

## Historical SAM Variability. Part II: Twentieth-Century Variability and Trends from Reconstructions, Observations, and the IPCC AR4 Models\*

RYAN L. FOGT,<sup>+,#</sup> JUDITH PERLWITZ,<sup>@</sup> ANDREW J. MONAGHAN,<sup>+,&</sup> & DAVID H. BROMWICH,<sup>+,\*\*</sup>  
 JULIE M. JONES,<sup>++</sup> AND GARETH J. MARSHALL<sup>##</sup>

<sup>+</sup> *Polar Meteorology Group, Byrd Polar Research Center, The Ohio State University, Columbus, Ohio*

<sup>@</sup> *NOAA/Earth System Research Laboratory, Physical Sciences Division, Boulder, Colorado*

<sup>\*\*</sup> *Atmospheric Sciences Program, Department of Geography, The Ohio State University, Columbus, Ohio*

<sup>++</sup> *Department of Geography, University of Sheffield, Sheffield, United Kingdom*

<sup>##</sup> *British Antarctic Survey, Cambridge, United Kingdom*

(Manuscript received 12 August 2008, in final form 10 June 2009)

### ABSTRACT

This second paper examines the Southern Hemisphere annular mode (SAM) variability from reconstructions, observed indices, and simulations from 17 Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) models from 1865 to 2005. Comparisons reveal the models do not fully simulate the duration of strong natural variability within the reconstructions during the 1930s and 1960s.

Seasonal indices are examined to understand the relative roles of forced and natural fluctuations. The models capture the recent (1957–2005) positive SAM trends in austral summer, which reconstructions indicate is the strongest trend during the last 150 yr; ozone depletion is the dominant mechanism driving these trends. In autumn, negative trends after 1930 in the reconstructions are stronger than the recent positive trend. Furthermore, model trends in autumn during 1957–2005 are the most different from observations. Both of these conditions suggest the recent autumn trend is most likely natural climate variability, with external forcing playing a secondary role. Many models also produce significant spring trends during this period not seen in observations. Although insignificant, these differences arise because of vastly different spatial structures in the Southern Hemisphere pressure trends. As the trend differences between models and observations in austral spring have been increasing over the last 30 yr, care must be exercised when examining the future SAM projections and their impacts in this season.

### 1. Introduction

In the companion paper, Jones et al. (2009, hereafter Part I) describe in detail various reconstructions of the Southern Hemisphere annular mode (SAM). Their study addresses the similarities between various reconstructions and discusses precise reasons for the differences among them and their known limitations. The full versions of the reconstructions extend back until at least

1905, and, despite different methodologies, the concatenated version of the Jones and Widmann (hereafter JWconcat) and Fogt reconstructions agree reasonably well with each other over the twentieth century and with observed indices starting in 1957. Given the reliability of these reconstructions as demonstrated in Part I, the historical SAM variability and trends throughout all of the twentieth century and the role of forced and natural variability in comparison with climate model simulations can be evaluated. These are the primary goals of this second paper. The analysis builds upon preliminary results in Fogt (2007) and does not use the Visbeck reconstructions included in Part I, as the variability and trends in that reconstruction are already assessed by Visbeck (2009).

Although recent observational studies have noted significant positive trends in the SAM during austral summer and autumn (Marshall 2007), the historical significance of these trends has remained largely unknown.

\* Byrd Polar Research Center Contribution Number 1381.

# Current affiliation: NOAA/Earth System Research Laboratory, Physical Sciences Division, Boulder, Colorado.

& Current affiliation: National Center for Atmospheric Research, Boulder, Colorado.

Corresponding author address: Ryan L. Fogt, NOAA/ESRL/PSD, 325 Broadway R/PSD, Boulder, CO 80305.  
 E-mail: ryan.fogt@noaa.gov

Notably, the summer trends in particular have been extensively associated with increasing greenhouse gas concentrations (Fyfe et al. 1999; Kushner et al. 2001; Stone et al. 2001; Cai et al. 2003; Marshall et al. 2004; Rauthe et al. 2004; Arblaster and Meehl 2006) and stratospheric ozone depletion over Antarctica (Sexton 2001; Thompson and Solomon 2002; Gillett and Thompson 2003; Shindell and Schmidt 2004; Miller et al. 2006; Cai and Cowan 2007; Perlwitz et al. 2008). Despite these strong recent trends, Jones and Widmann (2004) noted an “early peak” in their reconstructed December–January SAM index near 1960, thought to be due to internal climate processes. Other recent studies have found interactions between the summer SAM and tropical Pacific sea surface temperature variability (Zhou and Yu 2004; L’Heureux and Thompson 2006; Fogt and Bromwich 2006). Apart from the austral summer, relatively little work has been undertaken using seasonal multimodel comparisons of the SAM throughout the twentieth century.

Employing the reconstructions described in Part I, these recent trends can be placed in an historical context. By extending the observed indices by more than five decades, the relative roles of natural variability [historical spikes likely unrelated to these steady forcing mechanisms, as in the Jones and Widmann (2004) reconstruction] and the low-frequency anthropogenically amplified or forced variability from both ozone depletion and greenhouse gases most prominent in the latter half of the twentieth century (Solomon et al. 2005; Miller et al. 2006) can be assessed. Karoly (2003) states that untangling the separate contributions of forced versus internal processes in the SAM is “crucial” for understanding regional variations and future climate changes, and is thus of great importance.

To investigate the forced response, we make use of World Climate Research Programme’s (WCRP) Coupled Model Intercomparison Project phase 3 (CMIP3) multimodel dataset, which houses ensemble simulations conducted for scientific study in the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4; Solomon et al. 2007). The SAM indices from these ensemble simulations are compared to the reconstructions in a manner similar to Miller et al. (2006), but now for all seasons. Additionally, the reconstructions offer a better reflection of model performance and historical SAM variability than the First Hadley Centre Sea Level Pressure dataset (HadSLP1) used by Miller et al. (2006). With these reconstructions, this paper not only assesses the relative importance and the forcing behind the recent SAM trends in autumn and summer, but it also investigates the simulation of the SAM internal climate cycles by the models. This expands greatly

on previous studies (i.e., Arblaster and Meehl 2006; Miller et al. 2006; Cai and Cowan 2007) that did not assess the simulation of internal SAM variability due to the unavailability of long-term reliable SAM indices.

The paper is organized as follows: section 2 describes in more detail the various global climate models and the method used to obtain the SAM indices from the ensemble simulations. Section 3 examines the temporal variability in the SAM reconstructions and models and assesses the simulation of internal (unforced) SAM variability. Section 4 investigates the SAM trends, with special emphasis on attribution, during 1957–2005 when all indices overlap and ozone forcing becomes important. To put these recent trends in an historical perspective, section 5 examines the temporal evolution of the trends in the reconstructions and models. The analysis thus not only shows the overall picture of climate variability and change over the Southern Hemisphere—given the SAM’s strong influence on temperature, pressure, and precipitation there (Gillett et al. 2006; Karpechko et al. 2009)—but also assesses the models’ capabilities to predict this change. A summary and conclusions are offered in section 6.

## 2. Data and methods

### a. Global climate model characteristics

Table 1 describes the 17 fully coupled (with ocean, ice, and atmosphere components) global climate models (GCMs) included in this study, as well as the acronyms for each model. The mean sea level pressure (MSLP) field from the climate of the twentieth century (20C3m) experiment is the primary dataset employed, including all available simulations from models that have at least two ensemble members. All the simulations studied here begin prior to 1900 (although comparisons are conducted after 1865 when the earliest reconstructions begin) and run to 1999, 2000, or 2005. To ensure the magnitude of the recent trends are not influenced strongly by differing end dates of the 20C3m runs, the ensemble simulations are updated to 2005 using the Special Report on Emissions Scenarios (SRES) A1B runs, as in Miller et al. (2006) and Gillett et al. (2005). This scenario was chosen as it provides the greatest number of continued ensemble runs; however, because only the first five or so years are used, the choice is arbitrary as they differ little during this period (Gillett et al. 2005).

All of the models studied contain forcing from carbon dioxide and many other greenhouse gases, while 11 of the 17 GCMs employ time-variable ozone forcing (Table 1). Although each of the models is unique, there are many similarities between certain models worth

TABLE 1. Description of the IPCC AR4 GCMs used in the study. Members = number of ensemble runs used for 20C3m and A1B scenarios; ozone = indication of whether model contains time-variable ozone forcing; horizontal resolution = approximate output resolution of models; model top = height of atmospheric model top.

Model	Country	20C3m starting year	20C3m members	SRES A1B members	Ozone	Horizontal resolution (lon × lat, °)	Vertical levels	Model top (hPa)
CCCMA CGCM3.1	Canada	1865	5	5	No	2.80 × 2.80	31	1
Commonwealth Scientific and Industrial Research Organisation Mark version 3.0 (CSIRO Mk3.0)	Australia	1871	3	1	Yes	1.88 × 1.88	18	4.5
Institute of Atmospheric Physics (IAP) Flexible Global Ocean–Atmosphere–Land System Model (FGOALS)	China	1865	3	3	No	2.81 × 2.81	26	2.2
IPSL CM4	France	1865	2	1	No	3.75 × 2.50	19	4
Model for Interdisciplinary Research on Climate, medium-resolution version [MIROC(medres)]	Japan	1865	3	3	Yes	2.81 × 2.81	20	30 (km)
Meteorological Institute of the University of Bonn, ECHAM and the global Hamburg Ocean Primitive Equation (ECHO-G) Model (MIUBECHOG)	Germany/South Korea	1865	5	3	No	3.90 × 3.90	19	10
Max Planck Institute (MPI) ECHAM5	Germany	1865	4	4	Yes	1.88 × 1.88	31	10
MRI CGCM2	Japan	1865	5	5	No	2.81 × 2.81	30	0.4
NASA GISS Model E-H	United States	1880	5	5	Yes	5.00 × 4.00	20	0.1
NASA GISS Model E-R	United States	1880	9	5	Yes	5.00 × 4.00	20	0.1
NASA GISS Russell AOGCM	United States	1865	2	2	No	4.00 × 3.00	12	10
NCAR CCSM3.0	United States	1870	8	7	Yes	1.41 × 1.41	26	2.2
NCAR PCM1	United States	1890	4	4	Yes	2.81 × 2.81	18	2.2
NOAA GFDL CM2.0	United States	1865	3	1	Yes	2.50 × 2.00	24	3
NOAA GFDL CM2.1	United States	1865	3	1	Yes	2.50 × 2.00	24	3
UKMO third climate configuration of the Met Office Unified Model (HadCM3)	United Kingdom	1865	2	1	Yes	3.75 × 2.50	19	5
UKMO Hadley Centre Global Environmental Model version 1 (HadGEM1)	United Kingdom	1865	2	1	Yes	1.875 × 1.25	38	39.2 (km)

noting: the National Oceanic and Atmospheric Administration (NOAA) Geophysical Fluid Dynamics Laboratory (GFDL) Climate Model version 2.0 (CM2.0) and CM2.1 differ only in their dynamical cores, while the National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies (GISS) Model E-H and Model E-R differ in the ocean model employed (i.e., both have the same atmospheric general circulation model). In addition, the National Center for Atmospheric Research (NCAR) Community Climate System Model, version 3 (CCSM3.0) and Parallel Climate Model version 1.0 (PCM1.0), as well as the two Met Office (UKMO) models, share many physical parameterizations. Miller et al. (2006) and chapters 8–10 of the 2007 IPCC AR4 (Randall et al. 2007; Hegerl et al. 2007;

Meehl et al. 2007) provide additional information on the GCMs studied here.

