths ranging from 200 to 1156 years available to us (Table 2). In section 7, we examine twenty-first-century trends (2006–2100) in 20 RCP8.5 models and compare the results to previous CMIP3 results.

The statistical significance of trends presented below takes persistence into account (Power et al. 1998). The method used for calculating statistical significance, unless otherwise stated, is described in appendix A. We will refer to this as the P98 method.

The numerical values of significance provided below are defined as 100 $(1 - 2\alpha)$ where α is the probability of obtaining a trend by chance. At the 95% level α is 0.025 for a two-sided t test. Two trends are defined as statistically different at the 95% level if the slope of the trend difference is statistically different from zero at the 95% level (i.e., $\alpha \leq 0.025$).

4. Trends in observations and models for 1900–2012

We examine trends in observations and models over the full period 1900–2012 using two different sets of indices. In section 4a we examine trends in the box indices BoxW, BoxE, and the difference BoxΔP. In section 4b we examine trends in the station data at Darwin, Tahiti (which lie just south of BoxW and BoxE respectively), and the difference Tahiti minus Darwin. This provides two different ways of comparing the consistency between the model trends with the observed. It is known that station data from Darwin are more stable in time than a box index (Bunge and Clarke 2009) and the inclusion of Darwin MSLP in our analysis allows us to examine the possible role of observational error.

a. *Trends in BoxΔP, BoxW, and BoxE*

Observed (HadSLP2) and modeled trends in BoxΔP for the entire historical period 1900–2012 are shown in Fig. 1. Bars depicting observed trends are either a green (if positive) or pink (if negative). Note that a negative (positive) value in BoxΔP means that the WC is weakening (strengthening). Positive (negative) values in BoxW or BoxE mean that there is a trend toward higher (lower) pressure, and therefore sinking (ascending) motion. Results from this section are also summarized in Table 1.

The observed trends for BoxW ($+0.06$ Pa yr^{-1}) and BoxΔP (-0.2 Pa yr^{-1}) are not statistically significant, while the trend in BoxE (-0.13 Pa yr^{-1}) is only significant

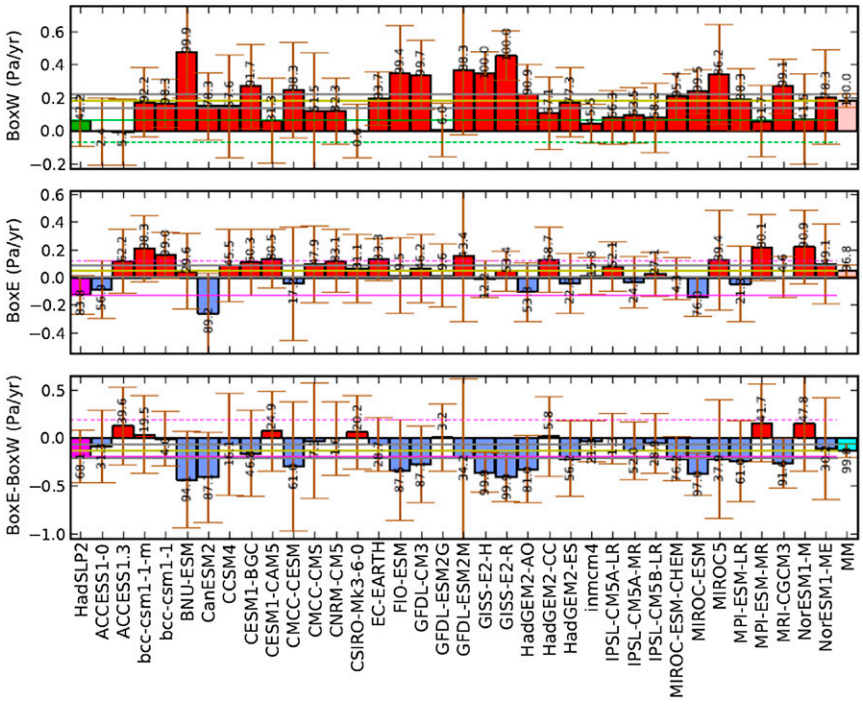

FIG. 1. Observed and modeled twentieth-century linear trends (Pa yr^{-1}) for each model listed in annual (top) BoxW (5°S–5°N, 80°–160°E), (middle) BoxE (5°S–5°N, 160°–80°W), and (bottom) BoxΔP (=BoxE − BoxW). The bars to the left of the plot are the observed (HadSLP2) trends. They are in green if the trend is positive and pink if the trend is negative. The dashed lines are the negative of the observed trends. The remaining bars represent a single run for each model. These bars are red if they are positive and blue if they are negative. The horizontal yellow line gives the multimodel mean (MM). The 95% confidence interval for MM is shown as gray horizontal lines. Trends are calculated for the period 1900–2012. All historical model runs were extended from 2006 to 2012 using RCP8.5 data. The confidence intervals are at the 95% level. The numbers on the plot are the probabilities of the trends not occurring by chance taking persistence into account.

at the 84% level for the period 1900–2012. This assessment is based on the method described by Power et al. (1998) in which persistence is taken into account, referred to here as the P98 method (see appendix A for details).

The bars depicting trends in each of the 35 models are red if positive or blue if negative. The multimodel mean (MMM) is the rightmost bar on the plot (labeled MM), and a yellow horizontal line is extended across all models at this value. The MMM is a model-based estimate of the externally forced component, consisting of natural and anthropogenically forced components.

In most models, the 1900–2012 trend in BoxW is larger than the observed positive trend, and the trend in nine of these models is significant at the 95% level (i.e., there is less than 5% chance of obtaining the trend by chance taking persistence into account). The MMM trend in BoxW (0.18 Pa yr^{-1}) is significant at the 99.99% level.

As noted above, the observed negative trend in BoxE is only statistically significant at the 84% level. Only two of the model trends in Box E are significant at the 95% level, but since there are 25 positive and 10 negative trends, it turns out that the MMM trend is small (0.05 Pa yr^{-1}) but significant at the 95% level.

Most models show a negative trend in BoxΔP, and three of these models show a negative trend at the 95% level. Despite the large internal variability, 27 out of 35 models have negative trends for BoxΔP consistent with a weakening of the WC over the period 1900–2012. Using a binomial distribution assuming that the chance of any individual trend being positive or negative is equally likely and are both 0.5 (see Power and Kociuba 2011a,b), then this result (i.e., 27 out of 35 models with trends of same sign) is significant at the 99.97% level, under the assumption that the models are independent (Power and Kociuba 2011b). The MMM trend for BoxΔP is significant at the 99.7% level according to the P98 method.

The MMM trend in BoxΔP is similar to the observed value. While, most of the models have a trend in BoxE with opposite sign to the observations, the differences between BoxE trends in the models and the observations

are not statistically significant at the 95% level. In summary, there is no firm evidence of inconsistency between models, taken as a whole, and the observations over the full period 1900–2012.

b. *Trends in station data (1900–2012)*

The previous method is now applied to Darwin, Tahiti, and Tahiti minus Darwin, and the results are also summarized in Table 1.

We begin by comparing observed trends in the Box indices with observed trends in Darwin and Tahiti MSLP. Darwin MSLP is known to have data coverage that is more stable in time than BoxW (Bunge and Clarke 2009), and hence might be considered more reliable than BoxW over the 1900–2012 period. The observed trend for Darwin ($0.28\,\mathrm{Pa\,yr^{-1}}$) is larger than BoxW ($0.06\,\mathrm{Pa\,yr^{-1}}$). However, neither trend is statistically significant, nor is the trend difference. The consistency between the trend at Darwin and BoxW adds confidence to the reliability of the BoxW trend.

