

## Comparing Trends in the Southern Annular Mode and Surface Westerly Jet

NEIL C. SWART, JOHN C. FYFE, AND NATHAN GILLETT

*Canadian Centre for Climate Modelling and Analysis, Victoria, British Columbia, Canada*

GARETH J. MARSHALL

*British Antarctic Survey, Cambridge, United Kingdom*

(Manuscript received 6 May 2015, in final form 29 July 2015)

### ABSTRACT

This paper examines trends in the southern annular mode (SAM) and the strength, position, and width of the Southern Hemisphere surface westerly wind jet in observations, reanalyses, and models from phase 5 of the Coupled Model Intercomparison Project (CMIP5). First the period over 1951–2011 is considered, and it is shown that there are differences in the SAM and jet trends between the CMIP5 models, the Hadley Centre gridded SLP (HadSLP2r) dataset, and the Twentieth Century Reanalysis. The relationships between these trends demonstrate that the SAM index cannot be used to directly infer changes in any one kinematic property of the jet. The spatial structure of the observed trends in SLP and zonal winds is shown to be largest, but also most uncertain, in the southeastern Pacific. To constrain this uncertainty six reanalyses are included and compared with station-based observations of SLP. The CMIP5 mean SLP trends generally agree well with the direct observations, despite some climatological biases, while some reanalyses exhibit spuriously large SLP trends. Similarly, over the more reliable satellite era the spatial pattern of CMIP5 SLP trends is in excellent agreement with HadSLP2r, whereas several reanalyses are not. Then surface winds are compared with a satellite-based product, and it is shown that the CMIP5 mean trend is similar to observations in the core region of the westerlies, but that several reanalyses overestimate recent trends. The authors caution that studies examining the impact of wind changes on the Southern Ocean could be biased by these spuriously large trends in reanalysis products.

### 1. Introduction

The Southern Hemisphere (SH) westerlies are the strongest time-averaged surface winds on the planet, and they exert a pronounced influence on the global climate system. They do so in part by driving upwelling of deep waters in the Southern Ocean, and thereby the upper limb of the Atlantic meridional overturning circulation (AMOC) (Toggweiler and Samuels 1995; Marshall and Speer 2012). The AMOC, in turn, strongly modulates the oceanic uptake of heat and carbon (Kostov et al. 2014; Frölicher et al. 2015) and also

controls global primary production through regulation of the nutrient supply to the ocean thermocline (Sarmiento et al. 2004; Marinov et al. 2006). Variability and changes in the westerlies are thus of central interest when considering human-induced climate change (Toggweiler and Russell 2008).

The dominant mode of atmospheric variability in the SH is the southern annular mode (SAM). The SAM index has alternately been characterized as the leading empirical orthogonal function (EOF) of sea level pressure in the SH (Thompson and Wallace 2000) and as the sea level pressure difference between 40° and 65°S (Gong and Wang 1999). Observations have shown a trend toward the positive phase of the SAM since about 1970 (Thompson and Solomon 2002; Marshall 2003). Modeling studies have attributed this trend to human influence from a combination of increasing greenhouse gases and ozone depletion (Fyfe et al. 1999; Son et al. 2010; Gillett et al. 2013). The influence of ozone depletion has a strong seasonal signal, being largest during

---

 Denotes Open Access content.

---

*Corresponding author address:* Neil Swart, Canadian Centre for Climate Modelling and Analysis, Environment Canada, University of Victoria, P.O. Box 1700 STN CSC, Victoria BC V8W 2Y2, Canada.  
E-mail: neil.swart@ec.gc.ca

TABLE 1. List of reanalyses used in this study.

Name	Abbreviation	Reference	Data source
NCEP–NCAR reanalysis	R1	Kalnay et al. (1996)	<a href="http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html">http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html</a>
NCEP–DOE AMIP-II reanalysis	R2	Kanamitsu et al. (2002)	<a href="http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html">http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html</a>
Twentieth Century Reanalysis (version 2)	20CR	Compo et al. (2011)	<a href="http://portal.nersc.gov/project/20C_Reanalysis/">http://portal.nersc.gov/project/20C_Reanalysis/</a>
ECMWF interim reanalysis	ERA-Interim	Dee et al. (2011)	<a href="http://apps.ecmwf.int/datasets/data/interim-full-mode">http://apps.ecmwf.int/datasets/data/interim-full-mode</a>
NCEP Climate Forecast System Reanalysis	CFSR	Saha et al. (2010)	<a href="http://rda.ucar.edu/datasets/ds093.2/">http://rda.ucar.edu/datasets/ds093.2/</a>
NASA MERRA	MERRA	Rienecker et al. (2011)	<a href="http://disc.sci.gsfc.nasa.gov/daac-bin/DataHoldings.pl">http://disc.sci.gsfc.nasa.gov/daac-bin/DataHoldings.pl</a>

austral summer [December–February (DJF)], whereas the greenhouse gas (GHG) forcing operates consistently year round (Son et al. 2010; Thompson et al. 2011; Gillett et al. 2013). As a result, historical trends in the SAM are largest during austral summer, but small and statistically insignificant during the austral winter (Thompson et al. 2011).

These recent trends in the SAM have been associated with changes in the tropospheric circulation and climate (Thompson and Solomon 2002; Thompson et al. 2011). Month-to-month changes in the polarity of the SAM index are primarily associated with nearly symmetrical north–south vacillations of the surface westerly jet (herein referred to simply as the jet) (Hartmann and Lo 1998; Thompson and Wallace 2000). The positive phase of the SAM is associated with a poleward shifted jet, such that the westerlies are stronger over much of the Southern Ocean (with a center near 60°S) and weaker to the north (with a center near 40°S) (Thompson et al. 2011). However, oscillations in the SAM are also associated with changes in the width of the westerly jet and the strength of the jet at its peak (Monahan and Fyfe 2006). Indeed, the historical trend toward the positive phase of the SAM during the austral summer has been concurrent with both a poleward shift and a strengthening at the peak of the westerly jet (Swart and Fyfe 2012).

The climate models participating in phases 3 and 5 of the Coupled Model Intercomparison Project (CMIP3 and CMIP5, respectively) show systematic biases in their simulation of the SH westerly jet. On average the models simulate a climatological jet position that is 2°–3° of latitude equatorward of the observed position over the historical period (Swart and Fyfe 2012; Bracegirdle et al. 2013). Swart and Fyfe (2012) also showed that the simulated trends in jet strength over 1979–2010 were significantly smaller at the 5% level than the trends seen in the average of four reanalysis products (R1, R2, 20CR, and ERA-Interim; see Table 1 for expansions and additional information) in all seasons except June–August

(JJA). However, they also cautioned that this result was potentially unreliable, given that the reanalyses showed a large spread of trends and were poorly constrained in the Southern Hemisphere (Swart and Fyfe 2012).

More recently Gillett and Fyfe (2013) showed that over 1951–2011 the CMIP5 models simulate a SAM trend which is consistent with observationally based estimates, at least during DJF. Since trends in the strength of the westerly jet may be closely related to those in the SAM index (or sea level pressure gradient) through geostrophy, the findings of Swart and Fyfe (2012) and Gillett and Fyfe (2013) appear to be contradictory. However, given that the studies covered different time frames and used different metrics, there are many potential reasons for the apparent contradiction. In this paper we will compare changes in both the SAM and the westerly jet over a common period to resolve this discrepancy.

The aims of this study are to address two principal questions: 1) What is the relationship between trends in the SAM index and the kinematic properties of the westerly jet? 2) How do historical trends in the SAM and westerly jet compare between the best available direct observations, common reanalysis products, and the CMIP5 climate models? The second question is designed to quantify any systematic biases in the reanalyses or CMIP5 models. A major difficulty is that the direct observational estimates of sea level pressure and winds are not available with comprehensive coverage in both space and time. Here we attempt to make the closest possible comparison with the best available observations, which requires comparing trends in the SAM and winds over several different periods, and at specific geographic locations.

In the following section we describe the data and methods used in this study. Section 3 begins by considering changes in the SAM index and kinematic properties of the westerly jet focusing on the historical period since 1951. We start with a long historical record (i.e., presatellite era) because it facilitates the robust

detection of long-term trends and it also allows us to compare our results with those of [Gillett and Fyfe \(2013\)](#). [Section 4](#) uses a simple theoretical model to establish the expected relationship between SAM changes and jet properties and shows that this simple description largely explains the relationships seen in the full CMIP5 models. The spatial pattern of trends is examined in [section 5](#). Then, in [section 6](#) we undertake a detailed intercomparison of changes in sea level pressure and surface winds in various observations, reanalysis products, and the CMIP5 models over the more recent and reliable satellite era. In the final section we synthesize our findings and draw some broader conclusions.

## 2. Data and methods

We use monthly mean sea level pressure, 10-m zonal wind speed fields ( $u_{10m}$ ), and surface eastward wind stress from ensemble member 1 from 30 CMIP5 models [ACCESS1.0, ACCESS1.3, BCC\_CSM1.1, BCC\_CSM1.1(m), BNU-ESM, CanESM2, CMCC-CM, CMCC-CMS, CNRM-CM5, CSIRO Mk3.6.0, GISS-E2-H, GISS-E2-H-CC, GISS-E2-R, GISS-E2-R-CC, HadCM3, HadGEM2-AO, HadGEM2-CC, HadGEM2-ES, INM-CM4, IPSL-CM5A-LR, IPSL-CM5A-MR, IPSL-CM5B-LR, MIROC5, MIROC-ESM, MIROC-ESM-CHEM, MPI-ESM-LR, MPI-ESM-MR, MRI-CGCM3, NorESM1-M, and NorESM1-ME; expansions of acronyms are available online at <http://www.ametsoc.org/PubsAcronymList>]. We also use the equivalent output from six reanalyses, listed with their abbreviations, references, and data sources in [Table 1](#). The Twentieth Century Reanalysis (20CR; [Compo et al. 2011](#)) is an ensemble reanalysis consisting of 56 members. The 20CR ensemble members are not “free running” like the CMIP5 models, but rather they are produced with an ensemble Kalman filter data assimilation system to estimate the state of the atmosphere every 6 h ([Compo et al. 2011](#)). The spread across the 20CR ensemble provides the uncertainty of that estimate, arising from “atmospheric dynamics ... imperfect observations and a finite-ensemble first guess generated using an imperfect NWP model” ([Compo et al. 2011](#), p. 4). The spread across the 20CR ensemble does not represent large-scale differences in internal variability (e.g., phase of the SAM or ENSO), since all ensemble members are constrained to follow the observations. Hence, we consider the spread across the 20CR ensemble to represent “observational uncertainty.” For both CMIP5 and 20CR we perform our analysis on the individual ensemble members, and then compute an ensemble mean with an associated uncertainty (see below).

We use the gridded observational sea level pressure dataset, HadSLP2r, with reduced variance ([Allan and](#)

[Ansell 2006](#)). HadSLP2 extends from 1850 to 2004 and is based on quality controlled marine and terrestrial pressure observations that have been blended, gridded, and made spatially complete using a reduced space optimal interpolation. HadSLP2r extends this from 2005 to 2012 based on R1 fields ([Table 1](#)), which have been adjusted to have the same mean and variance as HadSLP2. (This “reduced variance” version is available online at <http://www.metoffice.gov.uk/hadobs/hadslp2>.) We also use the observed sea level pressures over 1958–2011 updated from [Marshall \(2003\)](#). [Marshall \(2003\)](#) used 12 individual stations to compute the proxy zonal mean SLPs at 40°S and 65°S (six stations near each latitude circle). Additional observationally based SAM reconstructions exist (e.g., [Jones et al. 2009](#); [Visbeck 2009](#)), and have previously been compared with each other ([Ho et al. 2012](#)), but we do not make use of them here.

The cross-calibrated multiplatform (CCMP) ocean surface wind vector analyses of [Atlas et al. \(2011\)](#) is used for  $u_{10m}$  winds and pseudo-wind stress fields over the period 1988–2011. The data were downloaded from the Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colorado (available online at <http://rda.ucar.edu/datasets/ds744.9/>). The supplied zonal pseudo-wind stress ( $u^2$ ) is converted to wind stress as:  $\tau_x = \rho c_d u^2$ , where  $\rho = 1.2 \text{ kg m}^{-3}$  is the density of air and  $c_d = 1.4 \times 10^{-3}$  is a dimensionless drag coefficient. CCMP is created using a variational analysis method (VAM), which takes in data from satellite radiometers and scatterometers, as well as ship and buoy observations. Observations are adjusted to the 10-m level assuming neutral stability. The VAM combines the data in a best fit, while satisfying smoothness and dynamical constraints. The procedure also requires a first-guess field, which comes from the ERA-40 reanalysis from July 1987 to December 1998, and from ERA-Interim thereafter ([Atlas et al. 2011](#)). Here we refer to CCMP as “satellite observations,” while acknowledging the presence of other observational inputs, and the reanalysis-based first guess.