### b. SAM definition

Taking a slightly different approach from Part I, here we use the observationally derived SAM index of Marshall (2003; hereafter “Marshall”) as our reference SAM index. This index starts in 1957, and because it is not dependent upon atmospheric reanalyses, which have been shown to be problematic prior to 1979 in the high southern latitudes (Bromwich et al. 2007; Bromwich and Fogt 2004), it is considered the best estimate of the SAM variability prior to 1979. The index also has the most reliable trends from 1957 onward (Marshall 2003, 2007). As demonstrated in Part I, it correlates very well ( $r > 0.85$

for periods post-1979) with other monthly SAM indices using the reanalyses based on the leading empirical orthogonal function at various pressure levels throughout the troposphere, so similar results are expected using other definitions (Arblaster and Meehl 2006; Miller et al. 2006).

As another estimate of observed SAM variability, two SAM indices from the Hadley Centre gridded MSLP dataset (HadSLP2; Allan and Ansell 2006) are also used from 1865 to 2005. Although this dataset unavoidably suffers from sparse observations prior to 1957, its variability (and thus, trends) from 1957 to 1979 is more reliable than the reanalysis datasets, and it is of similar quality after 1979 (Jones and Lister 2007). As in Part I, a Gong and Wang (1999)-based definition (HadSLP2-GW99) and one from the leading principal component (PC) of area-weighted MSLP from 20° to 90°S (HadSLP2-PC1) are used throughout.

The SAM index for each model ensemble simulation is based on the leading principal component of area-weighted MSLP from 20° to 90°S. Before calculating the empirical orthogonal functions (EOFs) for each simulation, the data were interpolated to a 5° × 5° latitude-longitude grid. The EOFs were calculated for the 1950–99 period, and the PCs were calculated by projecting the EOFs onto the original data through time. A Gong and Wang (1999) index was also calculated for each simulation, and the results are very similar; thus our conclusions are insensitive to the precise model SAM definition.

Each seasonal index is rescaled as in Part I, using the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) means and standard deviations over 1979–2001 from their respective index [i.e., the Gong and Wang (1999) index for the Fogt reconstruction, Marshall, HadSLP2-GW99, and the leading ERA-40 MSLP PC for the JWconcat reconstruction and model indices]. The HadSLP-PC1 is standardized based on its own mean and variance, since it has much lower variability than the ERA-40 reanalysis (Allan and Ansell 2006). This effect only changes the magnitude of the trends (making them similar to the other indices), but has little impact otherwise. All seasons listed are in reference to the Southern Hemisphere.

### 3. Historical SAM variability from reconstructions and GCMs

Figure 1 displays the 9-yr smoothed (as in Part I, with a Hamming filter) SAM indices along with the interannual model grand ensemble mean. The light and dark gray shaded regions in Fig. 1 represent the 95/5 and 75/25

percentiles, based on the spread of the individual ensemble simulations, respectively. It is quickly evident in Fig. 1 that both HadSLP2 SAM indices are markedly lower than the other indices prior to 1957. Although the HadSLP2 index realigns before 1900 with the reconstructions in the summer [December–February (DJF)] and autumn [March–May (MAM)], it is more than 1 standard deviation lower in the 1920–50 period in all seasons. The negative early HadSLP2 index results from MSLP increases of 2–4 hPa, greatest in winter [June–August (JJA)] and spring [September–November (SON)], in the high southern latitudes near the year 1957 (not shown). This year corresponds to the starting date of most Antarctic stations. Therefore, the shift at this time is thought to arise from insufficient observations constraining the early high southern latitude HadSLP2 and is likely erroneous. Because of this uncertainty, the HadSLP2 data is only examined from 1957 to 2005. Nonetheless, much of the decadal variability throughout the whole period is similar in HadSLP2 and the reconstructions.

An important question to ask when evaluating model performance is how well the models simulate real-world internal climate variability. In terms of the SAM reconstructions, an important unforced signal is the pronounced early SAM peaks (i.e., early 1960s in DJF, 1890s and 1930s in MAM, 1930s and 1960s in SON), since these peaks occur prior to ozone depletion and when the external forcing from greenhouse gases is much weaker than in the latter part of the twentieth century. Part I (their Fig. 8) demonstrated that the 1930s peaks were a true hemispheric SAM signature and not just a regional pressure response captured by the stations used in the reconstructions; the 1960s peaks also show a strong SAM signature but with a few regional asymmetries in the midlatitudes. To investigate whether these strong SAM peaks occur in the model's internal variability, we compare the interannual (unsmoothed) SAM indices from each model's preindustrial control simulation with the reconstructions prior to 1980. Each index (including the reconstructions) is standardized and the model indices are further broken into lengths equivalent to the pre-1980 reconstructions (i.e., 115 yr in DJF and MAM, and 76 yr in SON and JJA), making the distributions approximately normal and easily comparable. The total number of events where an index remains at or above  $\pm 1$  standard deviation for a given duration (ranging from a single year up to 6 yr) are displayed in Fig. 2. Note that the vertical axis in Fig. 2 is on a logarithmic scale and is in terms of total counts plus 1, so that a count of zero is represented as "1." The shading represents the 5/95 percentiles from the individual control simulations.

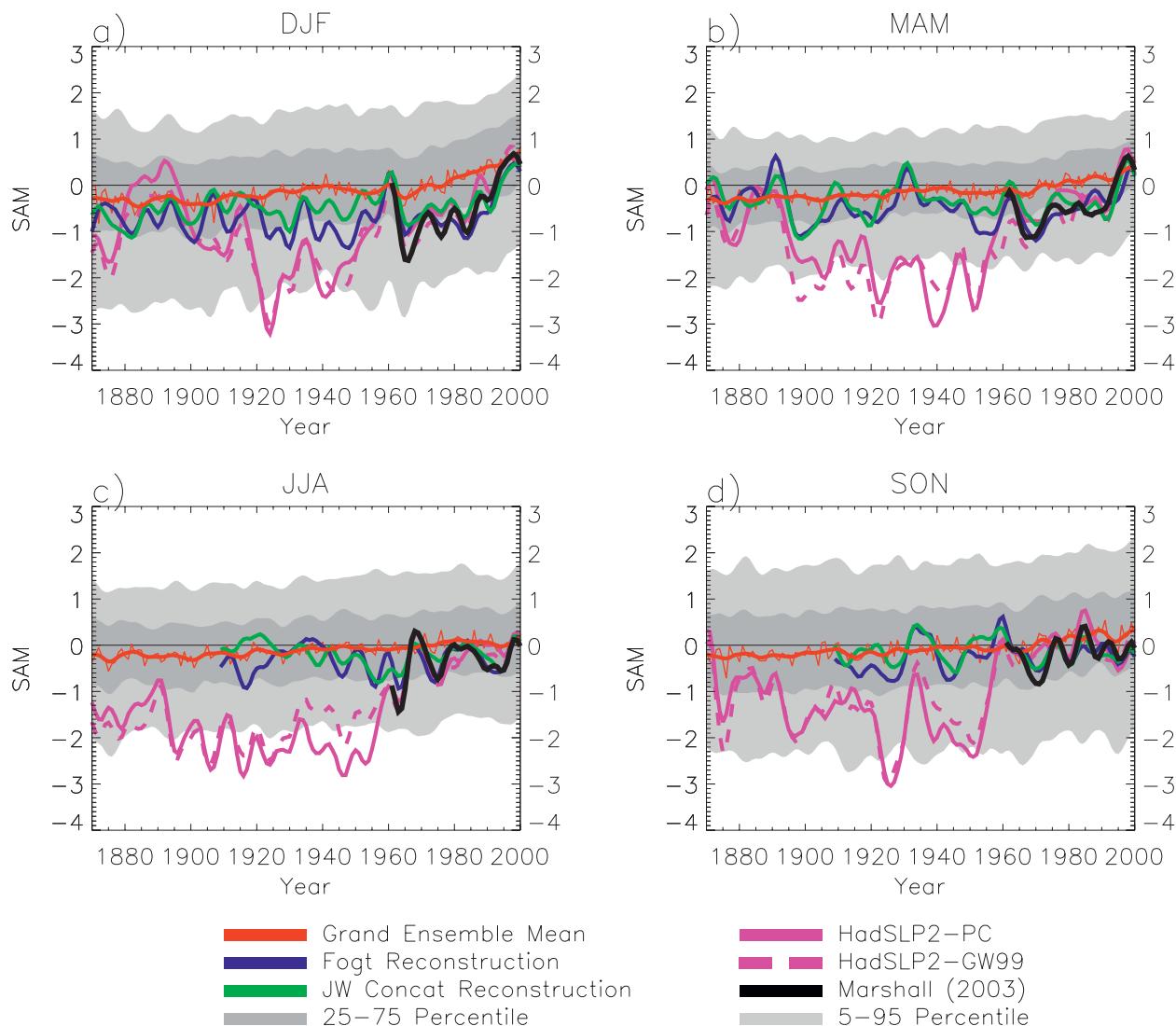


FIG. 1. Nine-yr smoothed (using a Hamming filter, as in Part I) SAM indices, along with the interannual grand ensemble mean (thin red line). The light and dark gray shaded regions represent the smoothed 95/5 and 75/25 percentiles, respectively, based on the spread of the individual ensemble runs. (a) DJF, (b) MAM, (c) JJA, and (d) SON.

In DJF and MAM, both reconstructions indicate multiple unsmoothed peaks lasting for 2–3 yr, clearly visible in the smoothed versions in Figs. 1a,b. In SON, the JWconcat reconstruction displays two peaks lasting exactly 3 yr each (positive in the 1930s, negative in the mid-1940s), while the Fogt reconstruction displays only a 4-yr period when the index was continuously above 1 standard deviation (in the early 1960s). The differences arise as 1 yr during each peak in the interannual reconstructions may fail to reach the threshold. However, because the smoothed versions in Fig. 1 agree very well, this value must still maintain a strong anomaly of the same sign as its neighbors. Therefore, the reconstructions are consistent despite slightly different magnitudes of their interannual values.