The observed trends at Tahiti and BoxE are opposite in sign, but neither trend is statistically significant at the 95% level. The observed trend in Tahiti minus Darwin ($-0.14\,\mathrm{Pa\,yr^{-1}}$) is slightly smaller than the observed trend in BoxΔP ($-19\,\mathrm{Pa\,yr^{-1}}$). Again results seem broadly consistent.

We now compare the consistency between the model trends and observations for Darwin and Tahiti. The MMM trend is about half of the observed positive trend for Darwin. Most of the models (26 out of 35) have the same sign as the trend in Darwin MSLP. The MMM trend in Tahiti MSLP is approximately equal to its observational counterpart, although the observed trend is not significant. There are 27 models that have the same sign as the observed trend. In the models there is no consensus of the sign of the trend (18 positive and 17 negative).

The analysis of the station data indicates that the observed trend at BoxW is smaller than the observed trend at Darwin, and both are positive. This adds some confidence to the reliability of the BoxW data. Most of the model trends agree with the sign of the trend for Darwin and Tahiti.

5. 33-yr trends

In this section, we first examine the trend in the box indices BoxΔP, BoxW, and BoxE for the recent 33-yr period 1980–2012 in the observations. These trends are compared to the trends over the same period in the historical climate models. This analysis is repeated for Tahiti minus Darwin, Darwin, and Tahiti, so that the consistency between model trends and the station data observations can be tested, and the results that follow

TABLE 1. Multimodel ensemble-mean sign consistency with observations. The numbers of positive and negative trends are counted for index 1 and index 2 and are displayed as boldface and italic numbers, respectively, in parentheses. Thirty-five models were used, and the magnitude of the ensemble mean (MM), its 95% confidence interval, and the associated probability of a trend not obtained by chance taking into account persistence following in parentheses, as defined in section 3, are displayed for each index. The same information is displayed for the observations (obs) corresponding to the associated indices.

	Trends ($\mathrm{Pa\,yr^{-1}}$)					
	BoxW	Darwin	BoxE	Tahiti	BoxΔP	Tahiti minus Darwin
1900–2012	(**32+**, **3−**)	(*26+*, *10−*)	(**25+**, **10−**)	(*28+*, *8−*)	(**8+**, **27−**)	(*19+*, *17−*)
MM	0.18 ± 0.04 (99.99%)	0.12 ± 0.05 (99.96%)	+0.05 ± 0.04 (97%)	0.12 ± 0.05 (98%)	−0.13 ± 0.07 (99.8%)	0.009 ± 0.08 (14%)
Obs	0.06 ± 0.2 (42%)	0.28 ± 0.4 (76%)	−0.13 ± 0.14 (84%)	0.14 ± 0.4 (48%)	−0.19 ± 0.3 (68%)	−0.14 ± 0.6 (27%)
1980–2012	(**25+**, **10−**)	(*23+*, *13−*)	(**25+**, **10−**)	(*32+*, *4−*)	(**17+**, **18−**)	(*29+*, *7−*)
MM	+0.50 ± 0.3 (96%)	0.39 ± 0.4 (88%)	0.34 ± 0.3 (94%)	1.02 ± 0.4 (98%)	−0.16 ± 0.6 (38%)	0.62 ± 0.7 (85%)
Obs	−1.74 ± 1.1 (94%)	−1.32 ± 2.2 (67%)	1.3 ± 0.9 (97%)	3.6 ± 2.3 (96%)	3.04 ± 1.9 (95%)	4.9 ± 4 (90%)

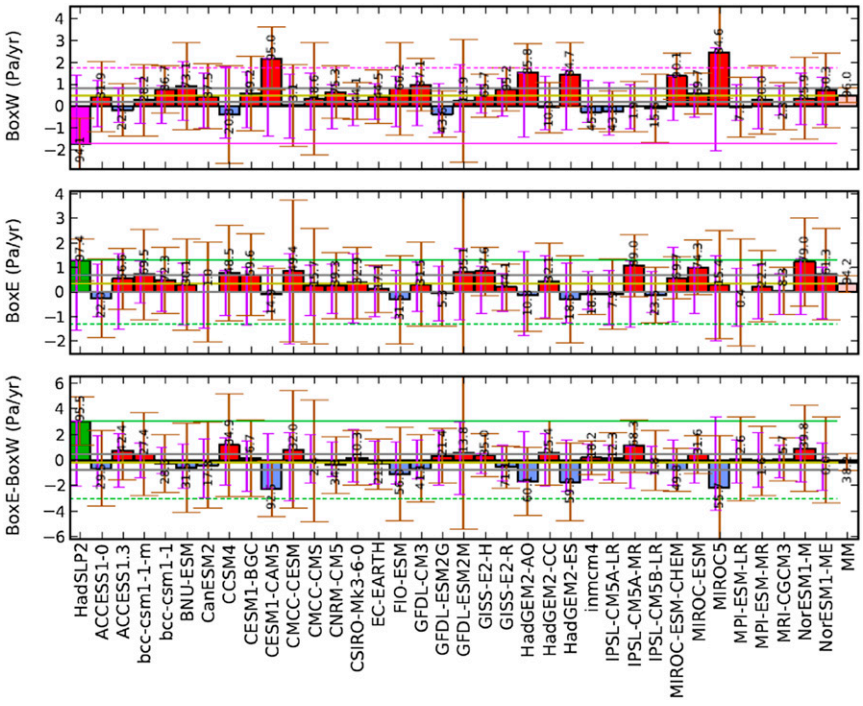

FIG. 2. As in Fig. 1, but for the years 1980–2012. The observational dataset HadSLP2 (1900–2012) was used. The confidence intervals are at the 95% level and the error bars are centered (brown) on each model name. The purple error bars to the left of the centered bars indicate the largest positive and negative 33-yr trends that can occur anywhere over 1900–2012.

are compared to the corresponding box indices. We then examine how unusual the observed recent 1980–2012 trend is by calculating all possible 33-yr trends over the 1900–2012 period. This approach is also applied to the historical models to determine whether trends larger than observed could be achieved at other 33-yr periods. Last, we examine how frequently the observed 1980–2012 trend could occur from internal variability alone, by examining trends in the preindustrial integrations.

a. Observed and model trends over the period 1980–2012

1) TRENDS IN BOX INDICES IN OBSERVATIONS AND MODELS (1980–2012)

Trends over the recent period 1980–2012 are shown as the first bar in Fig. 2. The trend in BoxΔP for the HadSLP2 dataset is 3.0 Pa yr^{-1}, which is statistically significant at the 95% level. The trends in BoxW (-1.7 Pa yr^{-1}) and BoxE (1.3 Pa yr^{-1}) are statistically significant at the 94% and 97% level respectively.

The model trends displayed in Fig. 2 for BoxΔP are all smaller in magnitude than the corresponding observations, and only one model has a trend that is statistically significant above the 90% level. Nor is there any consensus on the sign of the trends among the models, as there are an approximately equal number of positive and negative trends. The MMM trend in BoxΔP (0.16 Pa yr^{-1}, yellow line in Fig. 2) is not statistically significant. It is small compared to the observed trend (3.0 Pa yr^{-1}), which indicates that any externally forced component is very small compared with the observed trend in BoxΔP over the 1980–2012 period. The models have uncertainties in the trend as indicated by the (brown) confidence intervals in Fig. 2. So if this is taken into account when comparing the model trends in BoxΔP with the observed trend, we find that only 11 models have 95% confidence intervals that encompass the observed trend for BoxΔP.