Prior to any analysis, all the model, reanalysis, and observation data were remapped to a common  $1^\circ \times 1^\circ$  grid, using a distance weighting algorithm. The unitless SAM index is often calculated as the difference between the normalized sea level pressure at 40° and 65°S after [Gong and Wang \(1999\)](#). However, normalization (i.e., subtracting the mean and dividing by the standard deviation) removes systematic biases in the pressure at each latitude. Our nonnormalized SAM index is calculated as the zonal mean sea level pressure difference between 40° and 65°S in hectopascals (across all longitudes), as in [Gillett and Fyfe \(2013\)](#), except where noted.

Alternatively, where noted (and in Figs. 9 and 10), the SAM index is calculated in the same way but using only data from the 12 locations coincident with the stations used by Marshall (2003). The strength of the westerly jet is taken as the maximum of the zonal mean u10m between 20° and 70°S (in meters per second). The position of the jet is taken as the latitude, in degrees, at the jet maximum. The jet width is taken as the range of contiguous latitudes between 20° and 70°S (in degrees latitude), where the zonal mean u10m is positive. Where appropriate, seasonal averages were constructed as a simple (unweighted) mean over the 3-month periods DJF, March–May (MAM), JJA, and September–November (SON) respectively (seasonal means were computed from monthly indices, where applicable). Trends in the SAM index and jet properties were computed over various different time intervals (years), to allow for comparison with different observational products that each cover a limited period.

The analysis carried out in this paper was performed with the aid of IPython (Pérez and Granger 2007), and graphics were produced with matplotlib (Hunter 2007), version 1.4.3 (Droettboom et al. 2015). The analysis is fully reproducible with the open source code available from Swart (2015).

#### Ensembles, uncertainty, and statistics

Observed and simulated climate trends contain various sources of uncertainty that must be properly accounted for when formulating statistical tests (Fyfe et al. 2013; Santer et al. 2008). In this section we outline the sources of uncertainty in the CMIP5 ensemble of model simulations, the 20CR observational ensemble, and other observations. We then discuss appropriate statistical tests for (i) determining if observed and simulated trends are (in)consistent and (ii) determining whether an ensemble mean trend is significantly different from zero. A representation of the simulated and observed trends can be given by

$$b_{ij}^m = u^m + \text{Eint}_{ij}^m + E_i^m, \quad i = 1, \dots, n^m, \quad j = 1 \quad \text{and} \quad (1)$$

$$b_k^o = u^o + \text{Eint}_k^o + E_k^o, \quad k = 1, \dots, n^o, \quad (2)$$

where  $b_{ij}^m$  and  $b_k^o$  are trends calculated from single model runs or the observations;  $u^m$  and  $u^o$  are the true, unknown, deterministic trends due to external forcing in the model and observations (Fyfe et al. 2013);  $u^m$  is the component of the trend common to all models (in the limit as the collection of exchangeable models grows infinitely large); and  $\text{Eint}_{ij}^m$  and  $\text{Eint}_k^o$  are perturbations to  $b_{ij}^m$  and  $b_k^o$  respectively due to internal variability. For the models this is different for each run, but there is

essentially only one realization of internal variability for the observations. Also,  $E_i^m$  is the perturbation to  $b_{ij}^m$  that is introduced by model error in model  $i$ , and  $E_k^o$  is the observational error;  $n^m$  is the number of models (and here we have only used one realization for each model,  $j = 1$ ) and  $n^o$  is the size of the observational ensemble (which could be  $n^o = 56$  for 20CR but is  $n^o = 1$  for the other observations).

To assess (i) whether the observed and simulated trends are consistent, we formulate the null hypothesis that the observed and simulated trends are equal:

$$H_0: u^m = u^o. \quad (3)$$

An estimator of  $u^m - u^o$  is  $\overline{b^m} - \overline{b^o}$ , where the overbar represents the average over all ensemble members. A test of null hypothesis may be given by a test similar to the Student's  $t$  test for the difference in means:

$$t = \frac{\overline{b^m} - \overline{b^o}}{\sqrt{\frac{s_{b^m}^2}{n^m} + S^{o2}}}, \quad (4)$$

where  $s_{b^m}^2/n^m$  is an estimate of the variance of the mean of the model-mean trend, which arises due to model errors ( $E_i^m$ ) and internal variability ( $\text{Eint}_{ij}^m$ ) that are present in the CMIP5 ensemble (Santer et al. 2008). The uncertainty estimated from the CMIP5 ensemble in this way accounts for all the uncertainty terms in (1). Also,  $S^{o2}$  is an estimate of the variance of the observed trend. This term should account for the uncertainty due to observational error ( $E_k^o$ ) and internal variability ( $\text{Eint}_k^o$ ) present in the observations. Estimating the observational error requires more than a single observation (often not available) and the influence of internal variability is hard to estimate robustly given that there is only a single observed realization of this variability.

The uncertainty in the observed trend due to internal variability ( $\text{Eint}_k^o$ ) can be estimated using the standard error of the trend adjusted for autocorrelation (e.g., Santer et al. 2008). Alternatively, this uncertainty can be estimated by making the assumption that the variance of the observed and simulated trends is equal (i.e., the spread across the model ensemble is used as an estimate of the influence of internal variability on the observations). Under this assumption of equal variances with a single observational estimate yields

$$t = \frac{\overline{b^m} - \overline{b^o}}{s_{b^m} \sqrt{\frac{1}{n^m} + 1}} \approx \frac{\overline{b^m} - \overline{b^o}}{s_{b^m}}. \quad (5)$$

To reject the null hypothesis of equal trends at the 5% level requires  $|\overline{b^m} - \overline{b^o}| > cs_{b^m}$ , where  $c$  is the 97.5th

percentile of the Student's  $t$  distribution with  $n^m - 1$  degrees of freedom. Since  $c \approx 2$  (for  $n_m \geq 10$ ), a statistically significant difference requires that the observations lie outside of two standard deviations ( $2\sigma$ ) from the model mean trend. For a large enough sample size of normally distributed data, this is equivalent to saying that the observed trend should lie outside of the 2.5th–97.5th percentile of the simulated trends (which we shall show in all trend plots). The consistency of simulated and observed trends can thus be evaluating by asking whether the observations fall within the 2.5th–97.5th percentile of the simulated trends (Swart et al. 2015; Gillett and Fyfe 2013; Gillett et al. 2013).

In the case of the 20CR ensemble, the observational uncertainty ( $E_k^o$ ) may also be directly quantified as  $s_{b^o}^2/n^o$ , where  $n^o = 56$  is the number of members in the 20CR ensemble and  $s_{b^o}^2$  is the variance across the ensemble. Note, however, that we cannot simply replace  $S^{o2}$  in (4) with  $s_{b^o}^2/n^o$ . The reason for this is that on average over the free-running CMIP5 models the influence of internal variability is zero ( $\overline{Eint}_{ij}^m = 0$ ), but for each 20CR ensemble member the influence of internal variability is constrained to be the same by the observations ( $Eint_k^o \neq 0$ ). If we neglected to account for this, as the number of model and 20CR ensemble members increased, we would inevitably find significant differences,  $\overline{b^m} - \overline{b^o} \neq 0$ , even if the true underlying trends were equal ( $u^m - u^o = 0$ ), simply because of differences in internal variability. We could instead add the 20CR observational uncertainty into the test above to make it even more conservative:

$$t = \frac{\overline{b^m} - \overline{b^o}}{\sqrt{s_{b^m}^2 \left( \frac{1}{n^m} + 1 \right) + \frac{s_{b^o}^2}{n^o}}} \quad (6)$$

but since  $s_{b^m} \gg s_{b^o}^2/n^o$ , neglecting this term makes little practical difference. Therefore throughout we will consider the observed and simulated trends to be significantly different at the 5% level when the observed trend (ensemble mean for 20CR) falls outside of the 2.5th–97.5th percentile of the simulated trends.

The above tests relate to the question of whether the observed and simulated trends are consistent. The second topic (ii) that we are interested in assessing is whether a given ensemble mean trend is significantly different from zero. The appropriate Student's  $t$  test of the null hypothesis that the mean trend is zero is given by

$$t = \frac{\overline{b}}{s_b \sqrt{\frac{1}{n}}}, \quad (7)$$

which could be tested for the models or 20CR ensemble. The uncertainty in the mean trend is represented by the (95%) confidence interval, which is given by

$$\overline{b} \pm \frac{cs_b}{\sqrt{n}}, \quad (8)$$

where  $c$  is the 97.5th percentile of the Student's  $t$  distribution with  $n - 1$  degrees of freedom (von Storch and Zwiers 1999). We also plot this 95% confidence interval for the CMIP5 and 20CR trends. This definition of the 95% confidence interval is used for both time series (e.g., Fig. 1, shaded areas) and trends (e.g., Fig. 2, solid vertical bars).

### 3. Observed and simulated changes in the SAM and westerly jet

#### a. Time series

Over 1871–1950 the annual mean SAM index from 20CR, HadSLP2r, and the CMIP5 models hover around 25 hPa on average (Fig. 1a). Over this period, the CMIP5 ensemble mean has an equatorward biased jet position relative to 20CR (Fig. 1c), but the simulated jet strength and width are roughly equivalent to those in 20CR (Figs. 1b,d).

Prior to 1950, these metrics show pronounced interannual and decadal time scale variability, but no significant secular trends. From around 1950 onward, HadSLP2r and 20CR both show a clear shift toward larger values of the SAM index. Jet strength shows a simultaneous increase in 20CR over this period, while consistent changes in jet position and width are less evident. The CMIP5 models also show an increase in the SAM index and jet strength, although the simulated increase generally appears lower than that seen in the 20CR and HadSLP2.

To more closely compare changes between 20CR, HadSLP2r, and the CMIP5 models, we next consider linear trends in these metrics over 1951–2011. The R1 data is also available over this period, but we exclude it here because it is known to exhibit spurious trends in the SAM (Marshall 2003). However, in section 6 we will conduct a more thorough interobservational product comparison.

#### b. Linear trends by season over 1951–2011

Over 1951–2011 both HadSLP2r and 20CR show a positive SAM trend during all seasons (Fig. 2a). The HadSLP2r SAM trends are generally a little smaller than those in 20CR, and exhibit more seasonality. The CMIP5 models also exhibit positive trends on average during all seasons, but the model trends show the

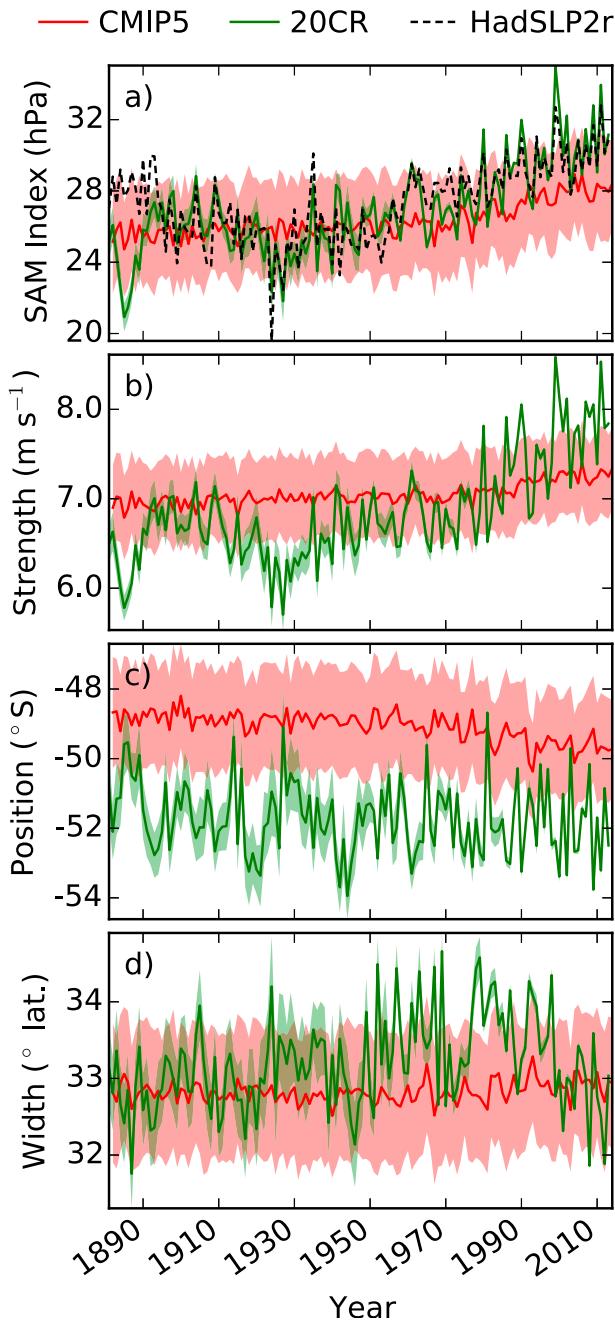


FIG. 1. Time series of the annual mean (a) SAM index and westerly jet (b) strength, (c) position, and (d) width over 1881–2013. The shaded areas envelop the 95% confidence interval about the mean for the CMIP5 and 20CR ensembles.

opposite seasonality to HadSLP2r and 20CR, being largest in DJF and smallest in JJA on average, as would be expected from the ozone-related forcing (Son et al. 2010; Thompson et al. 2011).