From Fig. 2, it is clear that the internal variability within the models simulates many fewer peaks lasting at least 2 yr (Fig. 2) in all seasons but JJA, as the total counts in the reconstructions lie outside 95% of the control simulations. It can therefore be said with high confidence that the models do not simulate such pronounced, multiyear SAM peaks in their internal variability. These tests, however, do not assess the general persistence of the SAM indices, since Fig. 2 only examines the duration of any isolated peak outside 1 standard deviation from its mean. To assess the general persistence of every peak (regardless of their magnitude), the ensemble mean lag-1 autocorrelations from the control simulations and their 5/95 percentiles are

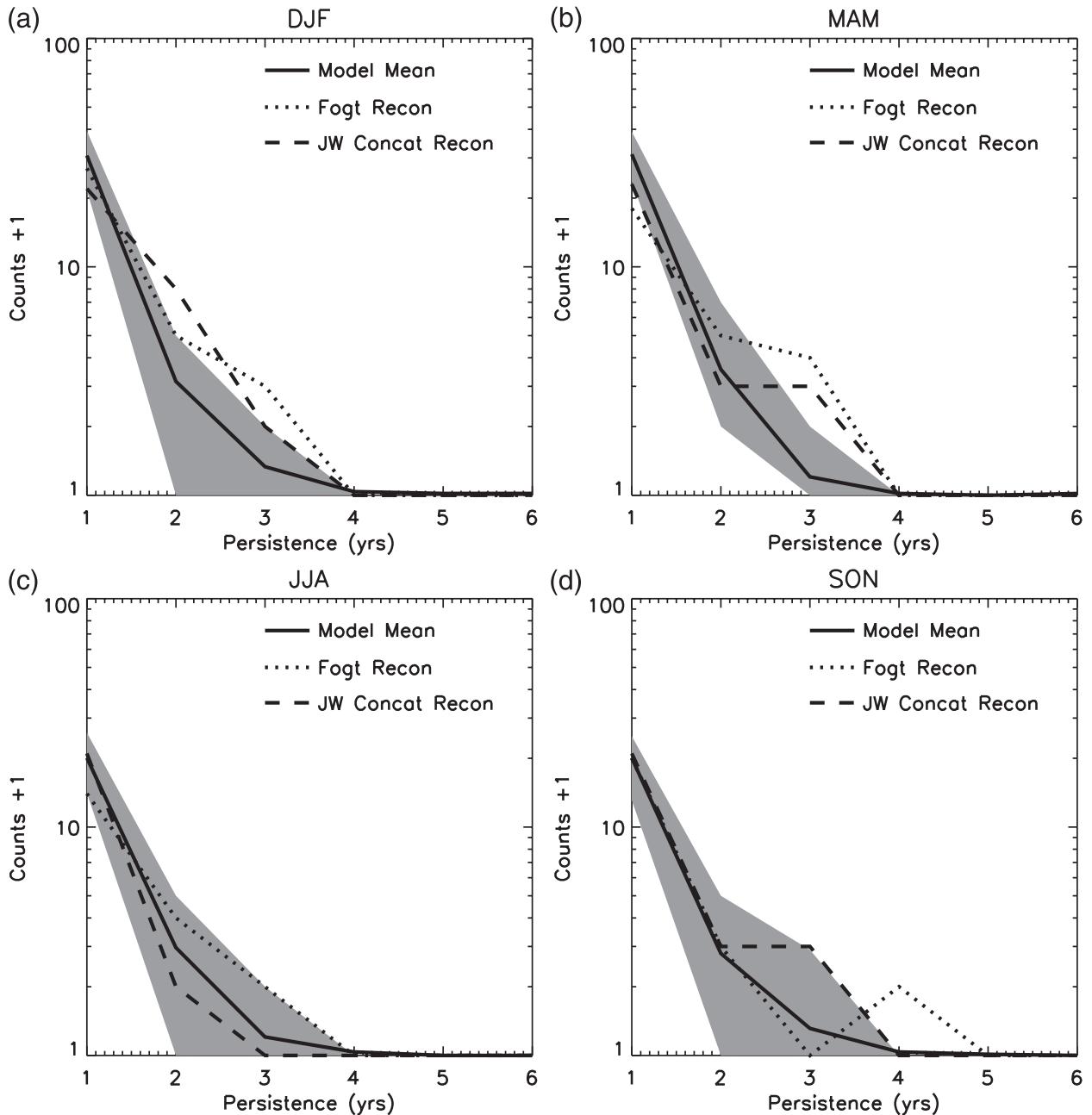


FIG. 2. Total counts for SAM peaks (defined as outside  $\pm 1$  standard deviation) based on the duration of the peak (in years). For example, a duration of 3 yr indicates a case where a particular SAM index remained  $>1$  (or  $<1$ ) standard deviation for 3 continuous years. The shaded region corresponds to the 5–95 percentiles based on the spread of the ensemble simulations. Note that the vertical axis is logarithmic and in units of “counts + 1.” (a) DJF, (b) MAM, (c) JJA, and (d) SON; 115 yr were used to determine the counts in (a) and (b) and 76 yr in (c) and (d).

displayed along with the reconstructions in Table 2. The time period is the same as in Fig. 2. Although still weak, only in MAM do the reconstructions indicate greater overall SAM persistence, which is undoubtedly influenced by the two strong multiyear events in reconstructions during this season not simulated by the models.

From current observational and modeling SAM studies, these large historical peaks likely result from periods of strong interaction from the tropics (Zhou and Yu 2004; L’Heureux and Thompson 2006; Fogt and Bromwich 2006) or strong stratospheric anomalies that later influence the troposphere through downward control

TABLE 2. Ensemble mean lag-1 autocorrelations and the 5/95 percentiles in the model control runs along with the Fogt and JWconcat reconstructions by season. Boldface values indicate the reconstructions are outside the model 5/95 percentiles.

	Model mean	5%	95%	Fogt	JWconcat
DJF	-0.02	-0.16	0.12	0.04	0.09
MAM	-0.03	-0.18	0.13	<b>0.21</b>	<b>0.18</b>
JJA	-0.01	-0.18	0.16	-0.12	-0.10
SON	-0.00	-0.20	0.20	-0.12	0.08

(Kindem and Christiansen 2001; Thompson et al. 2005). Tropical volcanic eruptions also may be a factor, although only the 1960s peaks align with a major volcanic eruption (Agung in 1963; Miller et al. 2006). Precisely detailing the causality of these historical SAM peaks would not only increase the scientific understanding of large-scale Southern Hemisphere climate variations but also would likely aid in improving model simulations to better capture these processes. The latter is especially true given that many of the possible mechanisms are not currently well represented by most GCMs (e.g., Kushner et al. 2001; Miller et al. 2006; Randall et al. 2007).

#### 4. SAM trends in the models and reconstructions, 1957–2005

##### a. Linear trends

Figure 1 clearly demonstrates that the grand ensemble mean captures the recent positive trends in DJF (Fig. 1a) and the lack thereof in JJA (Fig. 1c). In contrast, the recent trend in MAM (Fig. 1b) appears underestimated in the grand ensemble mean. Linear trends in the SAM indices from the model ensemble means, observations, and reconstructions are calculated to quantify these differences and address the relative role of forcing mechanisms. We investigate the 1957–2005 interval, the period of full overlap between all indices and when external forcing is the strongest. This period is comparable to the 1950–99 period chosen by Cai and Cowan (2007) and includes the 1960s and late 1990s peaks seen in many of the seasonal SAM indices (Fig. 1). Because this study considers all seasons and investigates a suite of models, observed indices, and the reconstructions from Part I, it is the most comprehensive study of the recent SAM trends to date.

Figure 3 displays the 1957–2005 seasonal trends for each index, with the 95% confidence intervals estimated from the standard deviation of equivalent length trends from the model's preindustrial control runs (scaled by the square root of the number of ensemble members for each model). Thus, externally forced trends are denoted

when these confidence intervals do not cross the zero line. The methodology conducted here is similar to other SAM modeling studies (e.g., Marshall et al. 2004; Cai and Cowan 2007). The model mean trends are the mean trend from each ensemble simulation within a given model, rather than the trend from the mean ensemble. Since the current focus is on attribution of these recent trends, the approach of Marshall et al. (2004) was used for the observed indices and reconstructions. This technique uses the mean standard deviation from the 17 models' control runs as a representation of the internal variability in the observed climate system. The resulting "observed" confidence intervals are then plotted about zero since the observations are not model simulations on which the confidence intervals are based. Of course, the observed confidence intervals could also be determined from the full-length reconstructions; however, this analysis is deferred until section 5, where the relative importance of the recent trends (model and reconstructions alike) is examined in an historical context. Table 3 lists the mean trends and includes the nonozone and ozone model mean trends not separately identified in Fig. 3.

In DJF, our results are in direct agreement with previous conclusions (e.g., Miller et al. 2006; Cai and Cowan 2007) despite different SAM definitions and averaging periods. Figure 3a clearly shows positive trends in nearly every model, most of which rise above internal variability and are comparable to the Marshall and HadSLP2 trends. The exception appears for models that do not contain time-variable ozone forcing, although two of these models [Canadian Centre for Climate Modelling and Analysis Coupled General Circulation Model version 3.1 (CCCMA CGCM3.1) and Meteorological Research Institute Coupled General Circulation Model, version 2.3.2a (MRI CGCM2.3.2a)] also produce trends above their own internal variability, albeit weaker than the ozone-containing models. The opposite (and significant) trend in the GISS Atmosphere–Ocean Model (AOM) is comparable to the negative trend in Cai and Cowan (2007) and near-zero trend in Miller et al. (2006). The observed indices of Marshall and HadSLP2-GW99 are also outside the range of internal variability and are very similar in magnitude to the ozone and grand ensemble means from Table 3. Although positive, both reconstructions produce weaker trends than observed; these differences will be discussed in section 4d. The relative magnitude of ozone versus nonozone forced trends in Table 3 is similar to previous studies (Arblaster and Meehl 2006; Cai and Cowan 2007). Given that all models that contain time-variable ozone produce significant trends in Fig. 3a and the ozone and nonozone means are significantly different at the  $p < 0.05$  level, it is clear that ozone depletion is a dominant mechanism in