Another approach to assessing consistency between model and observed trends is to calculate the statistical significance of the difference between the model trends and the observed trend. We find that six models have trends in BoxΔP that are different from the observed at the 95% level. Also, 14 models have trends statistically different from the observed trend at the 90% level. In addition, the difference between the MMM trend in BoxΔP and the observed trend is statistically significant at the 95% level. This suggests that there is an inconsistency between the model and observed trends in BoxΔP.

Among the model trends for BoxW (MMM 0.5 Pa yr^{-1}) there are 25 large positive values and 10 small negative values. The statistical significance of this asymmetry in the sign of trends is 99.7% (assuming a binomial distribution

as described previously). Using the P98 method the significance becomes 95%.

The statistical significance of the difference between the BoxW trends in the individual model and the observations is now considered. We find that 10 models have trends statistically different to the observed, at the 95%, while an additional five models have trends that are statistically significant at the 90% level only. The MMM trend in BoxW is statistically different from the observed trend at the 95% level. Furthermore, the sign of the trend in BoxW is opposite to the sign of the observed trend. This is in contrast to the MMM trend of BoxE ($0.34\,\mathrm{Pa\,yr^{-1}}$), which has the same sign as the observed trend ($1\,\mathrm{Pa\,yr^{-1}}$). There is more agreement between the model trends in BoxE and the observed, since only one model has a trend that is statistically significantly different to the observed trend at the 95% level. Note that the observed trend in BoxE lies within the confidence intervals of 28 of the 35 models. Twenty-five out of 35 models have a positive trend, which is significant at the 99.7% level using a binomial distribution as before. This result remains statistically significant (at the 95% level) if the significance is calculated using the P98 method. The trend results for the box indices are summarized in Table 1.

In summary, models cannot capture the 1980–2012 strengthening of the Walker circulation using the equatorial box difference index BoxΔP, even if confidence intervals are considered. The primary reason for this arises in the simulation of the trend in BoxW, rather than the trend in BoxE.

2) Trends in MSLP at Darwin and Tahiti in observations and models (1980–2012)

We repeat this analysis using Darwin and Tahiti to test the ability of the models to simulate the station data, and to make broad comparisons between the station data and the box index results.

The trend for Darwin is also negative ($-1.3\,\mathrm{Pa\,yr^{-1}}$), but the MMM trend is a smaller positive value ($0.4\,\mathrm{Pa\,yr^{-1}}$), and most of the model trends (23 models) are positive. However, only two models have trends that are statistically different to the observed at the 90% level. The trend for Tahiti ($3.6\,\mathrm{Pa\,yr^{-1}}$) is statistically significant at the 96% level. Most of the model trends are positive at Tahiti, and the MMM trend is less than one-third ($1\,\mathrm{Pa\,yr^{-1}}$) of the observed. This indicates that there is a significant external forcing component to the trend at Tahiti, and that the global warming signal has emerged in this 33-yr period. This is consistent with a previous study that showed that the MSLP increases more at Tahiti than at Darwin (Power and Kociuba 2011a). We also find the Tahiti minus Darwin trend is also positive ($4.9\,\mathrm{Pa\,yr^{-1}}$); however, this signal is dominated by the warming trend at Tahiti.

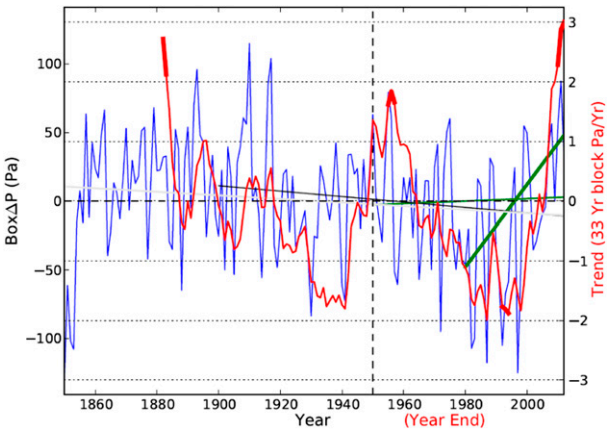

FIG. 3. Time series of BoxΔP for HadSLP2. Various trend lines are added to the time series (blue); 1850–2012 (gray), 1900–2012 (black), 1951–2012 (thin green), and 1980–2012 (thick green). Any trend calculated before 1951 and ending in 2012 is negative (WC weaken), and positive (WC strengthen) for calculations after 1951 and ending in 2012. A 33-yr running block trend is shown for each year end (i.e., last year in 33-yr block) in red, indicating the largest trend occurred over 1980–2012 and is significant at the 95% level. The thick red portions of the curves indicates the significance of the trend above the 90% level.

These results show that the model trends are consistent with the observed trends, and seem to capture the global warming trend at Tahiti over the 1980–2012 period.

b. All 33-yr trends

1) Observations

The magnitude of the 1980–2012 observed trend in BoxΔP is compared to other 33 periods to determine how unusual this event is. All possible 33-yr periods are displayed in Fig. 3 as a red curve. The 1980–2012 trend is the largest of all possible 33-yr trends, and has the highest significance level of 95%. Significance at the 90% level is indicated as a thicker red curve on top of the running trend (thin red) curve. Only two other periods post-1900 had trend significance at the 90% level, 1991 (i.e., 1959–91) and 1955 (i.e., 1923–55), and both of these trends had magnitudes less than $2\,\mathrm{Pa\,yr^{-1}}$.

2) Historical models

One possible explanation for why the climate models cannot reproduce the observed trend over the last 33 years is that short-term trends are very sensitive to El Niño and La Niña events. In the early part of 1980s in the observations, there were a number of successive El Niño events (1982/83 and 1986/87). Near the end of the 33-yr period there was a strong La Niña (2010–12) and a

TABLE 2. BoxΔP trends in preindustrial models exceeding observed trends (1980–2012). Trends were calculated using all possible blocks of 33 yr for 32 preindustrial climate models. Only the models presented here had 33-yr trends of absolute value greater than the 1980–2012 observed trend (3.04 Pa yr^{-1}), where the number of positive and negative trends that exceed the observed are displayed for each model. The lag-1 autocorrelation (ac), standard deviation (σ), and standard deviation of the 13-yr running average [σ(13 yra)] are also displayed. The same analysis is performed on 35 historical runs. Only two models exceeded the magnitude of the observed trend, and two were within 97% of the magnitude of the observed trend. The latter two trends are included as italic text in parentheses. The same statistics for the observations are also presented.

Models	No. of trends >3	No. of trends <−3	Length (yr)	Lag-1 ac	σ(Pa)	σ(13 yra) (Pa)
Preindustrial (32 models tested)						
ACCESS1–3 (r1i1p1)	1	0	500	0.124	67.815	14.306
CESM1-BGC (r1i1p1)	0	1	500	0.120	89.324	13.364
CESM1-WACCM (r1i1p1)	1	1	200	0.060	109.881	16.551
GFDL-ESM2M (r1i1p1)	1	4	500	0.221	117.025	17.704
MIROC5 (r1i1p1)	1	2	670	0.300	82.370	15.205
MPI-ESM-LR (r1i1p1)	4	1	1000	0.282	67.749	14.676
MPI-ESM-MR (r1i1p1)	1	5	1000	0.280	68.475	16.084
MPI-ESM-P (r1i1p1)	2	1	1156	0.204	66.925	15.560
All models used	11	15	17 357			
Historical (1900–2012) (35 models tested)						
CanESM2	*0*	*(−2.97) 1*	*113*	*0.09*	*82.5*	*14.56*
CMCC-CESM	0	1	113	−0.001	116.70	17.83
GFDL-ESM2M (r1i1p1)	*(2.98) 1*	*0*	*113*	*0.120*	*145.16*	*19.46*
MIROC5 (r1i1p1)	2	2	113	0.325	106.42	17.55
Observations HadSLP2	1	0	113	0.321	46.4	15.314

weak–moderate La Niña (2008/09). These events have a large impact on the strength of the trend. The phasing of ENSO and IPO events in the models will, in general, be unrelated to the phasing of these events in the real world. It is therefore useful to determine the largest 33-yr trend magnitude anywhere in a time series to see if trends rivalling the observed 1980–2012 trend are captured at any other time during the historical period.