During DJF the model-mean SAM trend is almost identical to that seen in 20CR, consistent with Gillett

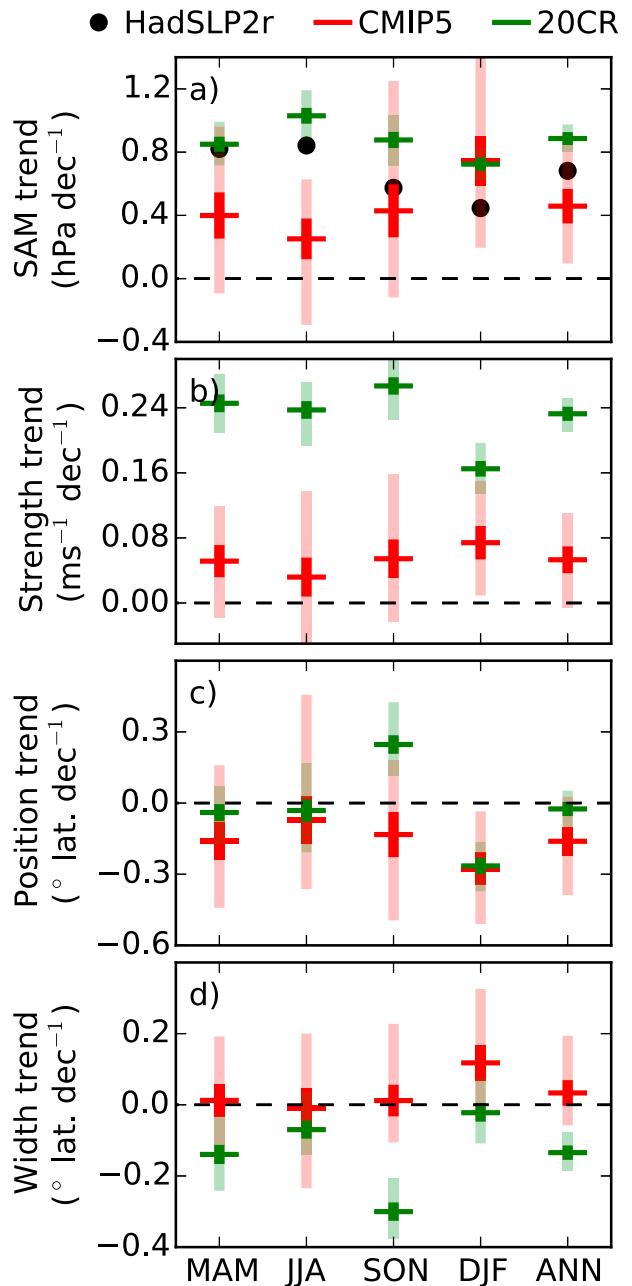


FIG. 2. Trends over 1951–2011 in (a) the SAM index and westerly jet (b) strength, (c) position, and (d) width by season in HadSLP2r, 20CR, and CMIP5. The CMIP5 and 20CR ensemble mean trends are given by the horizontal lines (red and green respectively), the 95% confidence interval is given by the solid vertical bars, and the 2.5th–97.5th percentile of trends in the individual ensemble members is given by the light vertical bars.

and Fyfe (2013). However, during the austral winter (JJA) the models significantly underestimate the SAM trend relative to 20CR and HadSLP2r. The models also significantly underestimate the annual (ANN) mean trend SAM relative to 20CR. Significance in this sense is

determined from the fact that the 20CR ensemble mean trend lies outside of the 2.5th–97.5th percentile of CMIP5 trends, and thus we can reject the null hypothesis that the 20CR and CMIP5 trends come from the same distribution, at the 5% level (see section 2).

For jet strength, 20CR exhibits a trend of between 0.15 and 0.25  $\text{m s}^{-1} \text{decade}^{-1}$  (Fig. 2b). The CMIP5 models also show positive jet strength trends in all seasons on average. Yet for jet strength, the modeled trends are significantly smaller than for 20CR in all seasons, with the annual mean trend being about 5 times weaker in the models. In all seasons, the 20CR-mean trends lie outside the 2.5th–97.5th percentile of CMIP5 trends.

Trends in jet position vary in sign over the seasons in 20CR (Fig. 2c), with a small, nonsignificant trend in the annual mean. The CMIP5 models show poleward trends in jet position that are significant at the 5% level during all seasons except JJA. The largest poleward trend in jet position occurs in DJF, with nearly identical trends in 20CR and the CMIP5 mean. Jet width does not exhibit any significant trends in the CMIP5 models except for in DJF, which has a broadening trend of about 0.1° latitude per decade on average. 20CR, by contrast, shows narrowing trends in all seasons, especially SON.

The disagreements between 20CR, HadSLP2r, and the CMIP5 models identified here at least partly reflect spuriously large trends in 20CR and HadSLP2r, rather than an underestimation of the “true” trend by the CMIP5 models, as we shall see in sections 5 and 6. Regardless, our key focus here is to highlight that over 1951–2011 the DJF jet strength trends differ by more than a factor of 2 between 20CR and CMIP5, while their SAM trends are similar. Indeed, it is not valid to assume that trends in the SAM index and jet properties are directly interchangeable, as we show in the following section.

#### 4. The relationship between changes in the SAM index and westerly jet properties

##### a. A simple theoretical model

To illustrate the relationship between the SAM index and the kinematic properties of the jet, we use a simple geostrophic model. The zonal mean zonal velocity  $U$  is given by a Gaussian, with a specified position  $\Phi$ , strength  $\eta$ , and width  $\sigma$ :

$$U(\phi) = \eta \exp \left[ -\frac{(\phi - \Phi)^2}{2\sigma^2} \right], \quad (9)$$

where  $\phi$  is latitude. In this model, the zonal jet velocity is related to the surface pressure field via geostrophy, such that

$$P(\phi) = -\rho \int fU \delta y, \quad (10)$$

where  $f = 2\omega \sin(\phi)$  is the Coriolis parameter, given the angular rotation rate of Earth,  $\omega = 7.3 \times 10^5 \text{ s}^{-1}$ , and  $\rho = 1.2 \text{ kg m}^{-3}$  is the density of air. We can use this idealized model to examine how the SAM changes are related to changes in an individual kinematic property of the jet. We start with default values of  $\eta = 7 \text{ m s}^{-1}$ ,  $\Phi = -48^\circ$ , and  $\sigma = 6^\circ$ , and then vary each of these three parameters individually, while keeping the other two fixed (Fig. 3).

Changes in jet strength and the SAM index are linearly related, such that an increasing SAM is associated with a strengthening jet (Fig. 3d). Changes in jet position and the SAM index are inversely related, with a poleward shifting jet corresponding to a strengthening SAM (Fig. 3e). However, the relationship is not linear. The increase in SAM is largest per unit of poleward shift for jets which are more equatorward. For example, for a jet that is centered at 45°S a poleward shift of 1° latitude is associated with an increase in the SAM index of about 1.7 hPa, while for a jet that is centered at 50°S the increase in SAM is less than 1 hPa for the same 1° poleward shift. Changes in the SAM index are also proportional to changes in jet width, but are generally more sensitive to jet narrowing than to jet widening.

The chief value of the model used here is to illustrate that changes in the SAM index can be influenced by changes in all three kinematic properties of the jet, as found previously (Monahan and Fyfe 2006, 2008). Changes in the SAM may be associated with changes in one kinematic property of the jet, while the other kinematic properties remain constant or even change in the opposite sense.

##### b. SAM–jet relationships in the CMIP5 models and 20CR

We first consider the relationships between the trends in the SAM index and the kinematic properties of the jet for a single season, DJF, when the simulated mean changes are largest. Trends in the DJF-mean SAM index over 1951–2011 are significantly correlated with trends in all three kinematic properties of the jet across the CMIP5 models (Figs. 4a,c,e). The sign of the relationships are as predicted by the simple geostrophic model.

The SAM index trend is also significantly correlated with the climatological position and inversely correlated with the climatological jet strength across the CMIP5 models (Figs. 4b,d). The correlation between SAM index trend and climatological position was also predicted by the simple geostrophic model: the change in SAM index is larger per degree poleward shift in jet position

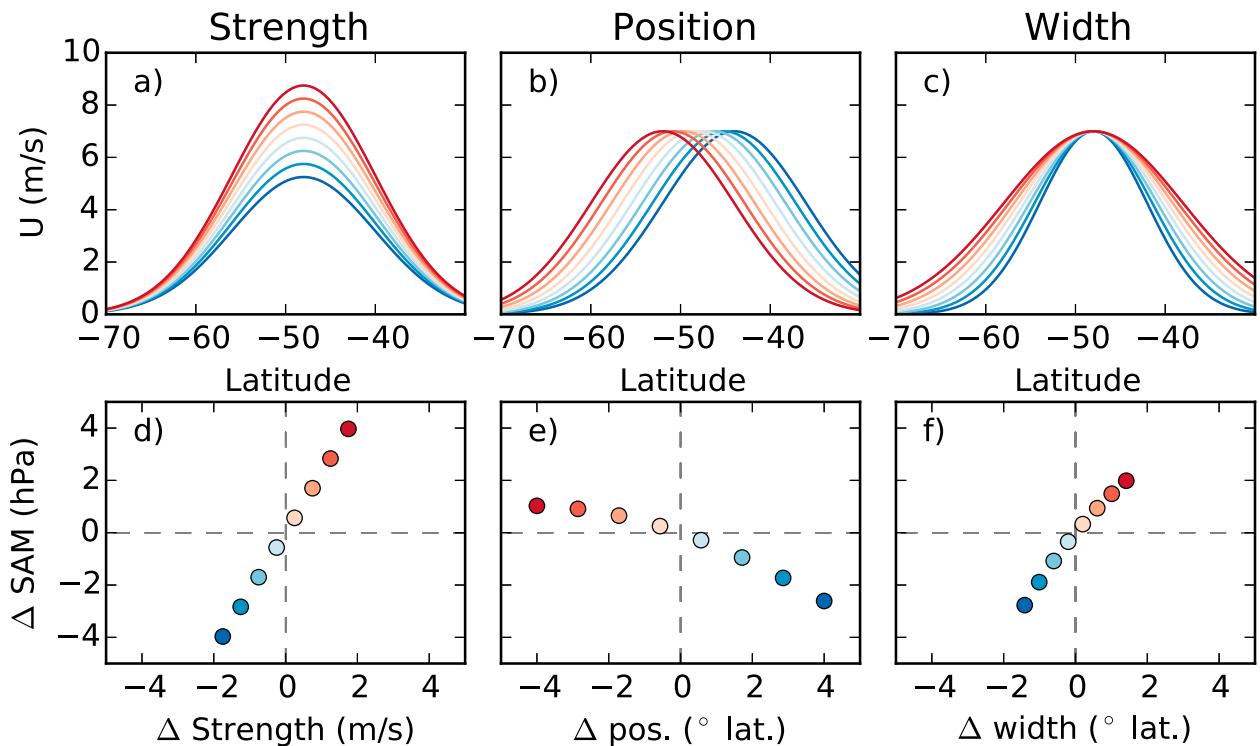


FIG. 3. The (a) strength, (b) position, and (c) width of an idealized Gaussian jet, and the relationship between changes in the (d) strength, (e) position, and (f) width of the idealized jet and changes in the corresponding SAM index.

for models that start with a more equatorward climatological position than for those with a more poleward climatological position (Fig. 3e). In addition, it is known that jets with a more equatorward climatological position experience larger historical poleward trends in position (Kidston and Gerber 2010; Bracegirdle et al. 2013).