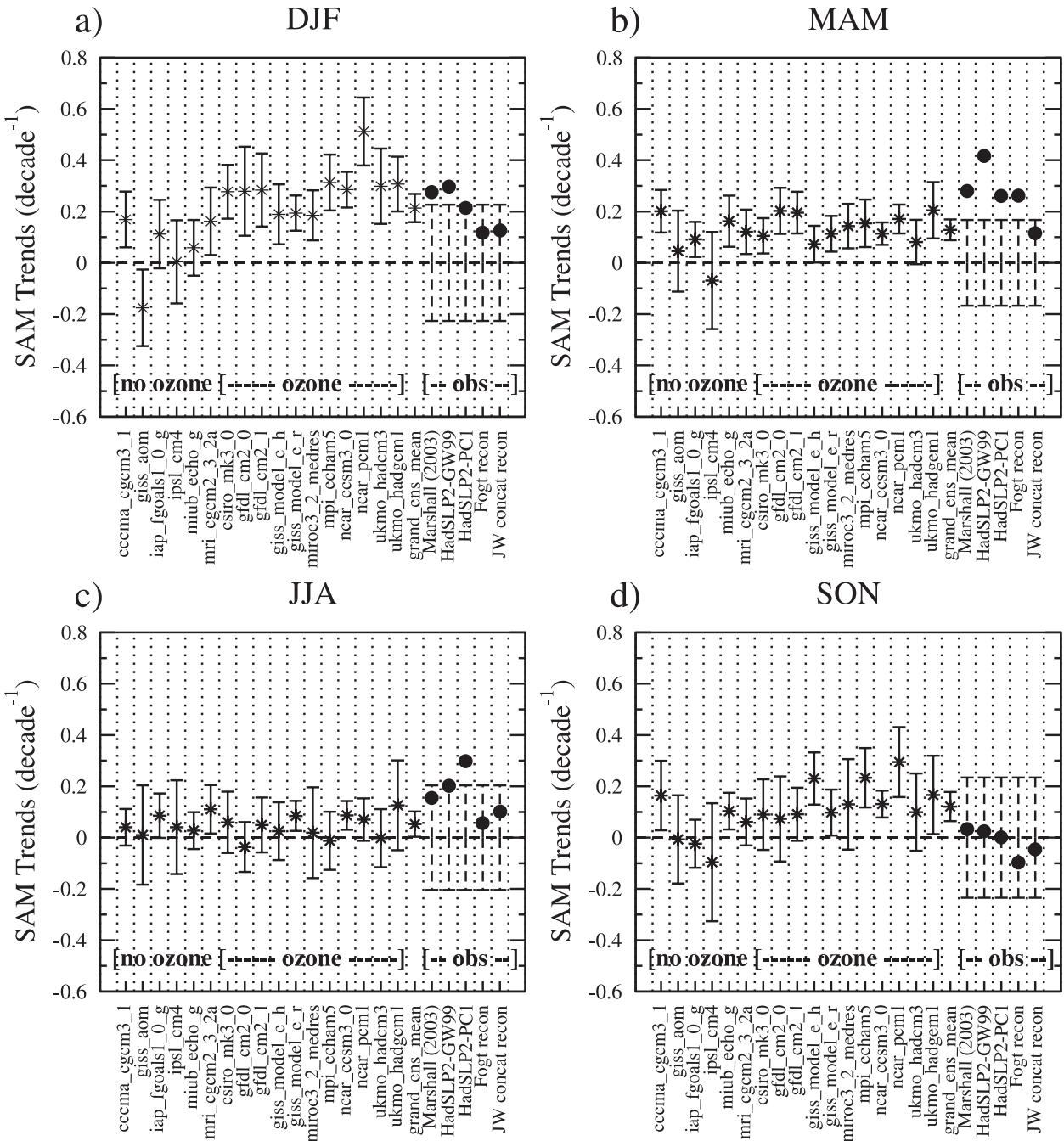


FIG. 3. Seasonal SAM trends in the model ensemble means, observations, and reconstructions for 1957–2005. For the models, the error bars are the 95% confidence intervals (scaled by the square root of the number of ensemble members for each model) from equivalent length trends from an unforced, preindustrial control run, and they represent the range of trends expected from internal climate variability. For the observations, the mean range of the control trends is centered on zero.

producing the recent summer SAM trends (Perlwitz et al. 2008).

Despite the significant trend in MAM in the Marshall index (Marshall 2007), this season has not been examined exclusively with climate models (many studies examine

December–May trends). Figure 3b indicates significant positive MAM trends in most indices during 1957–2005, with the ozone and no-ozone mean trend magnitudes being indistinguishable in MAM (Table 3). Nonetheless, the model trends in this season are considerably weaker

TABLE 3. Seasonal SAM trends (in units of decade<sup>-1</sup>) in 1957–2005 for models with (“ozone”) and without (“no ozone”) time-variable ozone forcing, the grand ensemble mean (“mean”) and the Marshall (2003) index. Insignificant trends within the 95% confidence interval expected from internal variability are marked with an asterisk.

	Ozone	No ozone	Mean	Marshall
DJF	0.27 ± 0.07	0.09 ± 0.11*	0.21 ± 0.05	0.27 ± 0.23
MAM	0.13 ± 0.05	0.12 ± 0.09	0.13 ± 0.04	0.28 ± 0.17
JJA	0.05 ± 0.07*	0.06 ± 0.09*	0.05 ± 0.05*	0.16 ± 0.20*
SON	0.15 ± 0.07	0.06 ± 0.11*	0.12 ± 0.06	0.03 ± 0.23*

than observed (Figs. 1b, 3b, and Table 3), although not statistically different. The Fogt reconstruction reproduces the observed trend, but the JWconcat reconstruction trend is weaker and within internal climate variability.

During JJA (Fig. 3c), the majority of trends lie within the range of internal climate variability; only three models indicate forced trends. The HadSLP2-PC1 also indicates a strong forced trend. However, given that even fewer early data are available in the high southern latitudes to constrain the early HadSLP2 (because of less ship traffic; Allan and Ansell 2006), the large HadSLP2 trends are likely to be overestimated. Furthermore, the positive JJA trend in the Marshall index is also misleading. Examining the smoothed Marshall index in Fig. 1c, it is clear that the index displays very little trend after 1970. The positive trend over the 1957–2005 period is brought about by negative values in the early 1960s, in particular the 1964 value, which is more than 3.5 standard deviations below the 1957–2005 mean (suggested in Part I to be related to the Agung eruption of the year before). Thus, the models correctly display weak SAM trends in this season, especially as ozone depletion should have little influence on the JJA trends because of the polar night (Cai and Cowan 2007; Roscoe et al. 2006). However, as the greenhouse gases increase throughout the twenty-first century, the SAM in May–July is projected to shift toward its positive polarity in all models (Miller et al. 2006; cf. their Fig. 12).

Another interesting and yet relatively unstudied issue are the SON SAM trends (Fig. 3d). Here, 6 of the 11 models with ozone forcing and a third of those without produce significant trends in the SAM, while the observations and reconstructions overall display near-zero trends clearly within internal variability. As in summer, there is a suggestion of a difference between the ozone and nonozone model means (Table 3), although these differences are not statistically significant. Nonetheless, this is the only season where all observed indices indicate weak trends and many models (although less than half) display forced trends.

Notably, the L’Institut Pierre-Simon Laplace Coupled Model version 4 (IPSL CM4) model consistently displays one of the largest differences in terms of the sign and magnitude of the trend (Fig. 3). This in agreement with Connolley and Bracegirdle (2007), who note that this model has the lowest Southern Hemisphere MSLP skill score in all of the models studied here. Similarly, their study also noted higher skill scores for many of the models that contain temporal ozone variations. This suggests the inclusion of time-variable ozone leads to better simulation of the Southern Hemisphere basic state and, hence, the better agreement of these models with observations in Fig. 3 and Table 3.

#### b. Equality of model trends with observed and reconstruction trends

Although the observed and modeled trends are not statistically distinguishable (at  $p < 0.05$  level) in any season, there are some intriguing apparent differences between the models and observations that deserve further study. Since there are over 30 degrees of freedom for each slope, their difference ( $b_1 - b_2$ ) is approximately normally distributed,  $N \sim (b_1 - b_2, \sigma_{b_1}^2 + \sigma_{b_2}^2)$ , where  $b_1$  and  $b_2$  are slopes from an ensemble simulation and observed index (or reconstruction), respectively, and  $\sigma_{b_1}$   $\sigma_{b_2}$  are the standard errors about each of the slopes. Using this property, the null hypothesis that each ensemble trend is equal to each of the other trends (i.e.,  $b_1 = b_2$ ) was tested. Figure 4 presents the probability distributions that the null hypothesis of equal trends is true (divided into ozone and nonozone groups) using traditional box plots.

In DJF (Fig. 4a), it is clear from the distribution that trends in models with nonozone have a greater probability of being different than the Marshall and HadSLP2 indices. More than 75% (50%) of the models with (without) ozone depletion have a >40% (<40%) chance that their trends are equal to these observed indices. For the reconstructions, where the trends are weaker, the reverse is true. This again highlights the importance of ozone depletion in producing trends that are consistent with observations.

In MAM (Fig. 4b), 75% of the model trends have less than a 50% chance that their trends are equal to the Marshall and HadSLP2 indices and the Fogt reconstruction. Notably, the HadSLP2-GW99 index, which displays the strongest positive trend in MAM, is statistically different ( $p < 0.10$ ) from more than 75% of the model ensemble simulations. Thus, although not meeting criteria for rejecting normal hypothesis tests, the probability distributions in Fig. 4b indicate there is a comparatively low probability that the observed and modeled SAM trends in MAM are equal. In fact, this

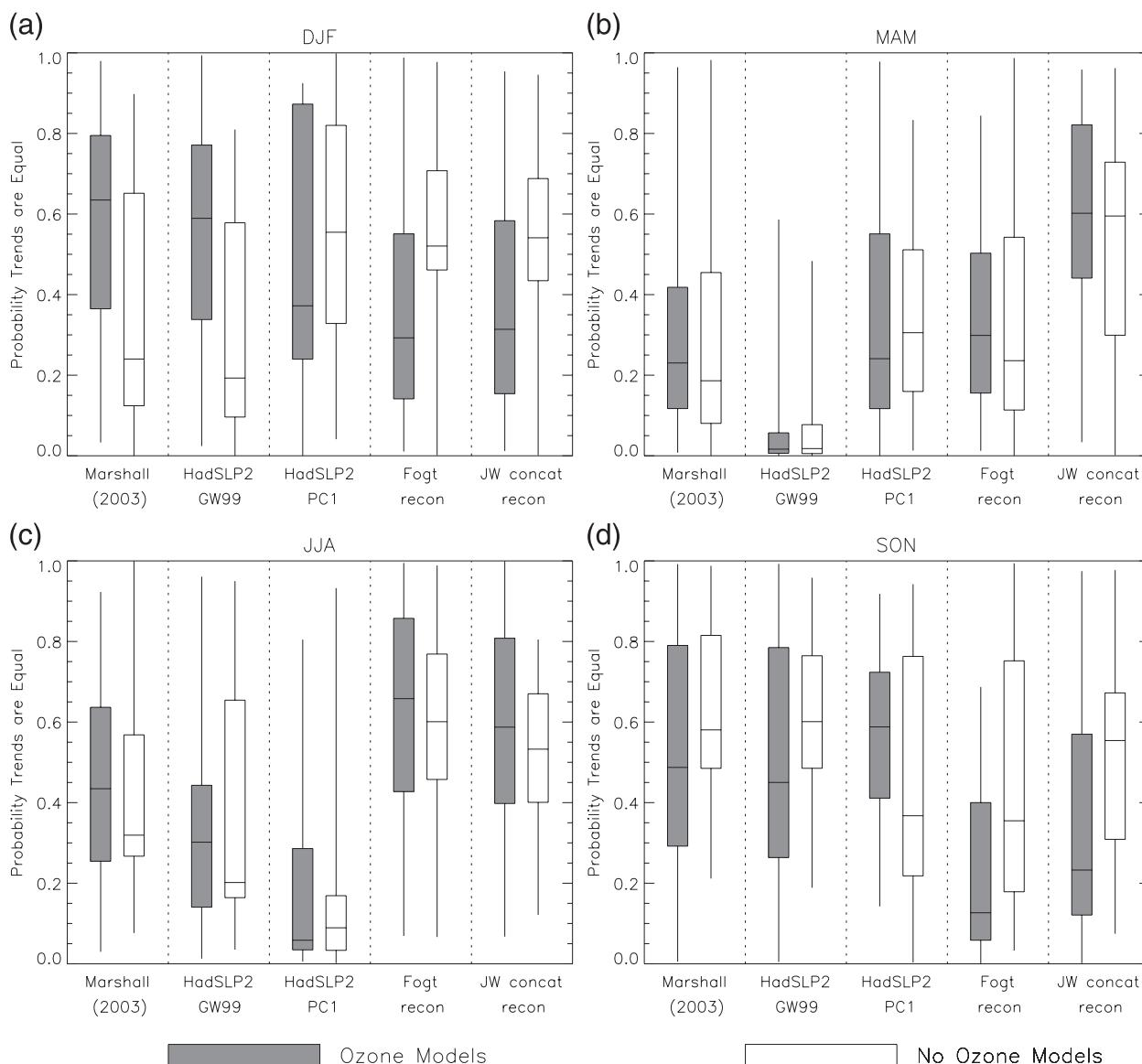


FIG. 4. Probability distributions that the null hypothesis (that model trends during 1957–2005 are equal to the indices defined on the horizontal axis) is true as represented by traditional box plots. The boxed region corresponds to the interquartile range, and the horizontal line within each box is the median of the distribution. The lines extending from the boxes reach to the maximum and minimum values of the distribution. The distributions are categorized into models containing temporal ozone variations (“ozone models”) and those without (“no-ozone models”). (a) DJF, (b) MAM, (c) JJA, and (d) SON.

probability is lower in MAM than in any other season. A similar but weaker argument could be made for the JJA SAM trends based on the distributions in Fig. 4c. However, recall that the HadSLP2 trends are potentially unreliable, and the Marshall trend is dominated by the presence of a strong negative outlier in 1964.