The largest magnitude (positive and negative) 33-yr trends from anywhere within 1900–2012 are presented in Fig. 2 as a purple bar left of the main (brown) bar. In this time period there are 81 possible 33-yr trends. There only two models that have a 33-yr trend that is larger in magnitude than observed, and two models are within 97% of the magnitude of the observed trend. These results are shown in Table 2 (note that expansions of model names in Table 2 are provide in appendix F).

Only two out of 35 historical models can simulate the magnitude of the 1980–2012 trend in BoxΔP even if 33-yr trends are searched everywhere in the historical period. Overall, the magnitude of the 1980–2012 observed trend is exceeded only once every 400 years on average, in the historical models.

3) PREINDUSTRIAL MODEL SIMULATIONS

We now examine the effect of internal variability only on 33-yr trends, and determine how frequently these trends have magnitudes that exceed the magnitude of the observed 1980–2012 trend in BoxΔP.

We use 32 preindustrial run models in which trends are entirely driven by internal natural variability. These runs ranged in length from 200 to 1156 years (Table 2), with an average length of approximately 500 yr for each model. Trends are calculated for every possible 33-yr window which can overlap. We find that only eight models out of 32 have a trend with a magnitude exceeding that of the observed. We define an event as a trend calculated from a particular 33-yr window that has a magnitude greater than the 1980–2012 observed trend. Only one model (GFDL-ESM2M) exhibited multiple events. Overall there are at most 11 events with positive trends and 15 events with negative trends. So conservatively there are at most 26 cases found in 17 357 yr of data tested; or 1.5 events per 1000 yr.

These results show that the models under preindustrial conditions only rarely reproduce trends that have magnitudes matching or exceeding the magnitude of the observed trend over 1980–2012.

6. Factors influencing internally driven trends

There are a variety of factors that can potentially influence the magnitude of internally driven variability and trends. A range of metrics can be used to assess various properties of ENSO variability (Bellenger et al. 2014; Stoner et al. 2009; Newman et al. 2003; Smith and Sardeshmukh 2000; Trenberth and Hoar 1996, 1997). Here we consider the level of interannual variability and

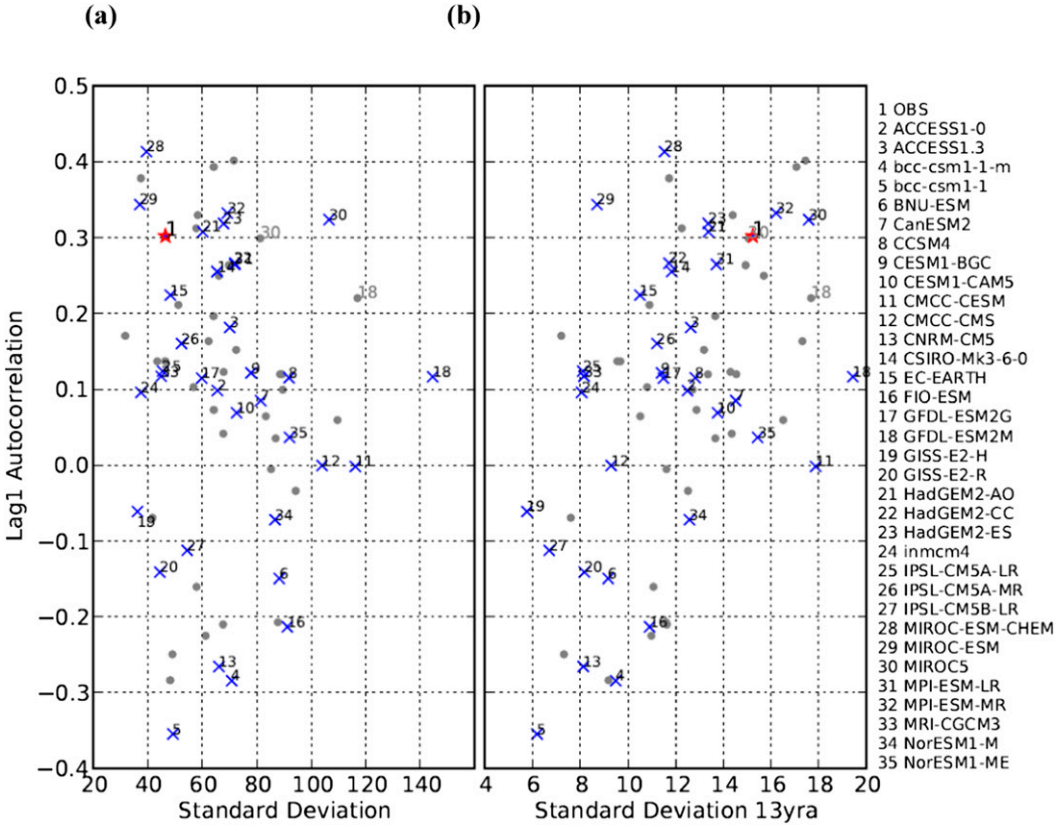

FIG. 4. Standard deviation (Pa), and lag-1-yr autocorrelation [$a(1)$] for observations and models. The standard deviation (σ) and $a(1)$, is calculated using the SLP time series (1900–2012) of the observations (red star labeled 1) and models (blue crosses labeled: 2–21). A 13-yr running average is applied before calculating the standard deviation [σ(13 yra)] in (b). The same quantities are calculated in the preindustrial runs and are shown as gray markers. Only two of these models are labeled as gray numbers. The number of model years available in each run is approximately 500.

the degree of persistence that exists in the interannual variability. The larger the interannual variability and the larger the persistence, then the larger the trends can be. Here we will examine the level of interannual variability. We will focus here on the standard deviation (σ) of both interannual and decadal variability [σ(13 yra), based on 13-yr running averages], the ratio σ/σ(13 yra), and the persistence $a(1)$. We will compare the value of the model statistics with observational estimates to see how realistically the models simulate them. We will see that the models lack persistence and this restricts their ability to simulate large trends.

a. Historical simulations and observations

Here we examine the ability of models to simulate the standard deviation, and lag-1-yr autocorrelation coefficient [$a(1)$] from 34 models (blue crosses) to the observations (red star labeled 1) for BoxΔP, in Fig. 4a.

It is clear that many of the models have a higher variance than the observed. Only five models have a lower variability. The σ_{obs} is 1 standard deviation of the multimodel mean of σ. Most of the models do not have enough persistence [where only six models have an $a(1)$ exceeding the observational value]. Two models have a larger persistence than the observations, but their σ is less than 80% of the observed. MIROC5 variability has the same persistence as the observations but approximately double the observed standard deviation σ_{obs}. The trend of this model was similar to the observed, but opposite in sign. GFDL-ESM2M variability has approximately triple σ_{obs}, but approximately half the persistence of the observations. This combination for GFDL contributes to the large confidence interval of the associated trend for this model in Fig. 2, although the trend itself turns out to be small compared to the observations. If all possible 33-yr trends in BoxΔP are calculated in the GFDL time series, then the maximum magnitude of the trend in this model is only 97% of the observed 1980–2012 trend, as discussed in section 5b(2). Therefore the standard deviation σ is not the dominant factor, since the GFDL model could only get within 97% of the observed trend, despite having σ approximately triple σ_{obs}.