Given these correlations showing that models with large SAM trends tend to have large trends in jet strength, position, and width, it might appear that the SAM index can be used to infer changes in the jet. However, the relationships between trends in the SAM and the kinematic properties of the jet change by season. This is demonstrated for the relationship between trends in the SAM index and jet strength (Fig. 5). Further, the relationships between the SAM and the jet differ between the CMIP5 and 20CR ensembles (Figs. 4 and 5), and also differ when comparing the six reanalyses in Table 1 to the CMIP5 models over the satellite era (not shown). The correlations between the SAM and jet properties within a given model also vary significantly over the CMIP5 ensemble. For example, the correlation between the SAM index and jet strength varies from  $r = 0.44$  in IPSL-CM5B-LR to  $r = 0.84$  in ACCESS1.0. Therefore, given the variability of these SAM jet relations across models and by season, trends in the SAM index cannot be used as a direct proxy for trends in the

jet, as previously shown (Thomas et al. 2015; Monahan and Fyfe 2006, 2008).

The reasoning above also explains how it is that 20CR-mean and the CMIP5-mean SAM trends can be similar, while the 20CR mean jet strength trend is much larger than seen in the models on average (Figs. 4a and 2). The poleward trend in jet position is similar between 20CR and the models on average (Fig. 4c); however, the models show a positive jet width trend (broadening) on average, while 20CR shows a small negative width trend on average (Fig. 4e). Thus, the broadening of the jet in the models makes it dynamically consistent for them to have the same SAM trend as 20CR, even though their jet strength trends are much weaker than in 20CR. In addition, the models have an equatorward biased climatological jet position relative to 20CR, and more equatorward jets are associated with larger changes in SAM (Fig. 4d) per unit poleward shift in jet position. The apparent discrepancies between trends in the SAM and jet strength are thus resolved.

## 5. Spatial structure of historical trends

The trends in monthly SH sea level pressure and winds also have important spatial structure. The SLP trend maps are shown over 1951–2004 (Fig. 6) because the

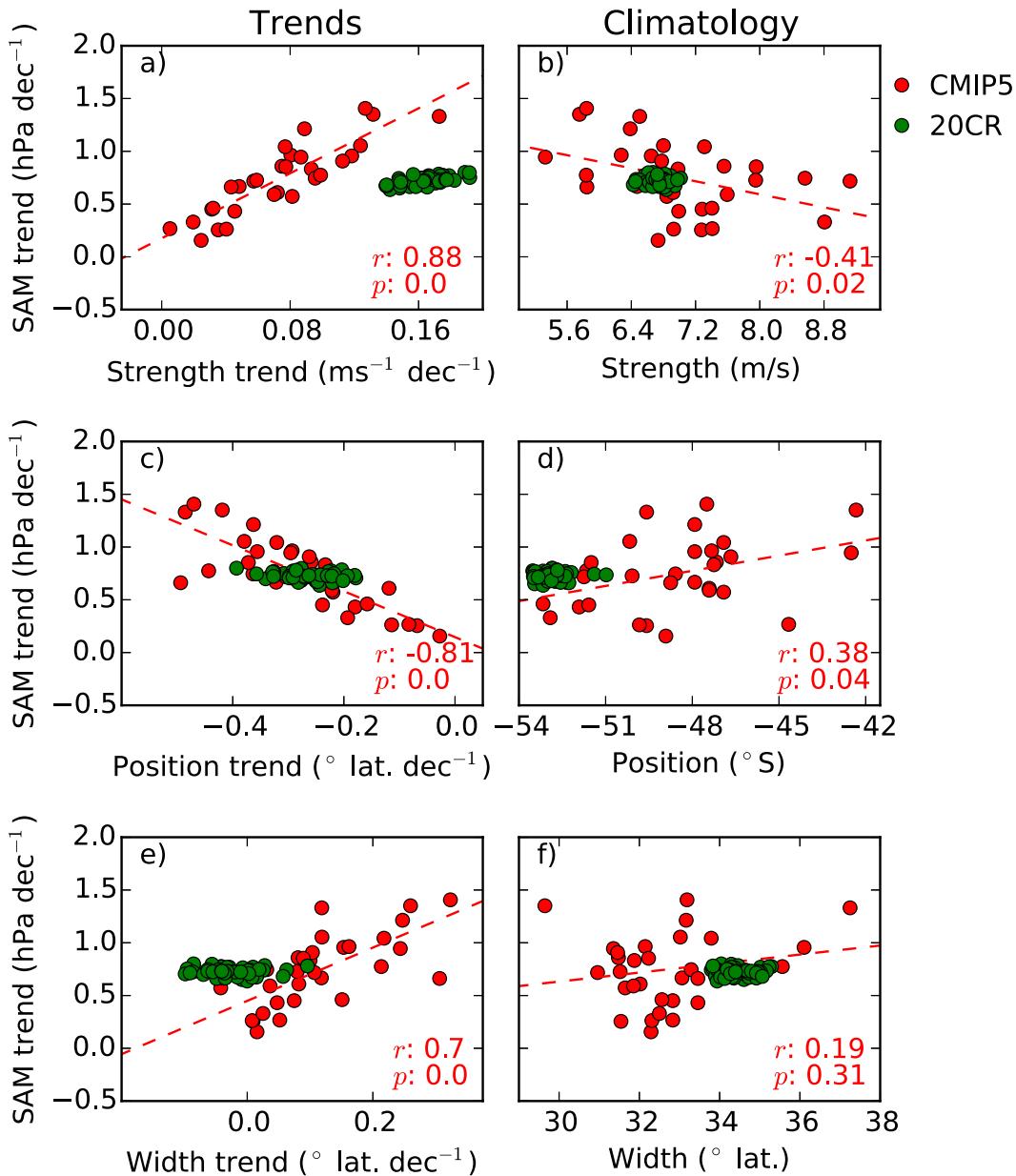


FIG. 4. The relationship between DJF-mean trends over 1951–2011 in the SAM index and trends in jet (a) strength, (c) position, and (e) width as well as climatological jet (b) strength, (d) position, and (f) width for the 30 individual CMIP5 simulations and 20CR. Numbers in the bottom right of the panels give the correlation coefficient  $r$  and  $p$  value of the relationship.

HadSLP2r data become unreliable after 2005, as we shall see below. The HadSLP2r trend pattern is dominated by circumpolar wide negative trends in SLP south of 50°S, with a bull's-eye of strong negative trends focused over the South Pacific. To the north HadSLP2r shows an increase in pressure near 40°S, focused south of Africa. The 20CR mean trends shows generally very similar patterns. In the CMIP5 mean trend, there are similar circumpolar bands of positive trends centered on

40°S, and negative trends south of 50°S. However, the CMIP5 models do not show the focused region of large negative trends in the South Pacific, or increasing SLP south of Africa. In both of these regions, the HadSLP2r trends lie outside the 2.5th–97.5th percentile of individual model trends, indicating that the differences are significant (Fig. 6). These differences may occur because the CMIP5 models have difficulty correctly simulating variations in the wavenumber-3 pattern around Antarctica

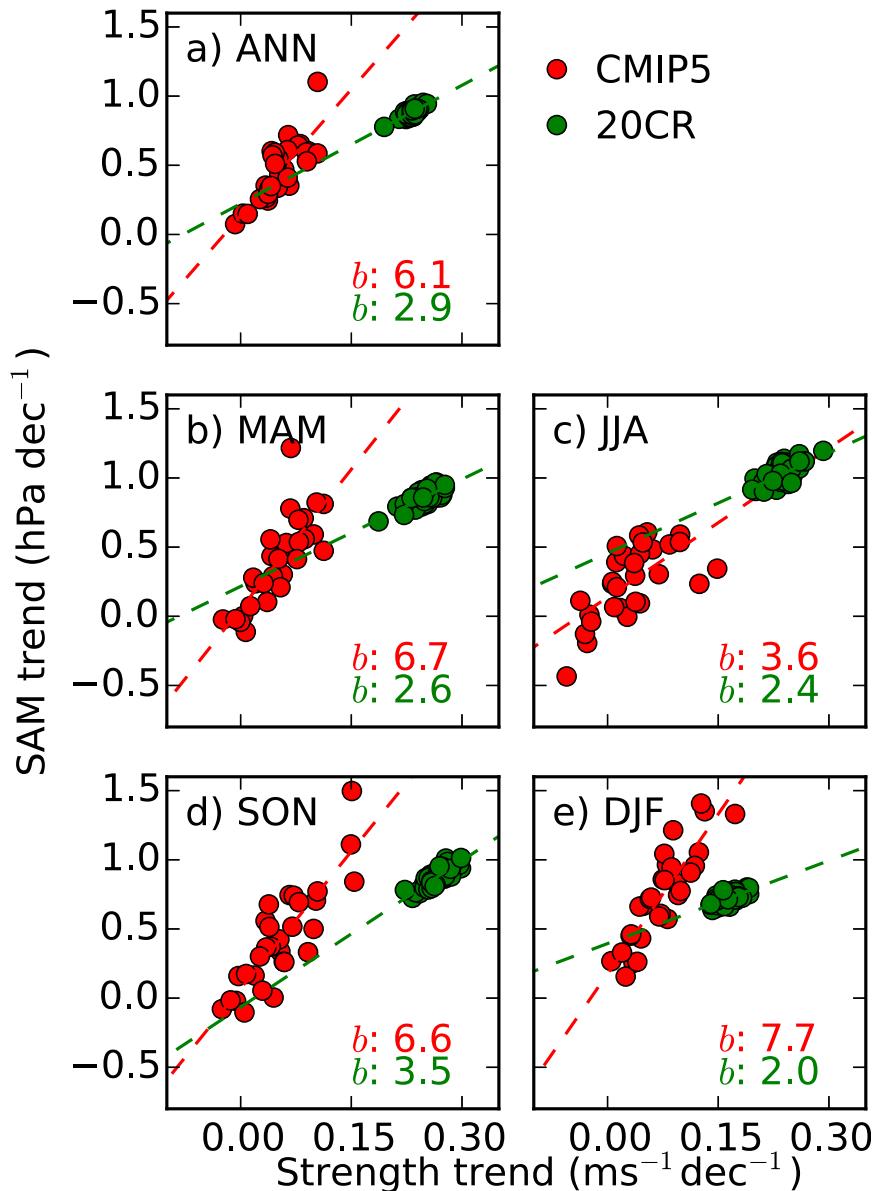


FIG. 5. The relationship between trends over 1951–2011 in the SAM index and trends in jet strength over various seasons. The  $b$  values given in the bottom right of the panels are the slope of the regression lines.

(Marshall and Bracegirdle 2015), or because of uncertainties in the observations described below.

Wind trends are shown for 20CR and the CMIP5 mean (Fig. 7). 20CR shows a band of large positive trends centered on the jet core near 50°S, with regions of negative trends on either side. The CMIP5 mean also shows strengthening trends, but they are much weaker and poleward displaced relative to the 20CR trends. Thus, the CMIP5 models show a strengthening on the poleward flank of the jet on average. The anomaly map shows a tripole of differences, indicating the shifted nature of the

trends in the CMIP5 mean, relative to 20CR, with the differences being significant nearly everywhere.

In the previous sections we have shown how the CMIP5 models have trends in SLP and surface winds that differ significantly from HadSLP2r and 20CR. These differences are evident in integrated metrics like the SAM index and zonal-mean jet strength, and as we have shown here are regionally focused in the southeastern Pacific. However, the southeastern Pacific is one of the most data-sparse regions and significant uncertainties exist in the observations, and from the

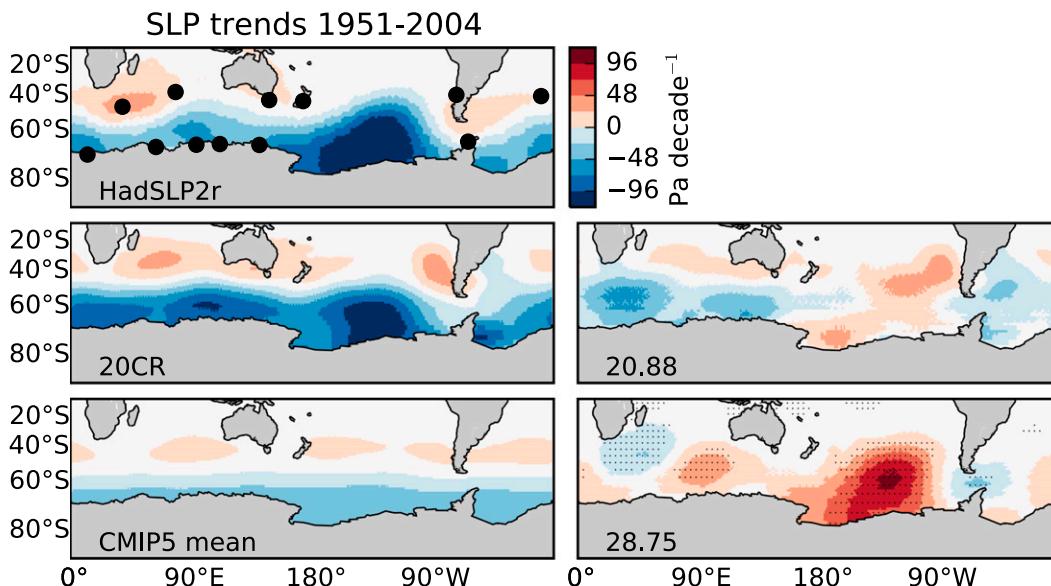


FIG. 6. (left) Trends in monthly sea level pressure over 1951–2004 for (top)–(bottom) HadSLP2r, 20CR, and the CMIP5 mean, and (right) anomalies relative to the HadSLP2r trends. The areas where the HadSLP2r trend lies outside the 2.5th–97.5th percentile of trends in the individual CMIP5 simulations is shown by stippling in the bottom-right panel. Numbers in the right-hand panels give the root-mean-square difference with HadSLP2r ( $\text{Pa decade}^{-1}$ ). Black dots in the panel for HadSLP2r show the position of the Marshall (2003) stations.

infilling methodologies associated with HadSLP2r (Allan and Ansell 2006).