Using these probability distributions further highlights the separation between the ozone models and nonozone models in SON (Fig. 4d). More than 75% of the nonozone models have more than a 50% chance of being equivalent to the Marshall and HadSLP2 SAM indices.

Furthermore, for all but HadSLP2 the median probability of the ozone models lies below the interquartile range of the probability distribution from the nonozone models. That is, the probability that the ozone and nonozone trends are equal, is less than 25% (Table 3).

Overall, Fig. 4 therefore suggests there are substantial (albeit, statistically insignificant) differences between ozone and nonozone models in both DJF and SON. It also demonstrates that all models produce notably weaker trends than the Marshall and HadSLP2 indices and the Fogt reconstruction in MAM. While the differences in DJF

highlight the relative importance of ozone forcing in this season, the overall weaker trends in MAM and stronger ozone model trends in SON have important implications for the underlying nature of these seasonal trends.

### c. Spatial MSLP trends during MAM and SON

To further investigate these apparent differences and highlight their potential causality, the spatial MSLP trends from the model with the strongest positive and negative mean ensemble SAM trend from Fig. 3 are displayed in Fig. 5, along with the HadSLP2 trends as a representation of observed trends. The patterns in other models (with SAM trends of the same sign in Fig. 3) are consistent with these representative models (not shown). The spatial correlation of the model trends with the HadSLP2 trends is given in the bottom right of Figs. 5a–d. Although there are uncertainties in the early HadSLP2 data, before 1979 its biases are smaller than the reanalyses in places where station data are available (Jones and Lister 2007). In places where no data are available, it is difficult to precisely know the actual MSLP trends, and the HadSLP2 pattern should be viewed only qualitatively.

In both seasons, models with negative (Figs. 5a,b) and positive (Figs. 5c,d) SAM trends project nearly the entire spatial structure in a marked annular, SAM-like pattern. In MAM (left column of Fig. 5), the annular pattern corresponds fairly well with the HadSLP2 pattern, giving a negative spatial correlation in Fig. 5a ( $r = -0.48$ ) and a positive spatial correlation in Fig. 5c ( $r = 0.71$ ). However, it is clear that the trend magnitude in both cases is much weaker than observed, giving rise to the lower overall weaker trends in Fig. 3b. Nonetheless, the strong correspondence between Fig. 5c and Fig. 5e suggests the models are correctly simulating the observed spatial structure but underpredicting its amplitude.

In contrast, during SON the strong annular structure of the model trends gives weak spatial correlations with the HadSLP2 pattern (Figs. 5b,d;  $r = -0.02$ ,  $r = -0.07$ ). This weak correlation results from the models missing the area of large observed negative SLP trends throughout the Pacific Ocean, especially in the South Pacific Ocean near the west Antarctic coast (Fig. 5f at 135°W). As the South Pacific is often strongly influenced by tropical ENSO teleconnections (Turner 2004), part of this regional observed trend is likely influenced from the tropics. The fact that the models miss this asymmetry suggests they may underestimate the tropical ENSO impacts on the SAM that are common in SON (Fogt and Bromwich 2006). These remote influences on the high latitudes alter the spatial trend patterns, making them less zonally symmetric, thereby partly explaining the

model differences in Figs. 3d and 4d. Another possible reason why the model and observed spatial trends are different could result from recent stratospheric temperature trends. During September and October, despite cooling from ozone depletion, portions of the stratosphere have warmed considerably (Johanson and Fu 2007). These regions of warming are due to an enhanced Brewer–Dobson circulation through tropospheric wave driving of the stratosphere. Notably, the IPCC models do not simulate this warming and only have cooling in the stratosphere, a response expected from ozone loss and greenhouse gas increases. The cooling is most marked in models with ozone loss, thus potentially explaining why these models exhibit larger differences than observations (Lin et al. 2009).

### d. Reconstruction trend differences

Of particular concern are the substantial differences (again statistically insignificant) between the reconstruction trends and the observed trends (Fig. 3). Although some of these differences are related to the period selected for trend comparisons, the majority of the differences can be explained by limitations of the predictors forming the reconstructions. Figure 6 shows the seasonal pressure trends for 1957–2005 using HadSLP2 (contoured every 0.2 hPa decade<sup>-1</sup>) as well as the sign of the observed station trends used in the Fogt reconstruction. Table 4 lists these trends as well as the correlation of each station with the Marshall index. As the JWconcat reconstruction uses many more stations over the 1957–2005 period, and because the JWconcat reconstruction trends agree better with observations in all but MAM, the few stations used in the Fogt reconstruction provide a better understanding of how the reconstruction trends differ. For the station data, positive (negative) trends are plotted as black (gray) circles, and their size is proportional to their trend (see figure legend).

In DJF (Fig. 6a), although the overall pressure trend pattern is consistent with a shift toward a positive SAM, there are notable regional differences in the trend pattern, which may arise from influences from the tropics in this season (L'Heureux and Thompson 2006; Fogt and Bromwich 2006). Of these, there are negative trends in New Zealand, where half the stations used in the Fogt reconstruction reside (Table 4). Since the reconstructions can be considered as a weighted sum of the anomalies at the predictor stations, the DJF reconstruction trends are reduced by the negative trends in the New Zealand stations. The trend reduction is further exacerbated since the New Zealand stations receive considerable weight because of their high correlation with the SAM index (Fig. 3 in Part I; Table 4).

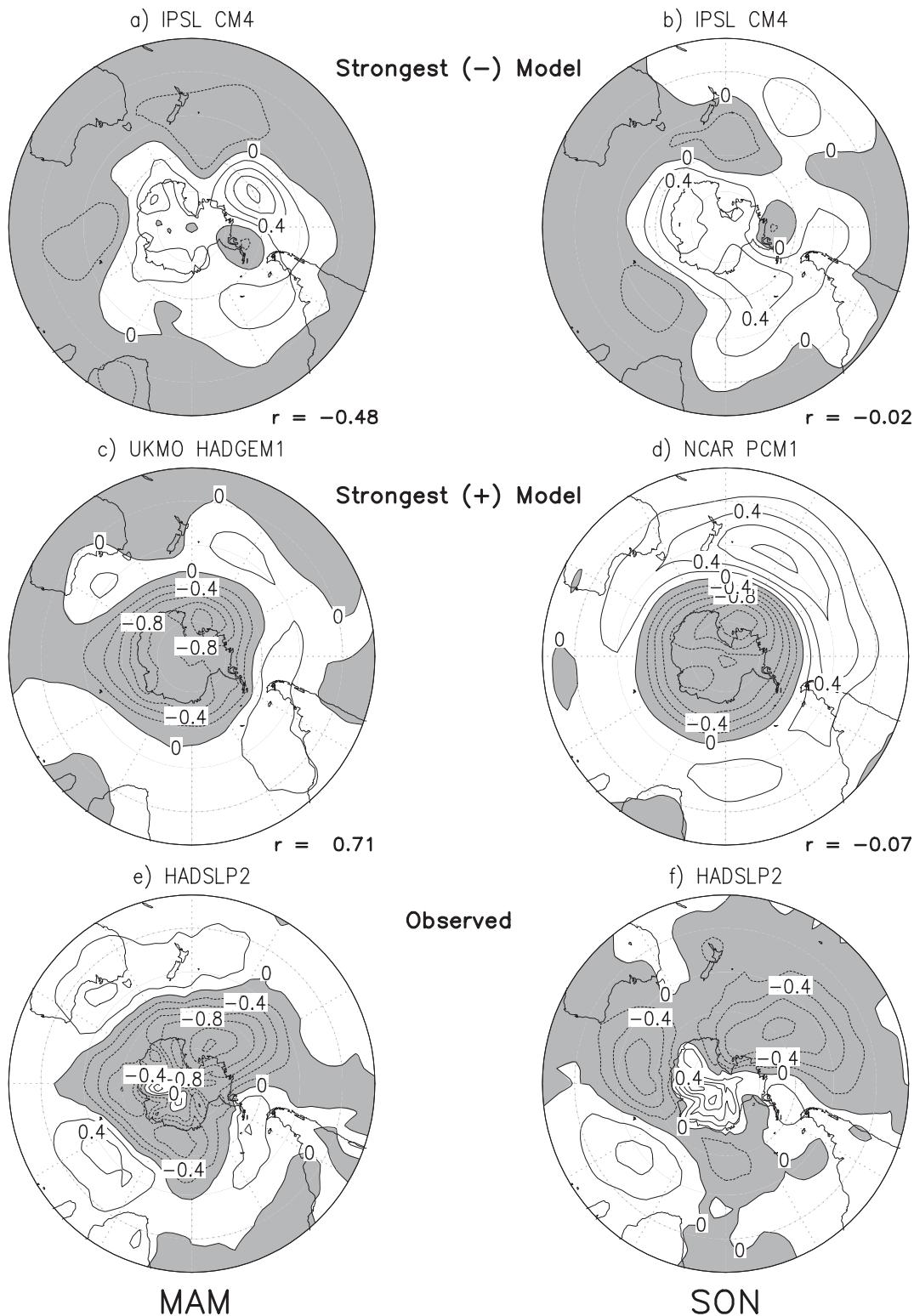


FIG. 5. (left) MAM and (right) SON MSLP trends in 1957–2005 from (a),(b) model with strongest negative SAM trends; (c),(d) model with strongest positive SAM trends; (e),(f) observed trends based on HadSLP2. Contour interval is 0.2 hPa decade<sup>-1</sup>, and negative contours are shaded. In bottom right of (a)–(d) is the spatial correlation with the HadSLP2 trend pattern for each season.