The analysis in the previous paragraph focuses on the properties of interannual variability. The magnitude of the variability on a longer time scale is particularly relevant to the statistical significance of 33-yr trends and the likelihood of obtaining a large trend from internal variability alone. We explore this by first applying a 13-yr running average [13 yra, based on the approach taken by Power et al. (1999) and Folland et al. (2002)] to the data before calculating the standard deviation [σ(13 yra)]. Only four models exceed the observed σ_{obs}(13 yra), which is 1.1 standard deviation from the multimodel mean of σ(13 yra). This is in contrast to a higher variance found at low frequency examined in the CMIP3 models by Stoner et al. (2009). Only two models (MIROC5 and CESM1-CAM5) have trends (1980–2012) comparable in magnitude to the observations. In these models a(1) is approximately equal to the observed, and σ(13 yra) is slightly larger than observed. There are two models (CMCC-CESM and MIROC5) that have a positive 33-yr trend exceeding the observations somewhere in the 1900–2012 time series (see the purple bars on the corresponding models in Fig. 2). These models all have σ and σ(13 yra) larger than the observed, and they have the ability to exceed the observed 1980–2012 trend once every 76 yr on average.

We also calculated the standard deviation for running averages of different lengths as a check of robustness of the results. The standard deviation of an 8-yr running average was examined and found there are approximately an equal number of models that are higher and lower than the observations. However when examining the standard deviation for a 10-, 13-, 17-, and 33-yr running average, we found that there were at most three models exceeding the corresponding observed standard deviation in each case. Since the 10-, 13-, 17-, and 33-yr running averages all exhibit a lower standard deviation than observed, we choose the 13 yra as the metric for low frequency variability, and to be consistent with Power et al. (1999) and Folland et al. (2002).

These results show that only when both σ and σ(13 yra) were comparable in size to the observed low-frequency variability could the observed trend be exceeded by the models. In those particular models the observed trend could be exceeded once in every 76 yr on average.

b. *Preindustrial climate simulations*

We now examine the influence of internal variability and persistence on the magnitude of the trends of the preindustrial climate models, using the above analysis.

The standard deviation and a(1) in the preindustrial runs are also presented in Fig. 4. Most of the preindustrial runs (30 out of 34 models) have a higher σ than observed (46.4 Pa), where this observed value is 1.1 standard deviation from the multimodel mean of σ. There are five models with a higher a(1) than observed (0.32) as shown as gray dots in Fig. 4a. More broadly, most of the models have a lower σ(13 yra) than the observed (15.3 Pa), shown as gray dots in Fig. 4b. The observed σ_{obs}(13 yra) is 0.9 standard deviation from the multimodel mean of σ(13 yra). The GFDL model is an exception as it has a very large σ (approximately triple the observed value), with a σ(13 yra) 15% larger than observed, and its a(1) is approximately 70% of the observed value. This particular model has the ability to exceed the observed 1980–2012 trend once every 125 yr on average.

These results show that there are few models which have values of σ, a(1), and σ(13 yra) that are approximately equal to the corresponding observed values. But even in these models the recent observed trend over the period 1980–2012 is still rare in their preindustrial models (on average a 1 in 212-yr event).

c. *Ratio of variances*

The results from the previous two sections show that the low-frequency variability is lower than the observations both in the historical and preindustrial runs, and the interannual variability is higher than observed. This becomes evident if the ratio σ/σ(13 yra) is plotted (not shown): all historical models have a ratio greater than the observed. We then calculated this ratio for the best-fit first autoregressive [AR(1)] process for the observations and all of the models. The fitting method is described in appendix B. The ratio σ/σ(13 yra) is approximately the same for the observations and the associated AR(1) processes. However, the models under historical forcing have a significantly higher σ/σ(13 yra) than the associated AR(1) processes and the observations. One explanation is that the vast majority of the models (31 out of 34) in both the historical and preindustrial runs, have a negative lag-2 autocorrelation (MMM is -0.3) that is larger in magnitude than the observed value (-0.09). The autocorrelations in the models are not generally significant for lags greater than two. Hence if we assume that the models behave as an AR(2) process, it can be shown that the decadal variability decreases as the lag-2 autocorrelation becomes more negative (see appendix E). We also calculated the frequency spectra for all the models and the observed data (not shown). The averaged model spectra was compared to the observed, and was found to be generally smaller than the observed for periods greater than approximately 10 yr. The averaged model spectrum was generally larger than the observed at periods less than 5 yr, and particularly at periods 2 and 4. The fact that decadal variability in the models tends to be lower than the observed might suggest to some readers that the primary driver of decadal variability, stochastic forcing (Liu 2012), may be too low. This is not the case, however,

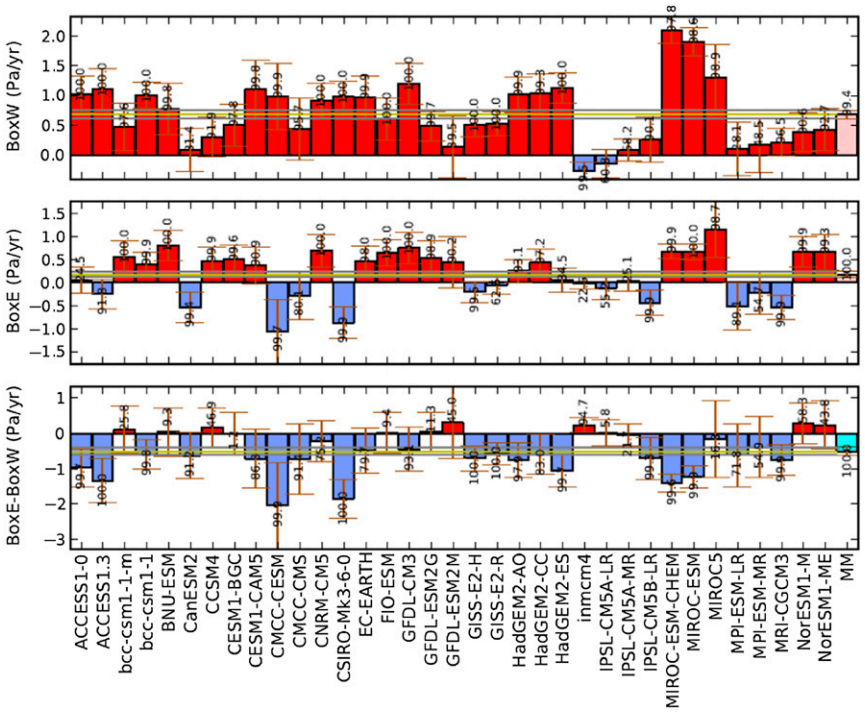

FIG. 5. Projected model trends (Pa yr^{-1}) for the twenty-first century 2013–99 for RCP8.5. This figure shows that 25 out of 35 models have a negative trend in BoxΔP, where 12 of these are significant at the 99% level. Only 1 model has a significant positive trend at the 95% level. The yellow line corresponds to the multimodel mean (MM in the figure), the gray lines represent the 95% confidence interval for MM.

since the interannual variability in the models tends to be larger than observed.

These results indicate that the model ENSOs are more oscillatory than the observed ENSO. Higher than observed oscillatory behavior will make it harder for models to simulate long-term trends even if the level of variability in interannual variability is realistic. In models with a high degree of oscillatory behavior they can have higher than observed levels of interannual variability and still find it difficult to simulate large trends.

7. Projected model trends for RCP8.5 (2013–99) compared to CMIP3

In sections 4, 5, and 6 we examined the ability of models to simulate past trend. In this section we look forward and examine trends over the period 2013–2100. We compare the results with those obtained previously by Power and Kociuba (2011b), who examined twenty-first-century trends in CMIP3 model.