To demonstrate this, the uncertainty in the 20CR SLP and u10m trends is shown as 2 times the standard deviation in trends across the 56-member 20CR ensemble (Fig. 8). The  $2\sigma$  spread is largest in the southeastern Pacific, and it represents about 20% of the magnitude of the mean trends. The 20CR ensemble also suffers from spurious trends associated with a changing observational network that are not fully quantified by the ensemble spread discussed above (Wang et al. 2013). In the

following sections, we address these issues by conducting an intercomparison of available observational and reanalysis products.

## 6. Intercomparison of changes across observational products and models

### a. SAM index computed at Marshall station locations

One of the most reliable records of changes in the SH SLP is from the station based estimates updated from

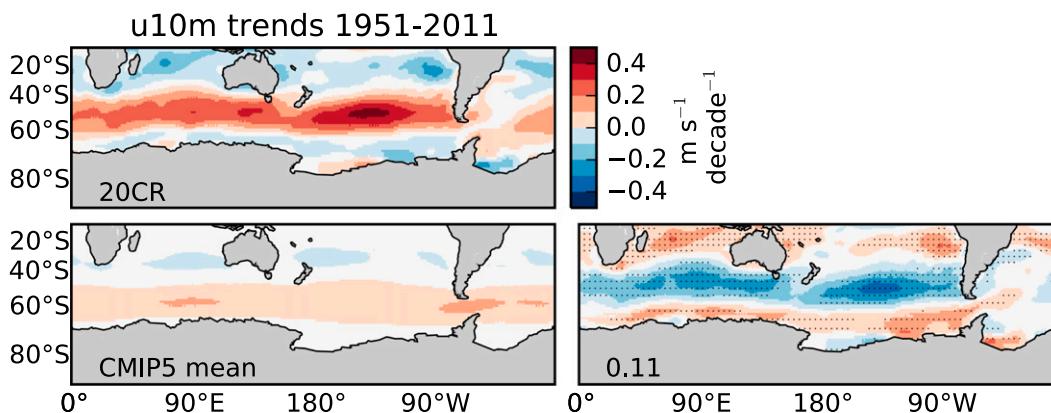


FIG. 7. (left) Trends in u10m over 1951–2011, and (right) anomalies relative to the 20CR trend. The areas where the 20CR trend lies outside the 2.5th–97.5th percentile of trends in the individual CMIP5 simulations is shown by stippling in the bottom-right panel. The number in the bottom left of the right panel gives the root-mean-square difference between the 20CR and CMIP5 mean trends ( $\text{m s}^{-1} \text{decade}^{-1}$ ).

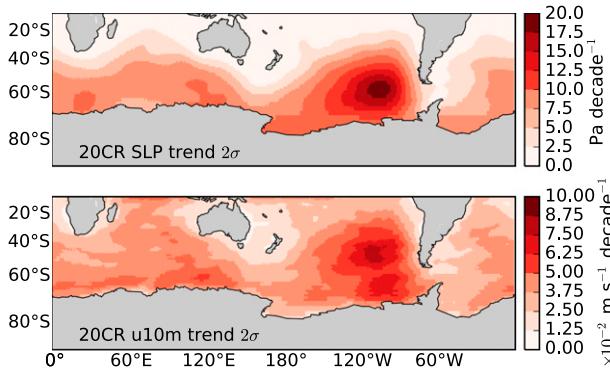


FIG. 8. The uncertainty in 20CR trends in SLP and  $u_{10m}$  over 1951–2011, given by 2 times the standard deviation of trends across the 56-member 20CR ensemble.

Marshall (2003). Data from six stations located near 40°S and an additional six stations near 65°S were averaged to give the mean SLP at those two latitudes respectively (for station positions see Fig. 6). Here, HadSLP2r, six reanalyses, and the CMIP5 models are subsampled at these same 12 locations in order to compare with the Marshall (2003) data.

In the time series of the mean pressure at 40°S it can be seen that the reanalyses and Marshall based observations have well-synchronized interannual variability (Fig. 9a). In all products a general long-term increase in SLP at 40°S is also evident. At 65°S, the observations and all six reanalyses show a long term decline in SLP (Fig. 9b). Biases here also occur principally in R1, which starts with a pressure that is about 8 hPa too high, and exhibits a large and spurious negative trend not seen in the observations of Marshall (2003) prior to about 1990. R2, which is a closely related product, has similar issues, and to a much lesser extent, 20CR. Since the 20CR spread is generally small after 1950 (Fig. 1), from here on we show only the 20CR ensemble mean. It can also clearly be seen that a large and spurious change occurs after 2005 at 65°S in HadSLP2r, coincident with when that product begins to be based on R1 output, and despite efforts to homogenize the dataset. Hence we limit all our spatial comparisons with HadSLP2r to the period before 2005.

The CMIP5 models on average have a pressure that is systematically low by about 1 hPa at 40°S and systematically high by about 4 hPa at 65°S, relative to the Marshall data (Figs. 9a,b). The SAM index shows the well-known long-term increase for the models, reanalyses, and observations (Fig. 9c). Biases, which largely stem from those at 65°S, are also clearly evident. To better assess the changes, SAM trends by season are also computed for two time periods (Fig. 10).

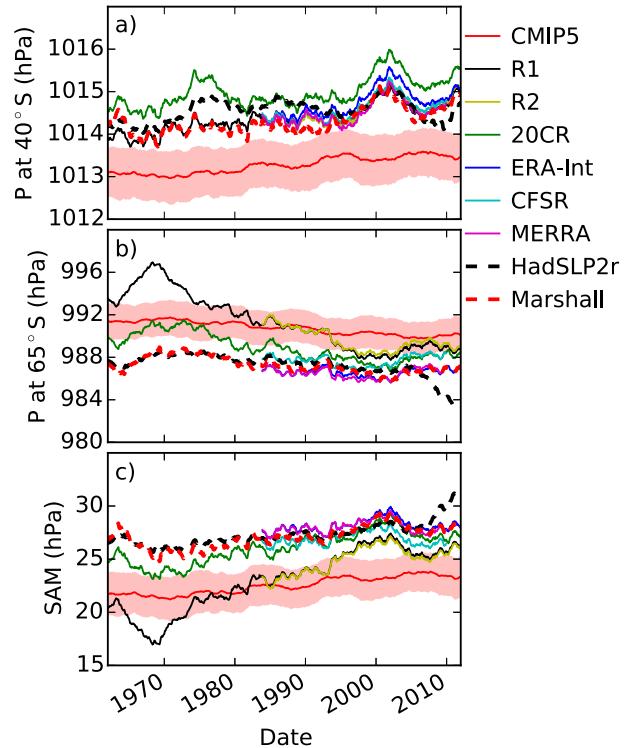


FIG. 9. Time series over 1962–2012 of pressure at (a) 40°S and (b) 65°S, and (c) the SAM index computed using only data from the locations of the stations used by Marshall (2003) and shown for the original Marshall (2003) data, HadSLP2r, six reanalyses, and the CMIP5 mean plus 95% confidence interval. All data have been smoothed with a 5-yr running mean.

Over 1958–2011, trends at 40°S are generally small and positive. Trends at 65°S are negative and larger, and show large biases for R1 relative to the Marshall and HadSLP2r observations. In the SAM index, the Marshall-based trend is positive in all seasons, except SON, when it is zero. The HadSLP2r trends also generally match the Marshall trends well, with the largest difference occurring in SON. The 20CR ensemble mean SAM trend is slightly larger than the trend observed in the Marshall data in all four seasons and the annual mean. The spread of CMIP5 trends over 1958–2011 includes the observed Marshall trend in all seasons except SON (Fig. 10). Interestingly, in the annual mean, the CMIP5 mean trend almost exactly matches the observed Marshall trend.

Over the shorter period from 1979 to 2009, most of the same conclusions hold. Trends at 40°S are small and positive, while trends at 65° are negative, larger, and less certain. The CMIP5 range of trends includes the Marshall observations in all seasons, and in the annual mean the CMIP5 mean trend is again almost identical to observed over the shorter satellite era.

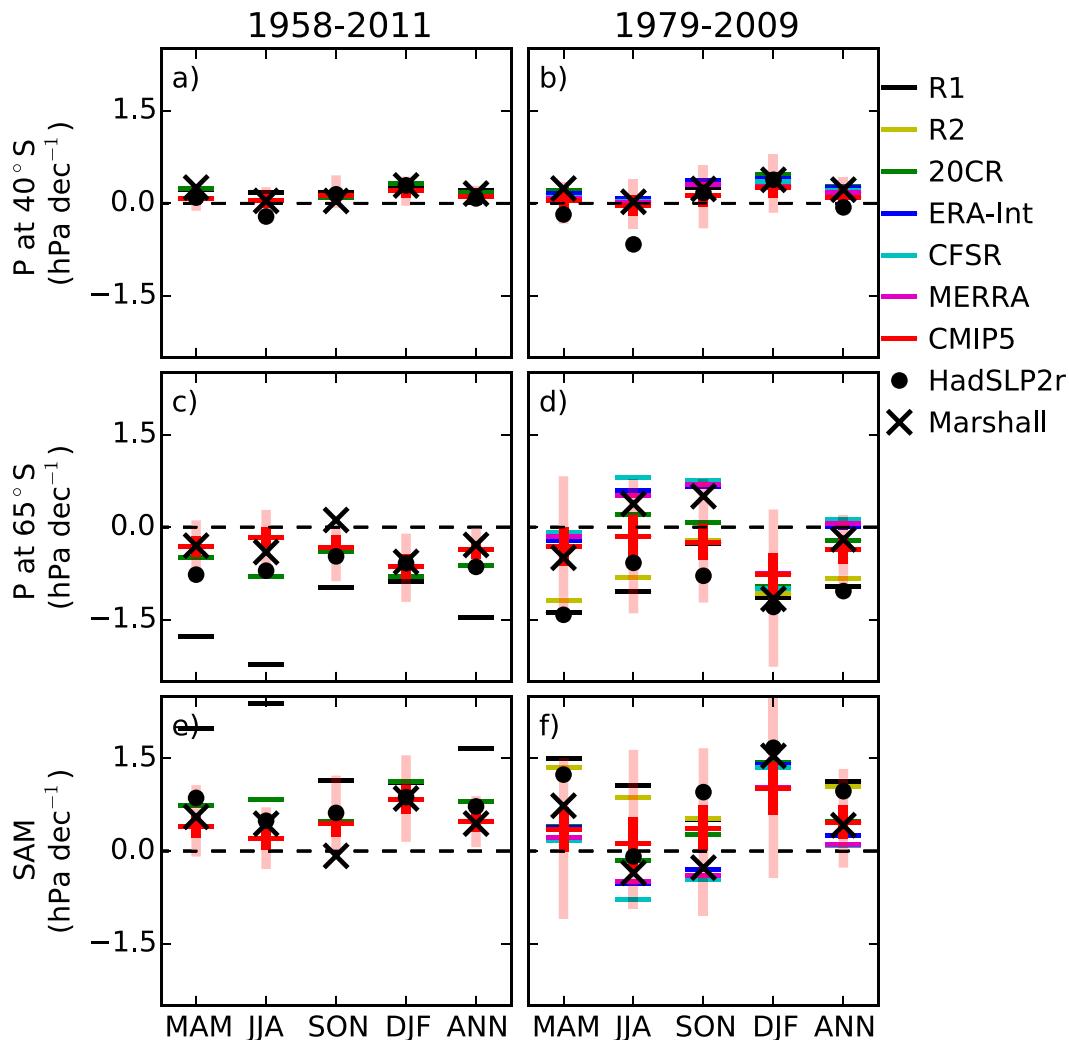


FIG. 10. Trends in (a),(b) pressure at  $40^{\circ}\text{S}$ , (c),(d) pressure at  $65^{\circ}\text{S}$ , and (e),(f) the SAM index computed using only data from the locations of the stations used by Marshall (2003) and shown for the periods (a),(c),(e) 1958–2011 and (b),(d),(f) 1979–2009. The CMIP5 mean trend is given by the horizontal red line, the 95% confidence interval is given by the solid red vertical bar, and the 2.5th–97.5th percentile of trends in the individual CMIP5 simulations is given by the light red vertical bar.