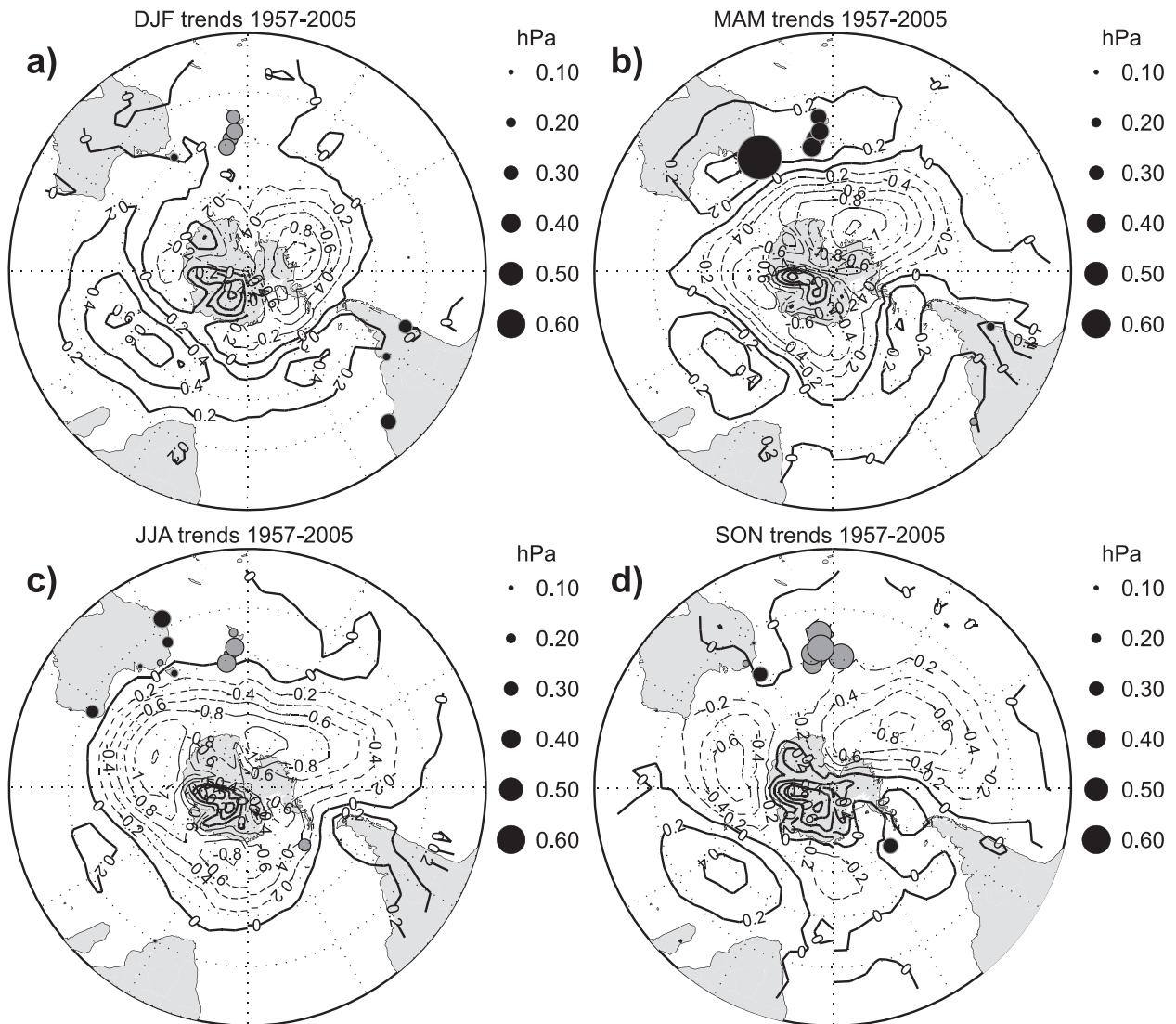


FIG. 6. HadSLP2 MSLP trends in 1957–2005. Contour interval is  $0.2 \text{ hPa decade}^{-1}$ , and negative contours are dashed. MSLP trends in the station data used in the Fogt reconstruction are plotted based on their sign: positive trends are black and negative are gray, with the size of the circle proportional to the trend (as given in the legend for each figure). (a) DJF, (b) MAM, (c) JJA, and (d) SON.

In MAM (Fig. 6b) and JJA (Fig. 6c), the pressure trend pattern is more zonally symmetric, and the station locations capture much of the hemispheric fluctuations, making the Fogt reconstruction trends better aligned with the observed trends (Figs. 3b,c). During MAM, the JWconcat reconstruction is much lower than observed ( $0.12 \text{ decade}^{-1}$  in JWconcat versus  $0.28 \text{ decade}^{-1}$  in Marshall). The weaker JWconcat trend is due to higher values prior to 1980 and lower values after this date (Fig. 1b) compared to the observed indices and the Fogt reconstruction (Fig. 2 in Part I). The weaker JJA trend in the Fogt reconstruction arises from a strong observed trend near Antarctica, which is better captured in the JWconcat reconstruction as it contains Antarctic stations

after 1957 (Fig. 5 in Part I). In SON, the distribution of predictor stations captures portions of this zonal wave-3 pattern, but the strong negative pressure trend in the New Zealand stations and the positive pressure trend at Orcadas (located near the Antarctic Peninsula) combine to produce slightly more negative reconstruction trends than observed.

Unfortunately, the limitations of the predictors in capturing a hemispheric pattern are unavoidable because of the geographic distribution of long-term stations. However, Part I determined that the major reconstruction peaks were representative of full hemispheric SAM-like anomalies patterns using all available station data and not simply regional responses captured by the predictors

TABLE 4. Station statistics for stations used in the Fogt reconstruction. For each season a station used, its correlation ( $r$ ) with the Marshall index and its trend from 1957 to 2005 ( $b$ , in units of hPa decade<sup>-1</sup>) are given.

Station	Lat	Lon	DJF		MAM		JJA		SON	
			$r$	$b$	$r$	$b$	$r$	$b$	$r$	$b$
Adelaide	-34.9	138.5			0.50	-0.38	0.44	0.19		
Alice Springs	-23.8	133.9					0.33	0.12		
Auckland	-36.9	174.8	0.53	-0.24	0.51	-0.11	0.57	-0.21	0.46	-0.60
Brisbane	-27.4	153.1					0.47	0.37		
Buenos Aires	-34.6	-58.6	0.48	0.17	0.61	0.38				
Chatham	-44.0	-176.6			0.57	0.46	0.66	0.07	0.69	-0.61
Christchurch	-43.5	172.6	0.66	-0.26			0.66	-0.31	0.71	-0.61
Dunedin	-45.9	170.5	0.66	-0.27	0.29	0.98	0.68	-0.38	0.74	-0.43
Durban	-30.0	31.0					0.40	0.10		
Hobart	-42.9	147.3	0.51	0.15			0.52	0.23	0.65	0.28
Hokitika	-42.7	171.0					0.64	-0.17	0.62	-0.37
Melbourne	-37.8	145.0					0.49	-0.11	0.44	-0.17
Orcadas	-60.7	-44.7					-0.19	-0.36	-0.56	0.30
Perth	-31.9	116.0			0.31	-0.16	0.54	0.28		
Port Elizabeth	-34.0	25.6			0.26	0.15	0.35	0.02		
Rio de Janeiro	-22.9	-43.2	0.41	0.34						
Santiago/Pudah	-33.4	-70.8	0.29	0.30	0.59	0.41				
Sydney	-33.9	151.2	0.39	0.02			0.50	0.25	0.34	0.06
Wellington	-41.3	174.8	0.62	-0.27			0.66	-0.41	0.66	-0.62

(cf. Fig. 10 of Part I). As will be discussed later, the trends in other periods between the reconstructions and observed indices are better aligned. Together, these facts provide strong evidence that the reconstructions are a reliable estimate of overall SAM variability (Part I), and only minor deviations from observed trends are expected (Fig. 3).

### 5. Historical significance of recent SAM trends

Up until this point, the analysis has not addressed the relative significance of the recent trends, which is a key aspect afforded by the new seasonal reconstructions, nor has it provided explanation regarding the weaker model trends (and their forcing mechanism) in MAM. To examine this, running 30-yr trends in the reconstructions and the Marshall index are plotted in Fig. 7. The 30-yr trends are chosen as they are sufficiently long enough to remove high-frequency climate noise but short enough to examine externally forced trends; consistent results were obtained using 40- and 50-yr trends. The 95% confidence intervals expected from internal climate variability (as in Fig. 3) are also plotted in Fig. 7.

In DJF (Fig. 7a), only 30-yr trends starting after 1965 are significant. Despite weaker trends from 1957 to 2005, the Fogt (JWconcat) reconstruction becomes significant at  $p < 0.05$  ( $p < 0.10$ ) after 1970. Furthermore, the running trends clearly highlight the historical importance of the recent DJF trends. In no other 30-yr period during the last 150 yr are the trends of equivalent or

greater magnitude. In light of this new historical perspective, it is clear that the ozone-induced changes primarily responsible for the significant DJF trends are unprecedented, thereby providing clear evidence of rapid anthropogenic impacts on climate.

In MAM (Fig. 7b), the recent trends also rise above the range from internal variability; however, there are also periods of significant ( $p < 0.05$ ) *negative* SAM trends, in particular the trend from 1928 to 1959 is indicated by both reconstructions. Examining this interval in Fig. 1b reveals these are the years following the 1930 peak when the SAM remained consistently negative. Interestingly, the magnitude of these negative trends in MAM is almost as strong as the highest positive trend in DJF (Fig. 7a) and larger than the recent trends in MAM. The dramatic historical trends in the reconstructions during MAM suggest that the model confidence intervals in Fig. 3 (and Figs. 7, 8) are too small, which in turn suggests that forced trends may be indicated more often than they occur. This is true for both the models and observations alike: with Fig. 7b, it is evident that the recent trends in observations are not unique, despite the indication of a forced response in Fig. 3b. Perhaps the underestimation of the confidence intervals is not surprising, given that the internal variability of the models simulate comparatively few SAM peaks of three or more years (Fig. 2), which induce these large trends. Although the models simulate an annular spatial MSLP trend pattern in MAM (left column of Fig. 5), the fact that a portion (or all) of the observed trend may be due

## 30-yr Observed Trends

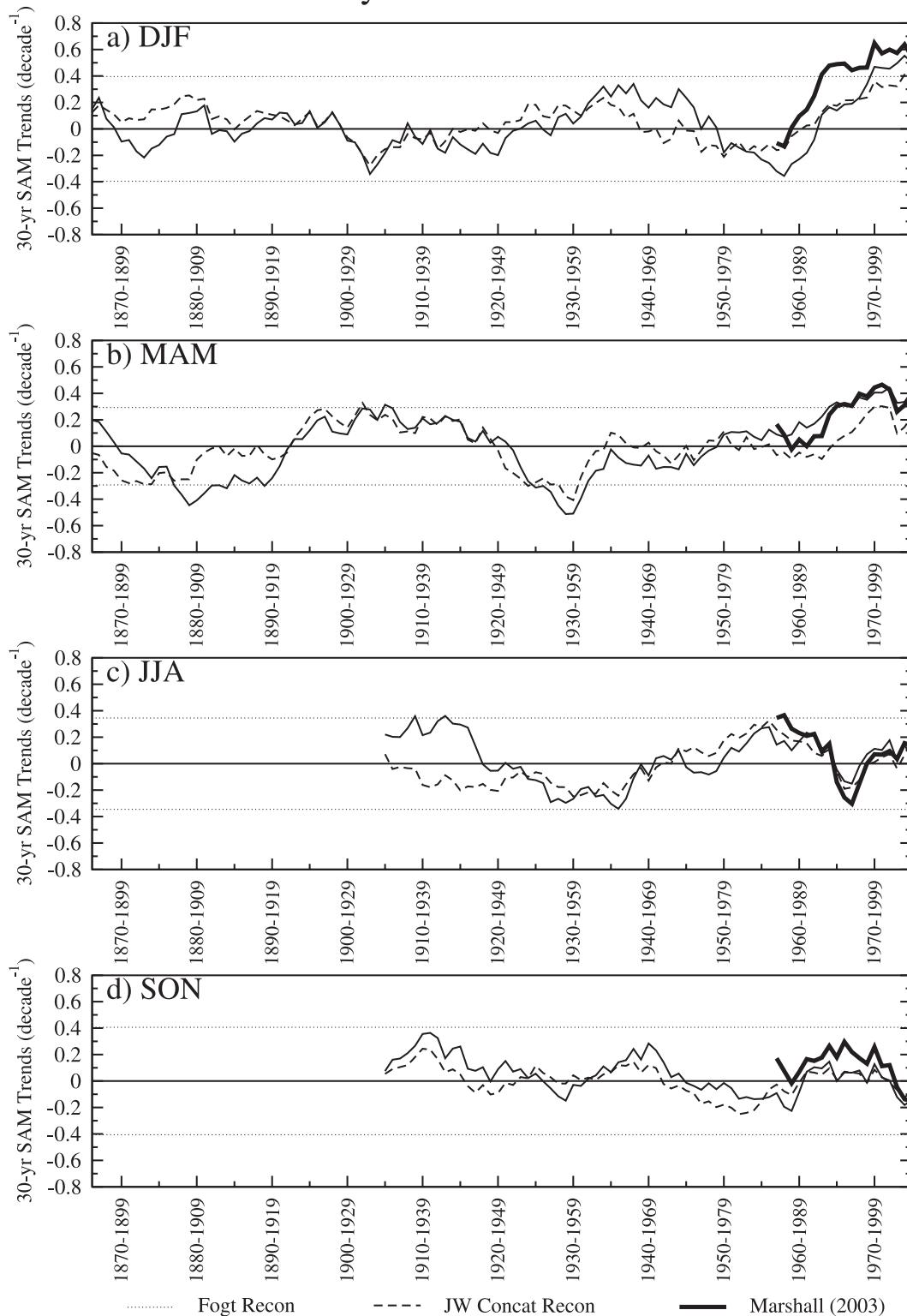


FIG. 7. Running magnitude of 30-yr trends from the reconstructions and the Marshall index. The dotted lines represent the range of internal climate variability from the mean of the models' preindustrial control runs. (a) DJF, (b) MAM, (c) JJA, and (d) SON.