The CMIP5 results are summarized in Fig. 5. Out of 35 models, 33 have positive trends in BoxW, and 25 of these have trends that are statistically significant at the 95% level. Only one model has a statistically significant negative trend (95% level). This is the same model that has a significant positive trend in BoxΔP. There is much less consensus on the sign of the trend for BoxE, as there are 22 positive and 13 negative trends. However, there are 16 models with positive trend significant at the 97% level and six models with negative trend significant at the 97% level. This results in a small (0.18 Pa yr^{-1}) but highly statistically significant MMM trend in BoxE at the 99.999% level. Twenty-five out of 35 models have a negative trend in BoxΔP (-0.50 Pa yr^{-1}), which corresponds to a statistical significance at the 99.7% level, assuming model independence (Power and Kociuba 2011b). Using the P98 method, the significance of the MMM is at the 99.999% level. Only one model has a positive trend that is statistically significant at the 95% level. Power and Kociuba (2011b) previously analyzed trends in the SRES A2 scenario and also found a strong consensus on the sign of the trend for BoxΔP, with 13 out of 17 models (2002–98) having a negative trend (-0.343 Pa yr^{-1}) for SRES A2. The trend is significant at the 98% level. Power and Kociuba (2011b) also found for SRES A1B that 14 out of 21 models have a negative trend (-0.122 Pa yr^{-1}) that is significant at the 95% level. The weakening of the WC in all scenarios tested here is due to a larger sinking motion of air in BoxW compared to BoxE.

These results show that despite an inability of the CMIP5 models to capture the recent strengthening of the WC, the future projections in CMIP5 models project a weakening WC, consistent with projections in CMIP3 models (Power and Kociuba 2011b).

8. Summary and discussion

a. The past

We investigated the strength of the Walker circulation (WC) in the observations and in CMIP5 climate models over the two time periods 1900–2012 and 1980–2012.

While there is a degree of agreement between models and observations in BoxΔP trends over the full historical period, there is no consensus on the sign of the trend over the 33-yr period 1980–2012. The observed trend over this 33-yr period is statistically significant at the 95% level, whereas none of the models have trends that are significant at or above the 95% level.

Our analysis indicates that the observed trend in BoxΔP over the period 1980–2012 is inconsistent with the modeled trends over the same period. The largest inconsistency in trends over the period 1980–2012 arises in the west Pacific because the models tend to exhibit an increase in MSLP in BoxW, whereas the observations show a decline over the same period (significant at the 94% level).

Four possible reasons can give rise to this situation: 1) the observed internal variability is rare (i.e., extremely large); 2) the models do not faithfully represent (i) the internal variability or (ii) the externally forced response; 3) the forcing applied to the models is deficient; or 4) the observations are in error. We will now discuss each of these candidates in turn.

The possibility that the observed trend over 1980–2012 is dominated by internal variability was examined using preindustrial runs of the climate models. It was found that 33-yr trends arising from internal variability with magnitudes greater than or equal to the magnitude of the observed 1980–2012 trend were very rare events. Such events were estimated to be 1.5 per 1000-yr events. If we only consider models that have a similar level of decadal variability to the observed, the trend becomes a 1-in-212-yr event.

We also performed Monte Carlo analysis using an AR(1) model (see appendix C), with parameters based on the BoxΔP observations, and we found that the observed trend 1980–2012 occurred in only 1.3% of all surrogate data. So if the inconsistency between observations and the model is due to internally generated variability in the observations, assuming the models faithfully represent internal variability, it would have to be very large internal variability. We know that the interdecadal Pacific oscillation switched from a positive phase to a cool phase (England et al. 2014; Flato et al. 2014), which would have strengthened the WC. However, we do not know if this change is sufficiently strong to cause the observed 1980–2012 trend. We hope to address this issue in a future study.

To examine possibility 2(i), we examined decadal variability in the models and compared it with the observations. We found that the level of decadal variability in the models is too weak, despite the fact that the interannual variability tends to be too high. The reason for this apparent inconsistency is that the models tend to be too oscillatory, and this makes it hard for the models to generate decadal anomalies.

Possibility 2(ii) has been examined previously by DiNezio et al. (2013). They concluded that the balance between the forced response to aerosol forcing and greenhouse gases may not be correct in the models as a larger aerosol component is known to generate a strengthening of the WC. It is possible, therefore, that the models either overestimate a weakening in the WC due to greenhouse gases or underestimate a strengthening due to sulfate aerosols. Furthermore, addressing possibility 3, the forcing itself might be in error (e.g., Schmidt et al. 2014; Flato et al. 2014; Kirtman et al. 2014).

To assess the possibility that there might be an error in the gridded MSLP that accentuated the observed trend over the period 1980–2012, we examined alternative proxies for the WC based on Tahiti and Darwin MSLP station data. While inconsistencies between observations and models were reduced, the results were mixed and so the extent to which observational error is influencing results obtained is unclear. It is interesting to note that England et al. (2014) concluded that models were not able to capture an observed strengthening of tropical winds in recent times. This strongly suggests that observational error (possibility 4) is not a major factor in leading to the inconsistency in the BoxΔP results we identify.

In summary, observed internal variability in the Pacific and deficiencies in the simulation of Pacific decadal variability, appear to be the main reasons for the apparent inconsistency between model and observed trends over the period 1980–2012. Other possible factors include errors in the response to, or in the representation of, external forcing. Further research is needed to provide a more accurate assessment of the relative importance of possibilities 1–4 in causing the model-to-observed inconsistency.

b. The future

We also examined the twenty-first century scenarios RCP2.6, RCP4.5, RCP6.0, and RCP8.5 and found there was a strong consensus among the models that

MSLP and BoxW increase, but the sign of change for BoxE is ambiguous, and
there is a robust reduction in BoxΔP, consistent with the weakening of the WC.

These findings are consistent with our previous results (Power and Kociuba 2011b) using CMIP3 data for scenarios SRES A1B and A2.

Given the inconsistencies discussed in section 8a, and that problems relating to possibilities 2(ii) and 3 cannot be ruled out, confidence in the twenty-first century projections of the Walker circulation using the same models is reduced.

Acknowledgments. This project was supported by the Australian Climate Change Science Program (ACCSP). We thank the anonymous reviewers for their constructive and helpful reviews.

APPENDIX A

The P98 Method

In this appendix, we define the t value taking persistence into account, derived by Power et al. (1998). The expression for the t value is

$$t = \frac{r}{f}\sqrt{\frac{(N-2)}{1-(r/f)^2}},$$

that has a t distribution, where N is the number of years, r is the temporal correlation between two time series $x(t)$ and $y(t)$, and the serial correlation is embedded in the quantity f:

$$f = \sqrt{\frac{(1+\gamma_x\gamma_y)}{(1-\gamma_x\gamma_y)}}, \quad \text{where}$$

$$\gamma_i = \frac{\sigma_i^2(1)}{\sigma_i^2(0)}, \quad \text{and } i \text{ is } x \text{ or } y,$$

where $\sigma_i(\tau)$ is the standard deviation at lag τ, which provides a measure of persistence:

$$\sigma_i(\tau) = \frac{1}{N-1}\sum_{i=1}^{N}[x_i(t)-\overline{x})][x_i(t-\tau)-\overline{x}].$$

Here $x(t) = t$, and $y(t)$ is a Box SLP metric as described in section 2.

The confidence interval, and significance level was based on using the t value here, and this is defined as the P98 method.

Other methods for estimating the confidence interval are described in appendix D; however, we use the P98 method throughout this paper as it is a stricter test for significance.

APPENDIX B

Standard Deviation of an AR(1) Process

This is the method used to calculate the standard deviation (annual and 13 yra) of an AR(1) process.