These findings suggest that there is little evidence that the CMIP5 models systematically underestimate the SAM trend. This is opposite to the conclusion in section 3, where the JJA (and annual mean) SAM trend (based on the zonal mean over all longitudes and over 1951–2011) in HadSLP2r and 20CR was found to be significantly larger than the CMIP5 trends. Much of the reason for this is that over 1951–2011 the largest SLP trends in 20CR and HadSLP2r occur in the southeastern Pacific, and this region contributes significantly to the overall SAM trend, but is also the most uncertain. In contrast, the Marshall-based SAM index considered in this section does not have any stations located in the southeastern Pacific (see Fig. 6) but has reliable trends due to

using a fixed observational network (Marshall 2003). In the following section we return to examining the spatial structure of trends over the full Southern Ocean.

#### *b. Spatial structure of trends over the recent past*

SLP trend maps were computed for 1979–2004, when all reanalysis products and HadSLP2r are available, and by ending in 2004 we avoid the continuity problems in HadSLP2r identified above (Fig. 11). The most prominent pattern in the HadSLP2r trends over this period is again the large negative and circumpolar trends in pressure south of about  $50^{\circ}\text{S}$ . Similar patterns are seen in R1, R2, and 20CR, but these products tend to overestimate the magnitude of the trends relative to

SLP trends 1979-2004

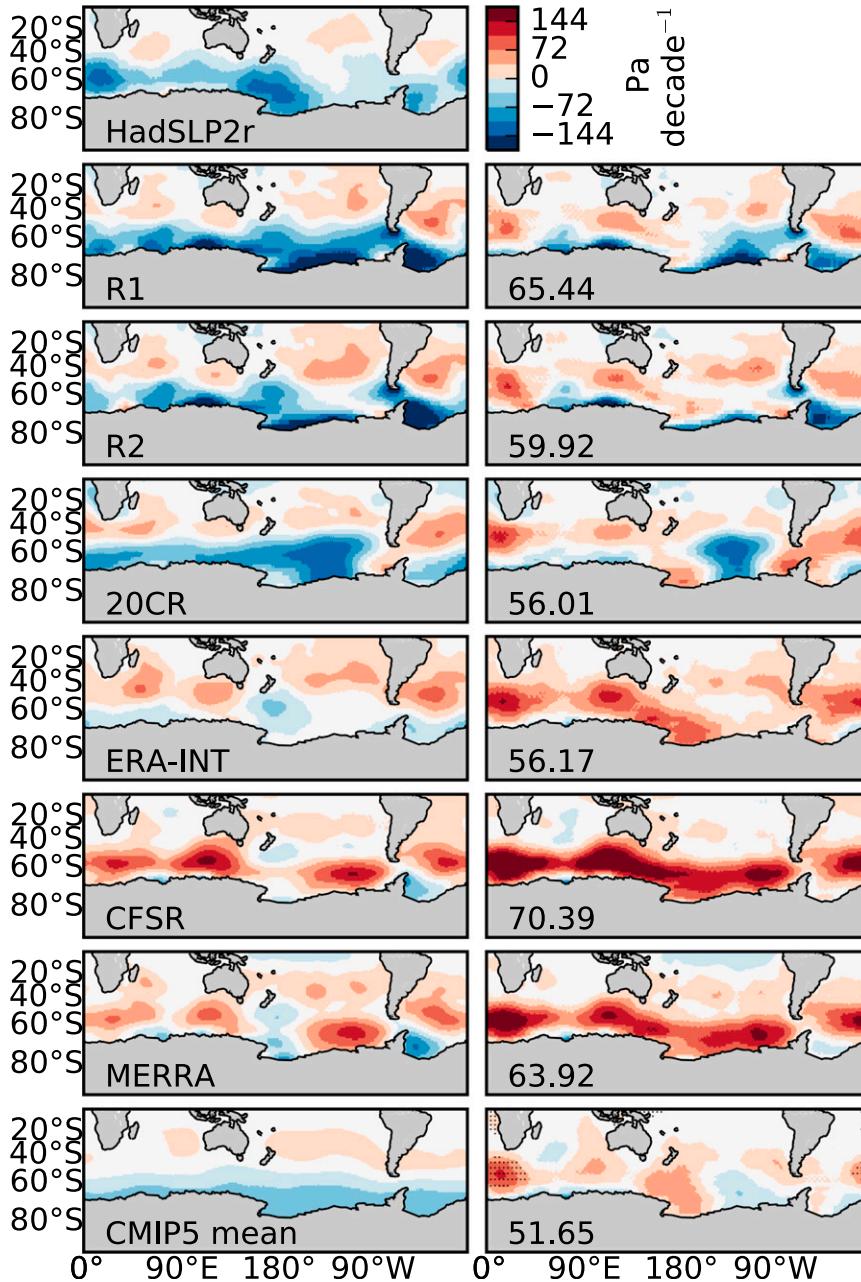


FIG. 11. (left) Trends in monthly mean sea level pressure over 1979–2004 for (top)–(bottom) HadSLP2r, six reanalyses, and the CMIP5 mean, and (right) anomalies relative to the HadSLP2r trends. The areas where the HadSLP2r trend lies outside the 5th–95th percentile of trends in the individual CMIP5 simulations is shown by stippling in the bottom-right panel. Numbers in the bottom left of the right panels give the root-mean-square difference with HadSLP2r (Pa decade<sup>-1</sup>).

HadSLP2r. CFSR and MERRA show the opposite, with large positive trends, and correspondingly, these products have the largest root-mean-square difference from the HadSLP2r observations. ERA-Interim also has SLP

trends that are a little too positive, but it has the best fit to the HadSLP2r observations after 20CR. The CMIP5 models show a similar pattern of trends to the observations, but generally with a weaker magnitude.

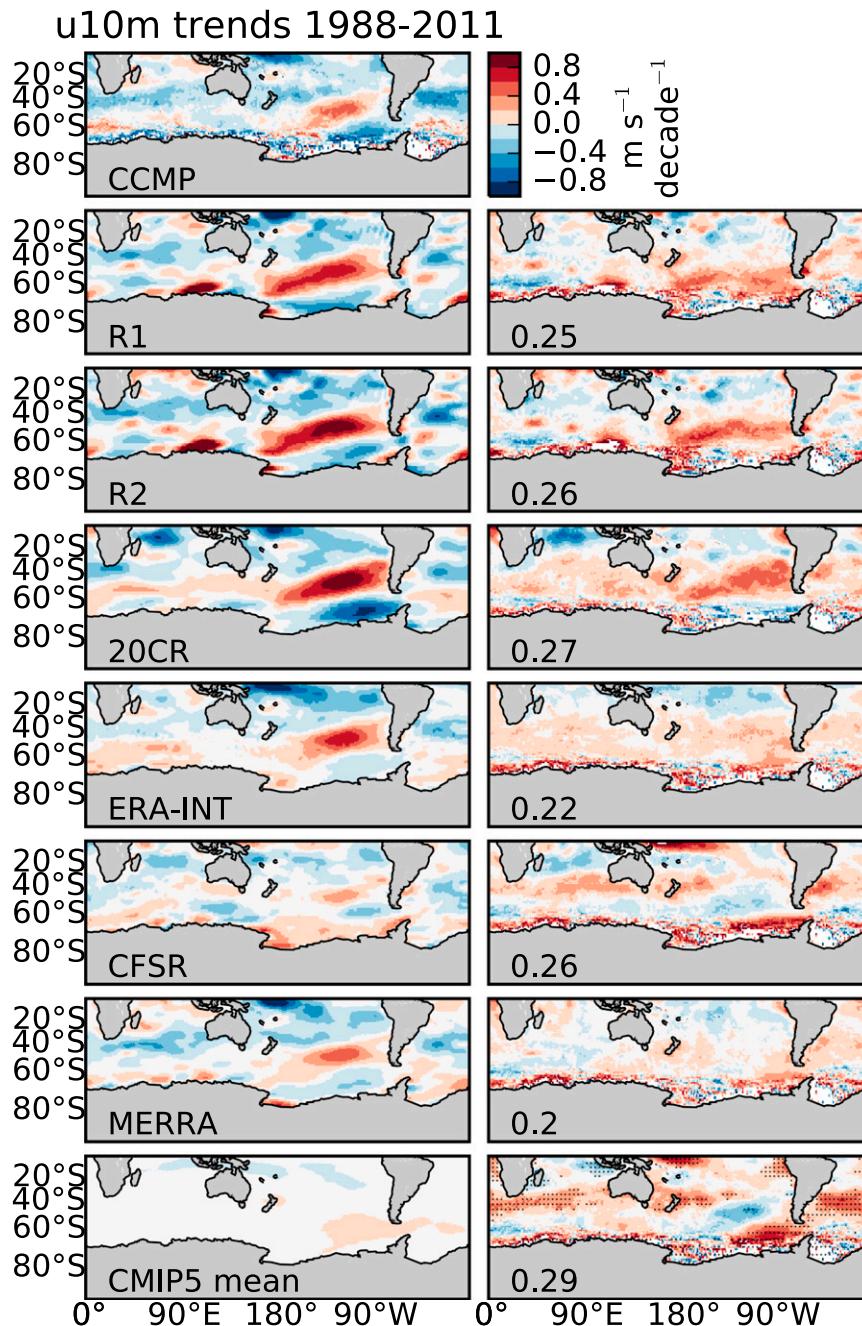


FIG. 12. (left) Trends in monthly mean u10m over 1988–2011 for (top)–(bottom) CCMP satellite winds, six reanalyses, and the CMIP5 mean, and (right) anomalies relative to the CCMP trends. The areas where the CCMP trend lies outside the 2.5th–97.5th percentile of trends in the individual CMIP5 simulations is shown by stippling in the bottom-right panel. Numbers in the bottom left of the right panels give the root-mean-square difference with CCMP ( $\text{m s}^{-1} \text{decade}^{-1}$ ).

Interestingly we note that the CMIP5 mean trend is more similar to the HadSLP2r observations than any of the six reanalyses, as seen by its smaller root-mean-square difference ( $51.65 \text{ Pa decade}^{-1}$ ).

