Skillful seasonal prediction of the Southern Annular Mode and Antarctic ozone

William J. M. Seviour *

Department of Physics, University of Oxford, UK

Steven C. Hardiman

Met Office Hadley Centre, Exeter, UK

Lesley J. Gray

National Centre for Atmospheric Science, University of Oxford, UK

Neal Butchart, Craig MacLachlan, Adam A. Scaife

Met Office Hadley Centre, Exeter, UK

* *Corresponding author address:* William Seviour, Atmospheric Physics, Clarendon Laboratory, Parks Road, Oxford, UK, OX1 3PU.

Email: seviour@atm.ox.ac.uk

ABSTRACT

Using a set of seasonal hindcast simulations produced by the Met Office Global Seasonal Forecast System 5 (GloSea5), significant predictability of the Southern Annular Mode (SAM) is demonstrated during the austral spring. The correlation of the September-November mean SAM with observed values is 0.64, which is statistically significant at the 99% confidence level, and is similar to that found recently for the North Atlantic Oscillation in the same system. Significant skill is also found in the prediction of the strength of the Antarctic stratospheric polar vortex at 1–4 month lead times. Due to the observed strong correlation between interannual variability in the strength of the Antarctic stratospheric circulation and ozone concentrations, it is possible to make skillful predictions of Antarctic column ozone amounts. By studying the variation of forecast skill with time and height, it is shown that skillful predictions of the SAM are significantly influenced by stratospheric anomalies which descend with time and are coupled with the troposphere. This effect allows skillful statistical forecasts of the October mean SAM to be produced based only on mid-stratosphere anomalies on 1st August. Together, these results both demonstrate a significant advance in the skill of seasonal forecasts of the Southern Hemisphere and highlight the importance of accurate modelling and observation of the stratosphere in producing long-range forecasts.

1. Introduction

Accurate prediction of the atmospheric circulation several months in advance relies on the presence of low-frequency predictable signals in the climate system. It has now been demonstrated that the stratosphere is an important pathway for the communication of predictable tropical signals across the globe; in particular, the El Niño-Southern Oscillation (ENSO) (Bell et al. 2009; Ineson and Scaife 2009; Hurwitz et al. 2011), Quasi-Biennial Oscillation (QBO) (Marshall and Scaife 2009; Garfinkel and Hartmann 2011), and 11-year solar cycle (Haigh 2003; Gray et al. 2013). These teleconnections allow for the possibility of significant predictability in regions remote from the direct effect of the signal. Despite this, many operational seasonal forecast models include only a poor representation of the stratosphere (Maycock et al. 2011), and it has been suggested that this contributes to their lack of seasonal forecast skill in the extratropics (Smith et al. 2012).

Furthermore, because stratospheric anomalies persist for longer than those in the troposphere and can influence surface weather patterns (e.g., Baldwin and Dunkerton 2001), the initial conditions of the stratosphere itself can act as a source of enhanced predictability (Baldwin et al. 2003; Charlton et al. 2003; Hardiman et al. 2011). The effect of the stratosphere on the troposphere is especially pronounced following a rapid midwinter breakdown of the strong westerly stratospheric polar vortex (known as a sudden stratospheric warming, SSW), and past work has focused on the influence of these events on forecast skill (Kuroda 2008; Sigmond et al. 2013). However, SSWs are highly nonlinear events which are currently not predictable beyond about two weeks in advance (Marshall and Scaife 2010), limiting their usefulness in seasonal prediction. SSWs also occur almost exclusively in the Northern Hemisphere (NH), with only one event in the approximately 60 year record having been observed in the Southern Hemisphere (SH), in September 2002 (Roscoe et al. 2005).

The rarity of SSWs in the SH is a result of less dynamical forcing from vertically propagating planetary waves in the SH relative to the NH stratosphere. This, in turn, comes about because of lesser SH orography and land-sea temperature contrasts which can excite planetary waves. This reduced variability also means that anomalies in the Antarctic stratosphere persist for longer than those in the Arctic (Simpson et al. 2011). Hence, the SH stratospheric circulation may be predictable on longer time scales, and thus more useful for seasonal forecasts despite the lack of SSWs. Indeed, Thompson et al. (2005) and Son et al. (2013) have found that smaller-amplitude variations in the Antarctic stratospheric polar vortex are followed by coherent temperature and pressure anomalies at the Earth's surface which resemble the Southern Annular Mode (SAM) pattern. These observations led Roff et al. (2011) to find that improved forecasts of the SAM up to 30 days ahead may be achieved with a stratosphere-resolving model. The SAM is the dominant mode of variability of the extratropical Southern Hemisphere sea-level pressure and affects the position of storm tracks, rainfall, surface air temperature, and ocean temperatures across the extratropics (e.g., Silvestri and Vera 2003; Reason and Rouault 2005; Hendon et al. 2007). As such, there are considerable societal benefits, and interests in its prediction (Lim et al. 2013).