First a Monte Carlo analysis is performed using an AR(1) model:

$$X_t = \rho x_{t-1} + \varepsilon_t.$$

Here ε is a white noise process with zero mean, standard deviation = σ, and $\rho = a(1)$, the autocorrelation coefficient 1-yr lag. We choose parameters ρ and σ based on the lag-1 autocorrelation and standard deviation of a time series being tested. We generated 2000 surrogate time series, each 113 yr long (which corresponds to the period 1900–2012). The standard deviation $\sigma(1)$ and $\sigma(13\,\text{yra})$ is calculated for each of the surrogates to generate a probability distribution. The most likely standard deviation $\sigma(1)$ and $\sigma(13\,\text{yra})$ is taken to be the median of the associated distribution.

APPENDIX C

Trend Significance Using an AR(1) Process

The method of calculating the trend significance is as follows.

We performed Monte Carlo analysis using an AR(1) model,

$$X_t = \rho x_{t-1} + \varepsilon_t.$$

Here ε is a white noise process with zero mean, standard deviation = σ, and $\rho = a(1)$, the autocorrelation coefficient 1-yr lag. We choose parameters $\rho = 0.321$ and $\sigma = 46.4$ based on HadSLP2 BoxΔP 1900–2012. We generated 100 000 surrogate time series, each 33 yr long (which corresponds to the period 1980–2012). A trend was calculated for each surrogate. Only 1.3% of the simulated time series had magnitudes greater than the magnitude of the observed trend ($3.0\,\text{Pa yr}^{-1}$), indicating that the observed trend is statistically significant at the 98.7% level.

APPENDIX D

Confidence Interval Estimation

The confidence intervals were recalculated using various methods to check for any variation. We find that if an AR(1) process is assumed for each of the models and observations, then the statistical significance of the trend is very similar to the values previously calculated. We also fitted an AR(2) process to each model and the observations. This increases the significance of the trend in all cases, although none is statistically significant above the 86% level. Similar results were found fitting an AR(3) process to each model and the observations. Finally, we determine the significance level empirically from the time series. This was achieved by counting how often the 1980–2012 trend appeared anywhere in the time series within the period 1900–2012, which we define here as a simple bootstrap method (SBS). This gave a higher estimate of statistical significance in most models, but only one model is significant at the 92% level (HadGEM2-ES with trend $-1.7\,\mathrm{Pa\,yr^{-1}}$). This model has a trend that is significant at the 56% level when using the P98 method.

APPENDIX E

Spectral Ratio of Two AR(2) Processes

In this appendix, we show that for a given lag-1 autocorrelation of an AR(2) process, the decadal variability decreases as the lag-2 autocorrelation value becomes more negative.

An AR(2) process with zero mean is defined as

$$Y_i(t,\rho_1,\rho_2) = C_1 Y(t-1) + C_2 Y(t-2) + \sigma_N \zeta_i(t),$$

where $\zeta_i(t)$ is a Gaussian white noise process with zero mean and unit standard deviation. The coefficients C_1 and C_2 can be calculated from the lag-1 and lag-2 autocorrelations ρ_1 and ρ_2 (Wilks (1995),

$$C_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \quad \text{and}$$

$$C_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2}.$$

The theoretical spectral density function is (Wilks 1995)

$$S(f,\rho_1,\rho_2) = \frac{4\sigma_N/n}{1+C_1^2+C_2^2-2C_1(1-C_2)\cos(2\pi f)-2C_2\cos(4\pi f)}.$$

The frequency f is positive and less than 0.5 and n is the number of points. If (ρ_1, ρ_2) is set to the observed values (0.3, −0.1) we find that that the ratio $S(f, 0.3, -0.1-\beta)/S(f, 0.3, -0.1)$ for any positive value β is <1 for periods greater than approximately 9 yr, and tends to form a peak >1 for periods shorter than approximately 9 yr. This peak steepens as β increases.

This results shows that for an AR(2) process with a fixed lag-1 autocorrelation, the decadal variability decreases as the lag-2 autocorrelation value becomes more negative.

APPENDIX F

Expansions of Model Names in Table 2

ACCESS	Australian Community Climate and Earth-System Simulator
CanESM2	Second Generation Canadian Earth System Model
CESM1-BGC	Community Earth System Model, version 1, with Biogeochemistry
CESM1-CAM5	CESM1 with Community Atmosphere Model, version 5
CESM1-WACCM	CESM1 with Whole Atmosphere Community Climate Model
CMCC-CESM	Centro Euro-Mediterraneo per i Cambiamenti Climatici Carbon Cycle Earth System Model
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory Earth System Model with Modular Ocean Model 4 (MOM4) component
MIROC5	Model for Interdisciplinary Research on Climate, version 5
MPI-ESM-LR	Max Planck Institute Earth System Model, low resolution
MPI-ESM-MR	Max Planck Institute Earth System Model, medium resolution
MPI-ESM-P	Max Planck Institute Earth System Model, paleo

REFERENCES

Allan, R. J., and T. J. Ansell, 2006: A new globally complete monthly historical mean sea level pressure data set (HadSLP2): 1850–2004. *J. Climate,* **19,** 5816–5842, doi:10.1175/JCLI3937.1.

Ashok, K., and T. Yamagata, 2009: The El Niño with a difference. *Nat. Climate Change,* **461,** 481–484, doi:10.1038/461481a.

Bellenger, H., E. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard, 2014: ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dyn.,* **42,** 1999–2018, doi:10.1007/s00382-013-1783-z.

Bunge, L., and A. J. Clarke, 2009: A verified estimation of the El Niño index Niño-3.4 since 1877. *J. Climate,* **22,** 3979–3992, doi:10.1175/2009JCLI2724.1.

Collins, M., and Coauthors, 2010: The impact of global warming on the tropical Pacific and El Niño. *Nat. Geosci.,* **3,** 391–397, doi:10.1038/NGEO868.

DiNezio, P. N., G. A. Vecchi, and A. C. Clement, 2013: Detectability of changes in the Walker circulation in response to global warming. *J. Climate,* **26,** 4038–4048, doi:10.1175/JCLI-D-12-00531.1.

England, M., and Coauthors, 2014: Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Climate Change,* **4,** 222–227, doi:10.1038/nclimate2106

Flato, G. J., and Coauthors, 2014: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis,* T. F. Stocker et al., Eds., Cambridge University Press, 741–866.

Folland, C. K., J. A. Renwick, M. J. Salinger, and A. B. Mullan, 2002: Relative influences on the interdecadal Pacific oscillation and ENSO on the South Pacific convergence zone. *Geophys. Res. Lett.,* **29,** 1643, doi:10.1029/2001GL014201.

Gill, A. E., 1982: *Atmosphere–Ocean Dynamics.* Academic Press, 662 pp.

Ham, Y.-G., and J.-S. Kug, 2012: How well do current climate models simulate two types of El Niño? *Climate Dyn.,* **39,** 383–398, doi:10.1007/s00382-011-1157-3.

Held, I. M., and B. J. Soden, 2006: Robust response of the hydrological cycle to global warming. *J. Climate,* **19,** 5686–5699, doi:10.1175/JCLI3990.1.

Kirtman, B., and Coauthors, 2014: Near-term climate change: Projections and predictability. *Climate Change 2013: The Physical Science Basis,* T. F. Stocker et al., Eds., Cambridge University Press, 953–1028.

Knutson, T. K., and S. Manabe, 1995: Time-mean response over the tropical Pacific to increased CO_2 in a coupled ocean–atmosphere model. *J. Climate,* **8,** 2181–2199, doi:10.1175/1520-0442(1995)008<2181:TMROTT>2.0.CO;2.