Maps of u10m trends from the CCMP satellite-based wind product are compared with the reanalyses and CMIP5 models for the available period of 1988–2011 (Fig. 12). CCMP generally shows negative trends in the

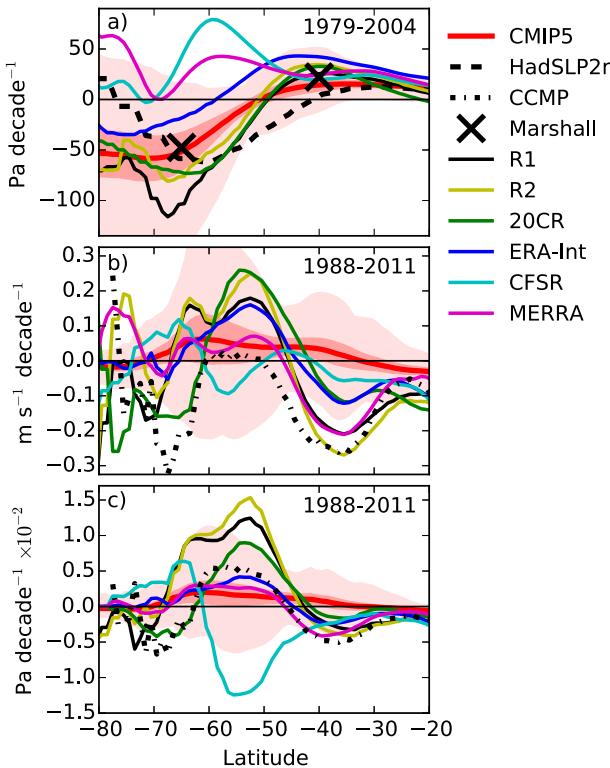


FIG. 13. Trends in zonal mean (a) sea level pressure over 1979–2004, (b) u10m over 1988–2011, and (c) the zonal component of the wind stress over 1988–2011. The solid red line shows the CMIP5 mean trend; the dark shading envelops the 95% confidence interval and the light shading envelops the 2.5th–97.5th percentile of trends in the individual CMIP5 simulations.

zonal winds (u10m) over the Southern Ocean during this period ( $-0.13 \text{ m s}^{-1} \text{ decade}^{-1}$  averaged south of  $35^\circ\text{S}$ ). Note that this contrasts with the surface wind speed trends in CCMP, which are generally positive ( $+0.27 \text{ m s}^{-1} \text{ decade}^{-1}$  averaged south of  $35^\circ\text{S}$ ) (Li et al. 2013; Wanninkhof et al. 2013). The CCMP u10m trend pattern is dominated by a large dipole-like feature in the South Pacific. All the reanalyses produce this pattern, but with varying degrees of magnitude. The trends are generally too large in R1, R2, and 20CR. MERRA is the best fit to the CCMP observations, followed by ERA-Interim, judged by their small root-mean-square difference with the CCMP trends. The CMIP5 models show only weak trends and do not reproduce the South Pacific dipole. This could reflect that fact that there is significant internal variability over the 23-yr period shown, or that the models are incapable of reproducing the correct response in the surface winds in this region, perhaps due to their inability to capture changes in the wavenumber-3 pattern as noted above (Marshall and Bracegirdle 2015).

To help compare the trends discussed above, zonal mean fields of the SLP and u10m trends were computed

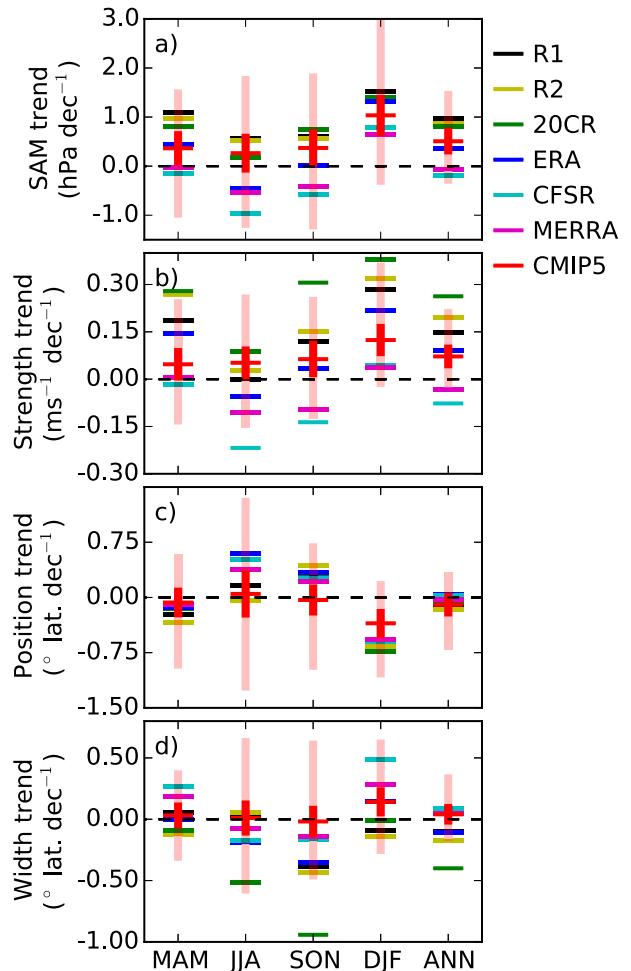


FIG. 14. Trends over 1979 to 2009 in (a) SAM index, (b) jet strength, (c) jet position, and (d) jet width by season in six reanalyses and CMIP5. The CMIP5 mean trend is given by the horizontal red line, the 95% confidence interval is given by the solid red vertical bar, and the 2.5th–97.5th percentile of trends in the individual CMIP5 simulations is given by the light red vertical bar. Since the spread in the 20CR ensemble is very small over this period it is not shown.

(Fig. 13). In the zonal means it is clear that the CMIP5 mean reproduces the available SLP observations very well (see red line and black crosses in Fig. 13a). In contrast, the positive SLP trends south of  $50^\circ\text{S}$  in CFSR and MERRA clearly stick out as spurious. The CCMP u10m trends interestingly show no positive trend near the peak of the jet ( $50^\circ\text{--}55^\circ\text{S}$ ; Fig. 13b). The MERRA u10m trends agree fairly well with the observations over this region, while R1, R2, and 20CR all seem to have trends that are too large. The CCMP observations and several reanalyses also show large negative u10m trends between about  $30^\circ\text{S}$  and  $50^\circ\text{S}$ . The CMIP5 model mean trend agrees well with the CCMP observations in the

region of the peak of the westerly jet near  $50^{\circ}$ – $60^{\circ}$ S. However the models do not simulate the negative trends on the equatorward flank of the jet near  $35^{\circ}$ S, where the CCMP trends fall outside the 2.5th–97.5th percentile of the CMIP5 trends.

Because  $u_{10m}$  winds depend on the formulation used to move winds to the reference height of 10 m (Kent et al. 2013), we also compare trends in surface zonal wind stress (Fig. 13c). Stress fields occur at a natural level (the surface), but themselves depend on the drag formulation employed. Nonetheless, in general the stress fields convey the same picture as  $u_{10m}$ , with R1, R2, and 20CR having larger than observed positive trends, with MERRA, ERA-Interim, and the CMIP5 mean being close to the CCMP values. Of note are the large negative trends evident in CFSR, consistent with a previous report (Swart et al. 2014).

Clearly, the best reanalysis product depends on the time period and variable of interest. One notable finding is that the CMIP5 models do not seem to underestimate the jet strengthening trend relative to the available observations, but R1, R2, and 20CR seem to overestimate the surface speed trends. In light of this it appears that the findings of section 2 that the models significantly underestimate the jet strength trends relative to 20CR should likely be interpreted as due to spuriously large trends in 20CR, not as a shortcoming in the models (although both could be in error). This indicates that a high degree of caution is required in using reanalysis products to validate simulated trends. Indeed, previous studies have also found a large spread between reanalysis products in the climatologies and trends of surface winds in the Southern Ocean (Kent et al. 2013; Li et al. 2013). In the final section, we reevaluate trends by season across all available products to search for robust features of change.

### c. Linear trends by season over 1979–2009

Here we consider trends over the 30-yr period between 1979 and 2009 (Fig. 14). This period has the advantage of being well observed, since it is within the satellite era. There are also six reanalysis products available for comparison, and the interproduct spread allows a determination of the observational uncertainty. The shorter 30-yr duration increases the ratio of noise in the trends due to internal variability, and reduces the statistical power relative to the 60-yr period (1951–2011) used previously. This is illustrated, for example, by the fact the 2.5th–97.5th percentile spread in DJF SAM trends across the CMIP5 ensemble increased from 1 hPa over 1951–2011 to over 3 hPa over 1979–2009.

During DJF, all six reanalysis products and the CMIP5 model mean show a significant positive trend in

the SAM. However, the SAM trends for the CMIP5 mean are smaller and not significant during the other seasons, and there is a large spread among the six reanalyses, which even differ their signs.

Similarly, the CMIP5 mean trend in jet strength is largest and statistically significant during DJF. All six reanalyses also show a positive trend during DJF, but the spread in magnitudes is large. Trends are smaller and more ambiguous during other seasons. Notably, in the annual mean, while the CMIP5 models show a significant positive trend on average, two reanalyses show negative trends, and the remaining four reanalyses have a factor of 3 spread in the magnitude of their trends.

Jet position trends show an important seasonality. The CMIP5 mean and all six reanalyses agree that the jet shifted poleward during DJF. However, during all the other seasons, and in the annual mean, the CMIP5 models do not show a significant trend in position. Indeed, all six reanalyses show a near a zero trend in annual mean jet position during this period. The annual mean trend is near zero in the reanalyses because the poleward trend during DJF is balanced by opposing equatorward trends in jet position during JJA and SON.

Jet width trends are not significant during any season for the CMIP5 mean. All six reanalyses do show negative trends (i.e., jet narrowing) during SON, but the spread in magnitude is large, and in the annual mean the reanalyses width trends are spread about zero.

The large spread among the reanalysis trends indicates the large degree of uncertainty in recently observed changes in the SH circulation. Similarly, the simulated changes have a large spread and are less certain than over the longer 60-yr period. Yet, despite the overall uncertainty, robust changes are clear during DJF, which is expected given the combination of ozone and GHG forcing (Son et al. 2010).

## 7. Discussion and conclusions

Over 1951–2011 the DJF trends in the 20CR ensemble mean SAM index and CMIP5 multimodel mean are nearly identical, yet over this same period the trend in the strength of the westerly jet in 20CR is much larger than the trends seen in the CMIP5 models (Figs. 2a,b). Using a simple geostrophic model we explained that trends in the SAM index and jet strength are not directly interchangeable, because trends in jet position and width combine with changes in jet strength to influence the SAM (Fig. 3). For this reason, trends in the SAM should not be used as a direct proxy for changes in any single kinematic property of the jet.

The CMIP5 models had an annual mean trend in the SAM index and jet strength that was significantly

smaller than seen in 20CR over 1951–2011 (Fig. 2b). However, this is partly due to spuriously large trends in 20CR, rather than the CMIP5 models underestimating the true trend. Indeed, the 20CR and HadSLP2r SAM trends since 1951 were largely driven by large negative trends in SLP in the South Pacific, a data-sparse region with a large uncertainty (Fig. 8; Allan and Ansell 2006).

Using sea level pressure data coincident with the 12 station locations used by Marshall (2003), we showed that the CMIP5 mean SLP trends at 40° and 65°S and the corresponding SAM index are consistent with the direct observations (Fig. 10). Surprisingly, the spatial pattern of CMIP5 model mean SLP trends was a better fit to HadSLP2r observed trends than any of six reanalysis products over the period 1979–2004 (Fig. 11). Similarly, in the zonal mean the CMIP5 trends in jet strength since 1988 were generally consistent with the CCMP satellite-based wind product near the core of the jet, although the models did not reproduce the spatial pattern of changes (Figs. 12 and 13). 20CR, R1, and R2 overestimated recent strengthening of the jet near its peak, relative to CCMP.

The best performing reanalysis product depends on the variable (SLP or u10m) and time period of choice, but in general 20CR best reproduced observed SLP trends while MERRA best reproduced surface wind trends relative to observations, and ERA-Interim performed best for surface winds and SLP combined. However, all the six reanalysis products experienced some spurious trends. The temporal continuity of reanalyses is inherently hampered by the evolving observational network that underlies these products. The resulting long-term trends in Southern Hemisphere sea level pressure and winds are unreliable, and as such reanalyses are likely inappropriate tools for validating these particular aspects of climate model simulations.

Many studies have used reanalysis-based forcing, and particularly R1, for forcing ocean-only models to investigate the role of Southern Ocean wind changes on ocean circulation (e.g., Biastoch et al. 2009; Screen et al. 2009) and the carbon cycle (e.g., Le Quéré et al. 2007; Lovenduski et al. 2008). The widely used surface forcing from the Co-ordinated Ocean–Ice Reference Experiments (CORE) Phases 1 and 2 (Danabasoglu et al. 2014; Large and Yeager 2009; Griffies et al. 2009) is itself primarily based on R1. However, as we have shown here, R1 has particularly large and spurious trends over the Southern Ocean, which might in turn bias studies using R1-derived products as surface forcing. Indeed, the impacts of atmospheric circulation changes on the Southern Ocean circulation and carbon cycle are highly sensitive to the choice of surface forcing (Swart et al. 2014), and the significant uncertainties associated with this forcing require further attention.

*Acknowledgments.* We thank Michael Sigmond and Slava Kharin for helpful comments on an earlier draft. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling 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 provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Support for the Twentieth Century Reanalysis Project dataset is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office. NCEP reanalysis data (R1 and R2) was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, from their website (Table 1). CFSR and CCMP data were made available from the Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. GJM was supported by the UK Natural Environment Research Council through the British Antarctic Survey programme Polar Science for Planet Earth.