Another reason for interest in the prediction of the Antarctic stratosphere is the interannual variability in springtime ozone depletion, which can significantly affect the amount of harmful ultraviolet radiation reaching the Earth's surface over the Southern Hemisphere. The magnitude of this interannual variability is a significant fraction of the magnitude of long-term depletion caused by emission of chlorofluorocarbons (CFCs) and other ozone-depleting substances. While ozone-depleted air is confined over the polar region by the stratospheric polar vortex during winter and spring (resulting in the ozone hole), this air is released to mid-latitudes following the ultimate breakdown of the vortex (final warming) in late spring/early summer. The extent of the resulting summertime ozone depletion is largely determined by the total deficit in ozone over the Antarctic during spring (Bodeker et al. 2005).

Salby et al. (2012) have shown that interannual variations in Antarctic ozone depletion are highly correlated with changes in planetary wave forcing of the stratosphere. They found that the anomalous vertical Eliassen-Palm (EP) flux (a measure of meridional eddy heat flux) at 70 hPa poleward of 40°S during August-September explains almost all the interannual variance of anomalous ozone depletion during September–November. Using this relationship, they postulate that accurate prediction of planetary wave forcing could allow skillful seasonal forecasts of ozone depletion.

The influence of planetary wave forcing on ozone depletion comes about through both chemical and dynamical mechanisms. Planetary wave breaking causes an increase of the strength of the stratospheric residual mean meridional circulation (Haynes et al. 1991), with a resultant increase in large-scale descent and adiabatic warming over the pole. This warming inhibits the formation of polar stratospheric clouds which have a vital role in the activation of halogen species that cause the chemical depletion of ozone. The increased meridional circulation as well as an enhancement of horizontal two-way mixing caused by planetary wave breaking, also cause an increase in the dynamical transport of tropical ozone-rich air to the polar regions, further increasing ozone concentrations. Breaking planetary waves can also modify the geometry of the stratospheric polar vortex, stripping away elements of ozone-depleted air (Waugh et al. 1994), or in the extreme case of the 2002 SSW causing the ozone hole to split in two (Charlton et al. 2005).

Here, we address directly the influence of the stratosphere on springtime Antarctic seasonal forecast skill using a set of historical hindcasts (or historical re-forecasts) from a new operational system with a fully stratosphere-resolving general circulation model. We find significant skill in the prediction of the Antarctic stratospheric polar vortex up to four months in advance, including for the 2002 SSW. Using the observed relationship between column ozone quantities and the stratospheric circulation, we are then able to infer skillful predictions of springtime ozone depletion, confirming the hypothesis of Salby et al. (2012). This exceeds the lead-time of other contemporary ozone forecasts, which are typically no more than two weeks (Eskes 2005). The forecast system also shows highly significant levels of skill in the prediction of the surface SAM at seasonal lead times. By studying the variation of hindcast skill with time and height, we demonstrate that this skill is significantly influenced by the descent of long-lived stratospheric circulation anomalies.

2. Seasonal forecast system

The analysis in this paper is based on results from a set of hindcast predictions produced by the Met Office Global Seasonal Forecast System 5 (GloSea5) (MacLachlan et al. 2014). This system is based upon the HadGEM3 coupled general circulation model (Hewitt et al. 2011), with an atmospheric resolution of 0.83° longitude by 0.56° latitude, 85 quasi-horizontal atmospheric levels and an upper boundary at 85 km. The ocean resolution is 0.25° in longitude and latitude, with 75 quasi-horizontal levels. A 15-member ensemble of hindcasts was run for each year in the period 1996–2009. The hindcast length is approximately four months from three separate start dates spaced two weeks apart and centered on 1st August (07/25, 08/01, 08/09), with 5 members initialized on each start date. Members initialized on the same start date differ only by stochastic parameterization of model physics (Tennant et al. 2011).