### 30-yr GCM Trends

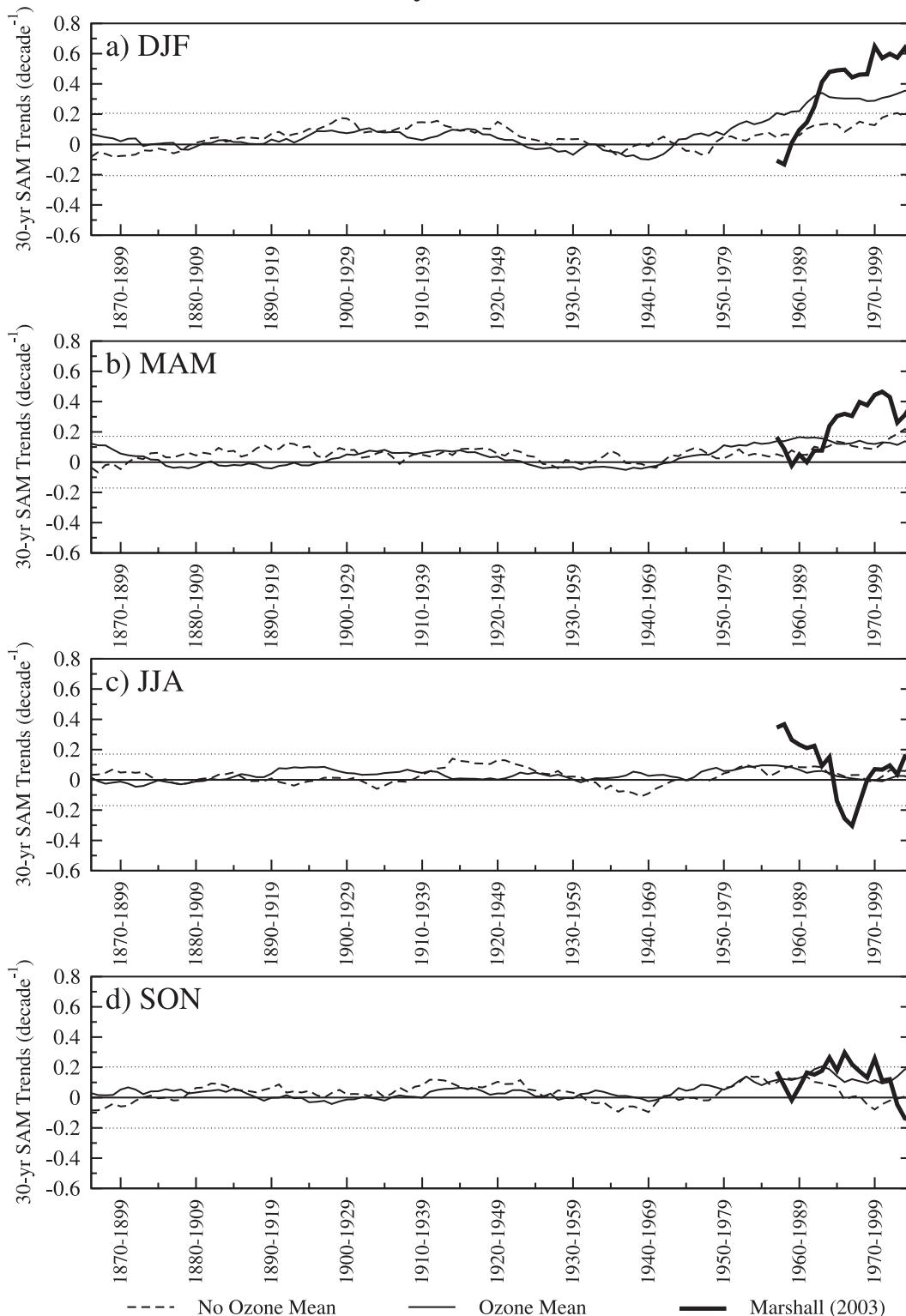


FIG. 8. As in Fig. 7, but for the no-ozone and ozone model means and the Marshall index. The confidence intervals are rescaled by square root of 6, the number of no-ozone models.

to natural climate variability helps to explain why the model trends are much weaker in this season.

There is marked multidecadal variability in the trends during JJA and SON (Figs. 7c,d). However, in these seasons the magnitude of the historical trends fails to emerge beyond the range of internal climate variability in either reconstruction. As in other seasons, the reconstruction trends throughout the twentieth century agree well with each other and with the Marshall index during the last 50 yr. An exception to this occurs in early JJA, when the reconstructions are most different. Part I suggests that the Fogt reconstruction is more reliable in this season, as it is not based on potentially unreliable winter reanalysis data (Bromwich and Fogt 2004).

How well do the AR4 models capture the relative significance of the recent trends? To answer this we separate the models into those that do and do not contain time-variable ozone forcing (as in Miller et al. 2006) and plot the 30-yr trends for these means along with the Marshall index in Fig. 8. Here, the confidence intervals are rescaled by  $\sqrt{6}$  (the square root of the number of nonozone models) so that they are an appropriate range for models without ozone forcing (and a conservative range for models with time-variable ozone forcing). Note that, prior to  $\sim 1979$ , ozone is prescribed as a (fixed) seasonal cycle in the models. Therefore, differences between the ozone and nonozone mean throughout much of the twentieth century are considered climate noise arising from different model sensitivities and configurations. Prior to 1880 the values in Fig. 8 should be examined with caution, as subtle shifts in the multimodel means occur when models with different starting dates (Table 1) are included.

During DJF (Fig. 8a), the trends in both means are similar throughout the twentieth century, with only subtle differences before the 1970s. The two model means diverge strongly when 1980 is included into the 30-yr trends ( $\sim 1950\text{--}79$  tick), corresponding to when some models begin to include decreasing stratospheric ozone. However, the fact that the nonozone mean does emerge above internal climate variability suggests that greenhouse gases have played a role in the summer SAM trends, albeit a much weaker one than ozone depletion (Arblaster and Meehl 2006; Cai and Cowan 2007; Roscoe and Haigh 2007). This also corresponds with a few nonozone models producing significant DJF trends (Fig. 3a).

In MAM, throughout all 30-yr periods both model means do not emerge strongly and continuously outside of the range of internal climate variability. However, the ozone model mean is consistently positive after the 1950–79 tick, just below the confidence interval. Similarly, the nonozone mean approaches the confidence interval during the last 30-yr period. These results sug-

gest that the SAM is becoming more positive recently in MAM, but 30-yr trends are still not strong enough to rise outside of the range of internal climate variability. Given that the majority of models underestimate the magnitude of the recent trends in this season (Figs. 3b, 4b), it is likely that the current trend primarily arises from a natural climate cycle as the models do not capture the duration of natural climate fluctuations in this season. However, as the trends are consistently positive in the last 40–50 yr, there are hints of a weaker forced component that will likely emerge in the next decade if these trends continue.

In JJA, the models agree well with the reconstructions and highlight the fact that essentially no strong SAM trend has been observed during the last 100 yr (Fig. 8c). Meanwhile, the 30-yr trends in the ozone models during SON (Fig. 8d) show a similar response as in MAM, with a continuous (although insignificant) positive trend since the 1970s. The Marshall index also displays a positive trend from the 1960s to the 1990s; however, its trend over the last 30 yr is negative. Although the differences between these trends are not statistically significant at  $p < 0.10$ , the probability that they are equal is even lower than that displayed in Fig. 4d. Isolated forcing mechanism runs from the PCM conducted by Arblaster and Meehl (2006) were investigated to understand why the models are producing the continuous positive SAM trend in SON. Over the last 50 yr, ozone-only runs produce negative SAM trends in SON, while greenhouse gas-only trends produce consistently positive trends, suggesting greenhouse gases are dominating the positive SAM trends. However, more tests are needed to increase the confidence in this assertion. Nonetheless, if greenhouse gases are the primary mechanism and tropical teleconnections continue to influence the SAM in this season, it is likely that the observed trend will soon be statistically different from the model trend in SON. Thus, caution should be warranted when examining future impacts of the SAM and its changes in SON from this particular subset of climate models.

## 6. Summary and conclusions

This paper has examined SAM variability during the twentieth century using the observation-based reconstructions detailed in Part I along with a suite of simulations from the IPCC AR4 archive. The reconstructions show decadal to multidecadal variability (Part I), with peaks of at least 2-yr duration at  $\sim 1960$  in SON and DJF and  $\sim 1930$  in MAM and SON (Fig. 1). Similar peaks are not captured in the internal climate cycles of the GCMs in the control simulations (Fig. 2). A possible mechanism leading to these peaks in the reconstructions could

be remote influences from the tropics, a feature not well simulated in most GCMs.

Features generally well resolved by the models and the reconstructions are the recent positive trends in DJF and lack thereof in JJA (Fig. 3). Both the model and observed trends were found to be outside the range of internal climate variability during DJF, suggesting that they are forced externally. While the reconstructions also display positive trends in DJF, they are weaker than observed because of regional asymmetries captured by the reconstructions' predictors (although the Fogt reconstruction emerges above a 90% confidence interval after 1970). The attribution work presented herein strongly suggests stratospheric ozone depletion is the dominant mechanism driving these trends, although forcing from the greenhouse gases also plays a smaller role, in agreement with Roscoe and Haigh (2007). These recent DJF trends are the strongest in the last 150 yr in model simulations and reconstructions alike, thereby providing clear evidence of rapid anthropogenic impacts on climate.

In MAM, the attribution analysis during 1957–2005 indicates that trends in the models, observations, and Fogt reconstruction have emerged outside the range of internal variability, possibly suggesting that external forcing is similarly playing a role in this season. However, because of the models missing the SAM peaks in MAM that last more than 2 yr, it is likely that the internal variability in this season is underestimated, which in turn increases the likelihood of incorrectly identifying a forced trend. As the reconstructions display historical trends of greater magnitude than the recent trend in the observations, we suggest that there is a high probability of a strong natural component in the recent trend, with external forcing playing a minor role. The models support this conclusion as well, since their trends agree the least with observations in MAM compared to any other season.