## REFERENCES

- Allan, R., and T. Ansell, 2006: A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *J. Climate*, **19**, 5816–5842, doi:10.1175/JCLI3937.1.
- Atlas, R., R. N. Hoffman, J. Ardizzone, S. Leidner, J. Jusem, D. Smith, and D. Gombos, 2011: A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. *Bull. Amer. Meteor. Soc.*, **92**, 157–174, doi:10.1175/2010BAMS2946.1.
- Biastoch, A., C. W. Boning, F. U. Schwarzkopf, and J. R. E. Lutjeharms, 2009: Increase in Agulhas leakage due to poleward shift of Southern Hemisphere westerlies. *Nature*, **462**, 495–498, doi:10.1038/nature08519.
- Bracegirdle, T. J., E. Shuckburgh, J.-B. Sallee, Z. Wang, A. J. S. Meijers, N. Bruneau, T. Phillips, and L. J. Wilcox, 2013: Assessment of surface winds over the Atlantic, Indian, and Pacific Ocean sectors of the Southern Ocean in CMIP5 models: Historical bias, forcing response, and state dependence. *J. Geophys. Res. Atmos.*, **118**, 547–562, doi:10.1002/jgrd.50153.
- Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, doi:10.1002/qj.776.
- Danabasoglu, G., and Coauthors, 2014: North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part I: Mean states. *Ocean Modell.*, **73**, 76–107, doi:10.1016/j.ocemod.2013.10.005.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation

- system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:[10.1002/qj.828](https://doi.org/10.1002/qj.828).
- Droettboom, M., and Coauthors, 2015: matplotlib version 1.4.3, doi:[10.5281/zenodo.15423](https://doi.org/10.5281/zenodo.15423).
- Frölicher, T., J. Sarmiento, D. Paynter, J. Dunne, J. Krasting, and M. Winton, 2015: Dominance of the Southern Ocean in anthropogenic carbon and heat uptake in CMIP5 models. *J. Climate*, **28**, 862–886, doi:[10.1175/JCLI-D-14-00117.1](https://doi.org/10.1175/JCLI-D-14-00117.1).
- Fyfe, J. C., G. J. Boer, and G. M. Flato, 1999: The Arctic and Antarctic oscillations and their projected changes under global warming. *Geophys. Res. Lett.*, **26**, 1601–1604, doi:[10.1029/1999GL900317](https://doi.org/10.1029/1999GL900317).
- , N. P. Gillett, and F. W. Zwiers, 2013: Overestimated global warming in the past 20 years. *Nat. Climate Change*, **3**, 767–769, doi:[10.1038/nclimate1972](https://doi.org/10.1038/nclimate1972).
- Gillett, N. P., and J. C. Fyfe, 2013: Annular mode changes in the CMIP5 simulations. *Geophys. Res. Lett.*, **40**, 1189–1193, doi:[10.1002/grl.50249](https://doi.org/10.1002/grl.50249).
- , —, and D. E. Parker, 2013: Attribution of observed sea level pressure trends to greenhouse gas, aerosol, and ozone changes. *Geophys. Res. Lett.*, **40**, 2302–2306, doi:[10.1002/grl.50500](https://doi.org/10.1002/grl.50500).
- Gong, D., and S. Wang, 1999: Definition of Antarctic Oscillation index. *Geophys. Res. Lett.*, **26**, 459–462, doi:[10.1029/1999GL900003](https://doi.org/10.1029/1999GL900003).
- Griffies, S. M., and Coauthors, 2009: Coordinated Ocean-Ice Reference Experiments (COREs). *Ocean Modell.*, **26**, 1–46, doi:[10.1016/j.ocemod.2008.08.007](https://doi.org/10.1016/j.ocemod.2008.08.007).
- Hartmann, D., and F. Lo, 1998: Wave-driven zonal flow vacillation in the Southern Hemisphere. *J. Atmos. Sci.*, **55**, 1303–1315, doi:[10.1175/1520-0469\(1998\)055<1303:WDZFVI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<1303:WDZFVI>2.0.CO;2).
- Ho, M., A. S. Kiem, and D. C. Verdon-Kidd, 2012: The southern annular mode: A comparison of indices. *Hydrol. Earth Syst. Sci.*, **16**, 967–982, doi:[10.5194/hess-16-967-2012](https://doi.org/10.5194/hess-16-967-2012).
- Hunter, J. D., 2007: Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.*, **9**, 90–95, doi:[10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- Jones, J. M., R. L. Fogt, M. Widmann, G. Marshall, P. Jones, and M. Visbeck, 2009: Historical SAM variability. Part I: Century-length seasonal reconstructions. *J. Climate*, **22**, 5319–5345, doi:[10.1175/2009JCLI2785.1](https://doi.org/10.1175/2009JCLI2785.1).
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:[10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643, doi:[10.1175/BAMS-83-11-1631](https://doi.org/10.1175/BAMS-83-11-1631).
- Kent, E. C., S. Fangohr, and D. I. Berry, 2013: A comparative assessment of monthly mean wind speed products over the global ocean. *Int. J. Climatol.*, **33**, 2520–2541, doi:[10.1002/joc.3606](https://doi.org/10.1002/joc.3606).
- Kidston, J., and E. Gerber, 2010: Intermodel variability of the poleward shift of the austral jet stream in the CMIP3 integrations linked to biases in 20th century climatology. *Geophys. Res. Lett.*, **37**, L09708, doi:[10.1029/2010GL042873](https://doi.org/10.1029/2010GL042873).
- Kostov, Y., K. C. Armour, and J. Marshall, 2014: Impact of the Atlantic meridional overturning circulation on ocean heat storage and transient climate change. *Geophys. Res. Lett.*, **41**, 2108–2116, doi:[10.1002/2013GL058998](https://doi.org/10.1002/2013GL058998).
- Large, W., and S. Yeager, 2009: The global climatology of an interannually varying air–sea flux data set. *Climate Dyn.*, **33**, 341–364, doi:[10.1007/s00382-008-0441-3](https://doi.org/10.1007/s00382-008-0441-3).
- Le Quéré, C., and Coauthors, 2007: Saturation of the Southern Ocean CO<sub>2</sub> sink due to recent climate change. *Science*, **316**, 1735–1738, doi:[10.1126/science.1136188](https://doi.org/10.1126/science.1136188).
- Li, M., J. Liu, Z. Wang, H. Wang, Z. Zhang, L. Zhang, and Q. Yang, 2013: Assessment of sea surface wind from NWP reanalyses and satellites in the Southern Ocean. *J. Atmos. Oceanic Technol.*, **30**, 1842–1853, doi:[10.1175/JTECH-D-12-00240.1](https://doi.org/10.1175/JTECH-D-12-00240.1).
- Lovenduski, N. S., N. Gruber, and S. C. Doney, 2008: Toward a mechanistic understanding of the decadal trends in the Southern Ocean carbon sink. *Global Biogeochem. Cycles*, **22**, GB3016, doi:[10.1029/2007GB003139](https://doi.org/10.1029/2007GB003139).
- Marinov, I., A. Gnanadesikan, J. Toggweiler, and J. Sarmiento, 2006: The Southern Ocean biogeochemical divide. *Nature*, **441**, 964–967, doi:[10.1038/nature04883](https://doi.org/10.1038/nature04883).
- Marshall, G. J., 2003: Trends in the southern annular mode from observations and reanalyses. *J. Climate*, **16**, 4134–4143, doi:[10.1175/1520-0442\(2003\)016<4134:TITSAM>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<4134:TITSAM>2.0.CO;2).
- , and T. Bracegirdle, 2015: An examination of the relationship between the southern annular mode and Antarctic surface air temperatures in the CMIP5 historical runs. *Climate Dyn.*, **45**, 1513–1535, doi:[10.1007/s00382-014-2406-z](https://doi.org/10.1007/s00382-014-2406-z).
- Marshall, J., and K. Speer, 2012: Closure of the meridional overturning circulation through Southern Ocean upwelling. *Nat. Geosci.*, **5**, 171–180, doi:[10.1038/ngeo1391](https://doi.org/10.1038/ngeo1391).
- Monahan, A. H., and J. C. Fyfe, 2006: On the nature of zonal jet EOFs. *J. Climate*, **19**, 6409–6424, doi:[10.1175/JCLI3960.1](https://doi.org/10.1175/JCLI3960.1).
- , and —, 2008: On annular modes and zonal jets. *J. Climate*, **21**, 1963–1978, doi:[10.1175/2007JCLI1841.1](https://doi.org/10.1175/2007JCLI1841.1).
- Pérez, F., and B. E. Granger, 2007: IPython: A system for interactive scientific computing. *Comput. Sci. Eng.*, **9**, 21–29, doi:[10.1109/MCSE.2007.53](https://doi.org/10.1109/MCSE.2007.53).
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA’s Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi:[10.1175/JCLI-D-11-00015.1](https://doi.org/10.1175/JCLI-D-11-00015.1).
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:[10.1175/2010BAMS3001.1](https://doi.org/10.1175/2010BAMS3001.1).
- Santer, B. D., and Coauthors, 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *Int. J. Climatol.*, **28**, 1703–1722, doi:[10.1002/joc.1756](https://doi.org/10.1002/joc.1756).
- Sarmiento, J., N. Gruber, M. Brzezinski, and J. Dunne, 2004: High-latitude controls of thermocline nutrients and low-latitude biological productivity. *Nature*, **427**, 56–60, doi:[10.1038/nature02127](https://doi.org/10.1038/nature02127).
- Screen, J., N. Gillett, D. Stevens, G. Marshall, and H. Roscoe, 2009: The role of eddies in the Southern Ocean temperature response to the southern annular mode. *J. Climate*, **22**, 806–818, doi:[10.1175/2008JCLI2416.1](https://doi.org/10.1175/2008JCLI2416.1).
- Son, S.-W., and Coauthors, 2010: Impact of stratospheric ozone on Southern Hemisphere circulation change: A multimodel assessment. *J. Geophys. Res.*, **115**, D00M07, doi:[10.1029/2010JD014271](https://doi.org/10.1029/2010JD014271).
- Swart, N., 2015: Analysis code for “Comparing trends in the Southern Annular Mode and surface westerly jet,” doi:[10.5281/zenodo.30103](https://doi.org/10.5281/zenodo.30103).
- , and J. C. Fyfe, 2012: Observed and simulated changes in the Southern Hemisphere surface westerly wind-stress. *Geophys. Res. Lett.*, **39**, L16711, doi:[10.1029/2012GL052810](https://doi.org/10.1029/2012GL052810).
- , —, O. A. Saenko, and M. Eby, 2014: Wind-driven changes in the ocean carbon sink. *Biogeosciences*, **11**, 6107–6117, doi:[10.5194/bg-11-6107-2014](https://doi.org/10.5194/bg-11-6107-2014).

- , —, E. Hawkins, J. E. Kay, and A. Jahn, 2015: Influence of internal variability on Arctic sea-ice trends. *Nat. Climate Change*, **5**, 86–89, doi:10.1038/nclimate2483.
- Thomas, J. L., D. W. Waugh, and A. Gnanadesikan, 2015: Southern Hemisphere extratropical circulation: Recent trends and natural variability. *Geophys. Res. Lett.*, **42**, 5508–5515, doi:10.1002/2015GL064521.
- Thompson, D. W. J., and J. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, **13**, 1000–1016, doi:10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2.
- , and S. Solomon, 2002: Interpretation of recent Southern Hemisphere climate change. *Science*, **296**, 895–899, doi:10.1126/science.1069270.
- , —, P. Kushner, M. England, K. Grise, and D. Karoly, 2011: Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nat. Geosci.*, **4**, 741–749, doi:10.1038/ngeo1296.
- Toggweiler, J. R., and B. Samuels, 1995: Effect of Drake Passage on the global thermohaline circulation. *Deep-Sea Res. I*, **42**, 477–500, doi:10.1016/0967-0637(95)00012-U.
- , and J. Russell, 2008: Ocean circulation in a warming climate. *Nature*, **451**, 286–288, doi:10.1038/nature06590.
- Visbeck, M., 2009: A station-based southern annular mode index from 1884 to 2005. *J. Climate*, **22**, 940–950, doi:10.1175/2008JCLI2260.1.
- von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.
- Wang, X., Y. Feng, G. P. Compo, V. R. Swail, F. W. Zwiers, R. J. Allan, and P. D. Sardeshmukh, 2013: Trends and low frequency variability of extra-tropical cyclone activity in the ensemble of twentieth century reanalysis. *Climate Dyn.*, **40**, 2775–2800, doi:10.1007/s00382-012-1450-9.
- Wanninkhof, R., and Coauthors, 2013: Global ocean carbon uptake: Magnitude, variability and trends. *Biogeosciences*, **10**, 1983–2000, doi:10.5194/bg-10-1983-2013.