L'Heureux, M. L., S. Lee, and B. Lyon, 2013: Recent multidecadal strengthening of the Walker circulation across the topical Pacific. *Nat. Climate Change,* **3,** 571–576, doi:10.1038/nclimate1840.

Liu, Z., 2012: Dynamics of interdecadal climate variability: A historical perspective. *J. Climate,* **25,** 1963–1995, doi:10.1175/2011JCLI3980.1.

Luo, J., W. Sasakia, and Y. Masumoto, 2012: Indian Ocean warming modulates Pacific climate change. *Proc. Natl. Acad. Sci. USA,* **109,** 18 701–18 706, doi:10.1073/pnas.1210239109.

Meehl, G. A., and J. M. Arblaster, 2012: Relating the strength of the tropospheric biennial oscillation (TBO) to the phase of the interdecadal Pacific oscillation (IPO). *Geophys. Res. Lett.,* **39,** L20716, doi:10.1029/2012GL053386.

——, and Coauthors, 2007: Global climate projections. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 747– 845.

——, A. Hu, J. M. Arblaster, J. Fasullo, and K. E. Trenberth, 2013: Externally forced and internally generated decadal climate variability associated with the interdecadal Pacific oscillation. *J. Climate,* **26,** 7298–7310, doi:10.1175/JCLI-D-12-00548.1.

Meng, Q., M. Latif, W. Park, N. S. Keenlyside, V. A. Semenov, and T. Martin, 2012: Twentieth century Walker circulation change: Data analysis and model experiments. *Climate Dyn.,* **38,** 1757–1773, doi:10.1007/s00382-011-1047-8.

Merrifield, M. A., 2011: A shift in western tropical Pacific sea level trends during the 1990s. *J. Climate,* **24,** 4126–4138, doi:10.1175/2011JCLI3932.1.

Newman, M. N., and Coauthors, 2003: ENSO-forced variability of the Pacific decadal oscillation. *J. Climate,* **16,** 3853–3857, doi:10.1175/1520-0442(2003)016<3853:EVOTPD>2.0.CO;2.

Nicholls, N., 2008: Recent trends in the seasonal and temporal behaviour of the El Niño–Southern Oscillation. *Geophys. Res. Lett.,* **35,** L19703, doi:10.1029/2008GL034499.

Power, S. B., and I. N. Smith, 2007: Weakening of the WC and apparent dominance of El Niño both reach record levels, but has ENSO really changed? *Geophys. Res. Lett.,* **34,** L18702, doi:10.1029/2007GL030854.

——, and G. Kociuba, 2011a: The impact of global warming on the Southern Oscillation Index. *Climate Dyn.,* **37,** 1745–1754, doi:10.1007/s00382-010-0951-7.

——, and ——, 2011b: What caused the observed twentieth-century weakening of the Walker circulation? *J. Climate,* **24,** 6501–6514, doi:10.1175/2011JCLI4101.1.

——, F. Tseitkin, S. Torok, B. Lavery, and B. McAvaney, 1998: Australian temperature, Australian rainfall, and the Southern Oscillation, 1910–1996: Coherent variability and recent changes. *Aust. Meteor. Mag.,* **47,** 85–101.

——, T. Casey, C. Folland, A. Colman, and V. Mehta, 1999: Inter-decadal modulation of the impact of ENSO on Australia. *Climate Dyn.,* **15,** 319–324, doi:10.1007/s003820050284.

——, F. Delage, C. Chung, G. Kociuba, and K. Keay, 2013: Robust twenty-first-century projections of El Niño and related precipitation variability. *Nature,* **502,** 541–545, doi:10.1038/nature12580.

Sandeep, S., F. Stordal, P. D. Sardeshmukh, and G. P. Compo, 2014: Pacific Walker circulation variability in coupled and uncoupled climate models. *Climate Dyn.,* **43,** 103–117, doi:10.1007/s00382-014-2135-3.

Schmidt, G. A., and Coauthors, 2014: Configuration and assessment of the GISS modelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst,* **6,** 141–184, doi:10.1002/2013MS000265.

Smith, C. A., and P. D. Sardeshmukh, 2000: The effect of ENSO on the intraseasonal variance of surface temperatures in winter. *Int. J. Climatol.,* **20,** 1543–1557, doi:10.1002/1097-0088(20001115)20:13<1543::AID-JOC579>3.0.CO;2-A.

Soden, B. J., D. L. Jackson, V. Ramaswamy, M. D. Schwarzkopf, and X. Huang, 2005: The radiative signature of upper tropospheric moistening. *Science,* **310,** 841–844, doi:10.1126/science.1115602.

Sohn, B. J., and S.-C. Park, 2010: Strengthened tropical circulations in past three decades inferred from water vapour transport. *J. Geophys. Res.,* **115,** D15112, doi:10.1029/2009JD013713.

——, S.-W. Yeh, J. Schmetz, and H.-J. Song, 2013: Observational evidences of Walker circulation change over the last 30 years contrasting with GCM results. *Climate Dyn.,* **40,** 1721–1732, doi:10.1007/s00382-012-1484-z.

Solomon, A., and M. Newman, 2012: Reconciling disparate twentieth-century Indo-Pacific ocean temperature trends in the instrumental record. *Nat. Climate Change,* **2,** 691–699, doi:10.1038/nclimate1591.

Stephens, G. L., and T. D. Ellis, 2008: Controls of global-mean precipitation increase in global warming GCM experiments. *J. Climate,* **21,** 6141–6155, doi:10.1175/2008JCLI2144.1.

Stoner, A. M. K., K. Hayhoe, and D. J. Wuebbles, 2009: Assessing general circulation model simulations of atmospheric teleconnection patterns. *J. Climate,* **22,** 4348–4372, doi:10.1175/2009JCLI2577.1.

Tanaka, H. L., N. Ishizaki, and A. Kitoh, 2004: Trend and interannual variability of Walker, monsoon and Hadley circulations defined by velocity potential in the upper troposphere. *Tellus,* **56A,** 250–269, doi:10.1111/j.1600-0870.2004.00049.x.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.,* **93,** 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tokinaga, H., S.-P. Xie, C. Deser, Y. Kosaka, and Y. M. Okumura, 2012a: Slowdown of the Walker circulation driven by tropical Indo-Pacific warming. *Nature,* **491,** 439–444, doi:10.1038/nature11576.

——, ——, A. Timmermann, S. McGregor, T. Ogata, H. Kubota, and Y. M. Okumura, 2012b: Regional patterns of tropical Indo-Pacific climate change: Evidence of the Walker circulation weakening. *J. Climate,* **25,** 1689–1710, doi:10.1175/JCLI-D-11-00263.1.

Trenberth, K. E., and T. J. Hoar, 1996: The 1990–95 El Niño–Southern Oscillation event: Longest on record. *Geophys. Res. Lett.,* **23,** 57–60, doi:10.1029/95GL03602.

——, and ——, 1997: El Niño and climate change. *Geophys. Res. Lett.,* **24,** 3057–3060, doi:10.1029/97GL03092.

Vecchi, G. A., and B. J. Soden, 2007: Global warming and the weakening of the tropical circulation. *J. Climate,* **20,** 4316–4340, doi:10.1175/JCLI4258.1.

——, and A. T. Wittenberg, 2010: El Niño and our future climate: Where do we stand? *Wiley Interdiscip. Rev.: Climate Change,* **1,** 260–270, doi:10.1002/wcc.33.

——, B. J. Soden, A. T. Wittenberg, I. A. Held, A. Leetma, and M. J. Harrison, 2006: Weakening of the topical atmospheric circulation due to anthropogenic forcing. *Nature,* **441,** 73–76, doi:10.1038/nature04744.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.