In contrast, many models that contain time-variable ozone forcing demonstrate forced positive trends in spring (SON) after 1957, absent in both observations and reconstructions. Spatially, observations show a zonal-wave 3 pattern in the MSLP trends, which may be partly induced from the tropics, while the models display very strong zonally symmetric pressure trends. Notably, the IPCC models miss regions of stratospheric warming during September and October that might also add to the differences in this season. These model and observation differences have been increasing over the last 30 yr, and if the opposing trends continue they will likely become significantly different in the near future. In turn, care should be exercised when examining the SAM changes and its impacts on Southern Hemisphere climate in this season.

The assessment and attribution of seasonal SAM trends will undoubtedly help with understanding the historical and future impacts this climate mode has across the Southern Hemisphere, especially the dramatic recent precipitation trends across southern Australia (i.e., Hendon et al. 2007). Additional work is needed to investigate the causality of the historical SAM peaks that last multiple years, including why the models do not adequately resolve these features. Turning to the future, the rate of ozone recovery will be a crucial issue to resolve for summer SAM projections (i.e., Shindell and Schmidt 2004; Perlwitz et al. 2008; Son et al. 2008). Also important is the role of the Southern Ocean in outgassing natural and absorbing anthropogenic CO<sub>2</sub> during positive SAM phases (i.e., Lovenduski et al. 2007), which may provide a positive feedback to future SAM trends. Similarly, the way the models handle the stratosphere–troposphere coupling (Fogt et al. 2009) and tropical interactions cannot be overlooked, especially since they are already generating differences between the models and the reconstructions and observations in austral spring.

*Acknowledgments.* We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison and the WCRP's Working Group on Coupled Modelling for their roles in making available the WCRP CMIP3 multimodel dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. The majority of the research was completed while the lead author held a National Research Council Research Associateship Award at NOAA. RLF, DHB, and AJM acknowledge partial support from the National Science Foundation Grant OPP-0337943, and DHB also acknowledges support from NSF Grant ATM-0751291. JP recognizes support from the NOAA Climate Program Office. We thank William Neff for comments on an earlier version of the manuscript, Julie Arblaster for assistance with the PCM simulations, and Martin Widmann for help with the statistical calculations in section 4b. Comments from two anonymous reviewers helped to clarify and strengthen the manuscript in several places, and are greatly appreciated.

#### 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.
- Arblaster, J. M., and G. A. Meehl, 2006: Contributions of external forcings to southern annular mode trends. *J. Climate*, **19**, 2896–2905.
- Bromwich, D. H., and R. L. Fogt, 2004: Strong trends in the skill of the ERA-40 and NCEP–NCAR reanalyses in the high and middle latitudes of the Southern Hemisphere, 1958–2001. *J. Climate*, **17**, 4603–4619.

- , —, K. I. Hodges, and J. E. Walsh, 2007: A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions. *J. Geophys. Res.*, **112**, D10111, doi:10.1029/2006JD007859.
- Cai, W. J., and T. Cowan, 2007: Trends in Southern Hemisphere circulation in IPCC AR4 models over 1950–99: Ozone depletion versus greenhouse gas forcing. *J. Climate*, **20**, 681–693.
- , P. H. Whetton, and D. J. Karoly, 2003: The response of the Antarctic Oscillation to increasing and stabilized atmospheric CO<sub>2</sub>. *J. Climate*, **16**, 1525–1538.
- Connolley, W. M., and T. J. Bracegirdle, 2007: An Antarctic assessment of IPCC AR4 coupled models. *Geophys. Res. Lett.*, **34**, L22505, doi:10.1029/2007GL031648.
- Fogt, R. L., 2007: Investigation of the southern annular mode and the El Niño–Southern Oscillation interactions. Ph.D. dissertation, Atmospheric Sciences Program, Department of Geography, The Ohio State University, 212 pp.
- , and D. H. Bromwich, 2006: Decadal variability of the ENSO teleconnection to the high-latitude South Pacific governed by coupling with the southern annular mode. *J. Climate*, **19**, 979–997.
- , J. Perlwitz, S. Pawson, and M. A. Olsen, 2009: Intra-annual relationships between polar ozone and the SAM. *Geophys. Res. Lett.*, **36**, L04707, doi:10.1029/2008GL036627.
- 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.
- Gillett, N. P., and D. W. J. Thompson, 2003: Simulation of recent Southern Hemisphere climate change. *Science*, **302**, 273–275.
- , R. J. Allan, and T. J. Ansell, 2005: Detection of external influence on sea level pressure with a multi-model ensemble. *Geophys. Res. Lett.*, **32**, L19714, doi:10.1029/2005GL023640.
- , T. D. Kell, and P. D. Jones, 2006: Regional climate impacts of the Southern Annular Mode. *Geophys. Res. Lett.*, **33**, L23704, doi:10.1029/2006GL027721.
- Gong, D., and S. Wang, 1999: Definition of Antarctic Oscillation index. *Geophys. Res. Lett.*, **26**, 459–462.
- Hegerl, G. C., and Coauthors, 2007: Understanding and attributing climate change. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 663–746.
- Hendon, H. H., D. W. J. Thompson, and M. C. Wheeler, 2007: Australian rainfall and surface temperature variations associated with the Southern Hemisphere annular mode. *J. Climate*, **20**, 2452–2467.
- Johanson, C. M., and Q. Fu, 2007: Antarctic atmospheric temperature trend patterns from satellite observations. *Geophys. Res. Lett.*, **34**, L12703, doi:10.1029/2006GL029108.
- Jones, J. M., and M. W. Widmann, 2004: Atmospheric science—Early peak in Antarctic oscillation index. *Nature*, **432**, 290–291.
- , R. L. Fogt, M. Widmann, G. J. Marshall, P. D. Jones, and M. Visbeck, 2009: Historical SAM variability. Part I: Century-length seasonal reconstructions. *J. Climate*, **22**, 5319–5345.
- Jones, P. D., and D. H. Lister, 2007: Intercomparison of four different Southern Hemisphere sea level pressure datasets. *Geophys. Res. Lett.*, **34**, L10704, doi:10.1029/2007GL029251.
- Karoly, D. J., 2003: Ozone and climate change. *Science*, **302**, 236–237.
- Karpechko, A. Y., N. P. Gillett, G. J. Marshall, and J. A. Screen, 2009: Climate impacts of the southern annular mode simulated by the CMIP3 models. *J. Climate*, **22**, 3751–3768.
- Kindem, I. T., and B. Christiansen, 2001: Tropospheric response to stratospheric ozone loss. *Geophys. Res. Lett.*, **28**, 1547–1550.
- Kushner, P. J., I. M. Held, and T. L. Delworth, 2001: Southern Hemisphere atmospheric circulation response to global warming. *J. Climate*, **14**, 2238–2249.
- L’Heureux, M. L., and D. W. J. Thompson, 2006: Observed relationships between the El Niño–Southern Oscillation and the extratropical zonal-mean circulation. *J. Climate*, **19**, 276–287.
- Lin, P., Q. Fu, S. Solomon, and J. M. Wallace, 2009: Temperature trend patterns in Southern Hemisphere high latitudes: Novel indicators of stratospheric change. *J. Climate*, in press.
- Lovenduski, N. S., N. Gruber, S. C. Doney, and I. D. Lima, 2007: Enhanced CO<sub>2</sub> outgassing in the Southern Ocean from a positive phase of the Southern Annular Mode. *Global Biogeochem. Cycles*, **21**, GB2026, doi:10.1029/2006GB002900.
- Marshall, G. J., 2003: Trends in the southern annular mode from observations and reanalyses. *J. Climate*, **16**, 4134–4143.
- , 2007: Half-century seasonal relationships between the Southern Annular Mode and Antarctic temperatures. *Int. J. Climatol.*, **27**, 373–383.
- , P. A. Stott, J. Turner, W. M. Connolley, J. C. King, and T. A. Lachlan-Cope, 2004: Causes of exceptional atmospheric circulation changes in the Southern Hemisphere. *Geophys. Res. Lett.*, **31**, L14205, doi:10.1029/2004GL019952.
- Meehl, G. A., and Coauthors, 2007: Global climate projections. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 747–846.
- Miller, R. L., G. A. Schmidt, and D. T. Shindell, 2006: Forced annual variations in the 20th century Intergovernmental Panel on Climate Change Fourth Assessment Report models. *J. Geophys. Res.*, **111**, D18101, doi:10.1029/2005JD006323.
- Perlwitz, J., S. Pawson, R. L. Fogt, J. E. Nielsen, and W. D. Neff, 2008: The impact of stratospheric ozone hole recovery on Antarctic climate. *Geophys. Res. Lett.*, **35**, L08714, doi:10.1029/2008GL033317.
- Randall, D. A., and Coauthors, 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 592–662.
- Rauthe, M., A. Hense, and H. Paeth, 2004: A model intercomparison study of climate change-signals in extratropical circulation. *Int. J. Climatol.*, **24**, 643–662.
- Roscoe, H. K., and J. D. Haigh, 2007: Influences of ozone depletion, the solar cycle and the QBO on the Southern Annular Mode. *Quart. J. Roy. Meteor. Soc.*, **133**, 1855–1864.
- , G. J. Marshall, and J. C. King, 2006: Low potential for stratospheric dynamical to be implicated in the large winter warming in the central Antarctic Peninsula. *Quart. J. Roy. Meteor. Soc.*, **132**, 803–820.
- Sexton, D. M. H., 2001: The effect of stratospheric ozone depletion on the phase of the Antarctic Oscillation. *Geophys. Res. Lett.*, **28**, 3697–3700.
- Shindell, D. T., and G. A. Schmidt, 2004: Southern Hemisphere climate response to ozone changes and greenhouse gas increases. *Geophys. Res. Lett.*, **31**, L18209, doi:10.1029/2004GL020724.
- Solomon, S., R. W. Portmann, T. Sasaki, D. J. Hofmann, and D. W. J. Thompson, 2005: Four decades of ozonesonde measurements over Antarctica. *J. Geophys. Res.*, **110**, D21311, doi:10.1029/2005JD005917.
- , and Coauthors, 2007: Technical summary. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 19–91.
- Son, S.-W., and Coauthors, 2008: The impact of stratospheric ozone recovery on the Southern Hemisphere westerly jet. *Science*, **320**, 1486–1489.

- Stone, D. A., A. J. Weaver, and R. J. Stouffer, 2001: Projection of climate change onto modes of atmospheric variability. *J. Climate*, **14**, 3551–3565.
- Thompson, D. W. J., and S. Solomon, 2002: Interpretation of recent Southern Hemisphere climate change. *Science*, **296**, 895–899.
- , M. P. Baldwin, and S. Solomon, 2005: Stratosphere–troposphere coupling in the Southern Hemisphere. *J. Atmos. Sci.*, **62**, 708–715.
- Turner, J., 2004: Review: The El Niño–Southern Oscillation and Antarctica. *Int. J. Climatol.*, **24**, 1–31.
- Visbeck, M., 2009: A station-based southern annular mode index from 1884 to 2005. *J. Climate*, **22**, 940–950.
- Zhou, T., and R. Yu, 2004: Sea-surface temperature induced variability of the Southern Annular Mode in an atmospheric general circulation model. *Geophys. Res. Lett.*, **31**, L24206, doi:10.1029/2004GL021473.