Initial conditions for the atmosphere and land surface were taken from the ERA-Interim reanalysis (Dee et al. 2011), and initial ocean and sea-ice concentrations from the GloSea5 Ocean and Sea Ice Analysis, based on the FOAM data assimilation system (Blockley et al. 2013). Beyond initialization the model takes no further observational data, and contains no flux corrections or relaxations to climatology. The model lacks interactive chemistry, and ozone concentrations are fixed to observed climatological values averaged over 1994–2005, including a seasonal cycle (Cionni et al. 2011).

Scaife et al. (2014) have shown that this seasonal forecast system produces unprecedented skillful forecasts of the North Atlantic Oscillation during the Northern Hemisphere winter. The combined effects of ENSO, QBO and sea-ice teleconnections, as well as the increased ocean resolution which has improved the representation of Northern Hemisphere blocking events (Scaife et al. 2011), contribute to this skill.

Hindcast accuracy is verified by comparison to the ERA-Interim reanalysis (Dee et al. 2011). This provides a 'clean comparison' since the hindcasts exactly match ERA-Interim at the initialization date. The ERA-Interim data set has been demonstrated to have realistic representation of the stratospheric meridional circulation (Seviour et al. 2012; Monge-Sanz et al. 2013). It also assimilates observations of ozone concentrations, and this assimilation has been demonstrated to be in close agreement with independent satellite data (Dragani 2011).

3. Seasonal forecast results

a. Stratospheric polar vortex

The climatology of Antarctic stratospheric polar vortex winds in the GloSea5 hindcasts is compared to the ERA-Interim reanalysis climatology in Fig. 1. The strength of the stratospheric polar vortex is measured by the zonal-mean zonal wind ($\overline{U}$) at 60°S and 10 hPa, which is approximately the center of the mean position of the vortex in the mid-stratosphere. The composite for the GloSea5 hindcasts is formed from all the individual ensemble members over 1996–2009 (a total of 210 realizations), while that from ERA-Interim is a composite of all years from 1979–2010 (a total of 32 years). It can be seen that the mean of the GloSea5 hindcasts agrees very closely with ERA-Interim throughout the spring, with only a slight bias towards weaker winds in August and September. The interquartile and 95th percentile ranges of GloSea5 and ERA-Interim also agree well, although the ERA-Interim values are noisier as would be expected from a sample size consisting of fewer years.

The GloSea5 hindcast predictions of interannual variability of the Antarctic stratospheric polar vortex winds are shown in Fig. 2(a). Anomalies are defined from the relevant climatology of either GloSea5 or ERA-Interim. For GloSea5, this climatology is calculated from the mean of each day across all ensemble members in all years, while for ERA-Interim the climatology is the mean of each day, smoothed with a 30-day running mean (in order to account for its increased noise due to the fact it consists of only a single realization). Results are shown for September–November (SON) averages, corresponding to a 1–4 month lead time. The correlation between the GloSea5 ensemble mean and ERA-Interim is 0.73, which is statistically significant at the 99% confidence level. This correlation does not depend strongly on particular years; the correlation remains significant at the 95% level ($r = 0.57$) if the year 2002 is excluded. Significance is calculated using a two-tailed bootstrap test, whereby the percentile of the observed correlation is calculated from the distribution of correlations of a large number ($\sim 10,000$) of pairs of time series formed by re-sampling with replacement from the original time series. These significance tests make fewer assumptions about the underlying structure of the data than parametric tests (Wilks 2006), and are used throughout this study.

Two SSW events were simulated in the GloSea5 hindcasts; in 1997 and 2002. Time series of stratospheric polar vortex winds for these events are shown in Fig. 1(a) along with the observed 2002 SSW in Fig. 1(b). Note that although $\overline{U}$ at 60°S and 10 hPa does not quite become easterly for the 1997 event, it does become easterly *poleward* of 60°S, which satisfies the World Meteorological Organization definition of a SSW. Given the total of 210 ensemble hindcasts, these two simulated events suggest a frequency of Southern Hemisphere SSW events of approximately one in 100 years in the current climate (making the assumption that the model can accurately simulate the probability of these events). It can also be seen in 2(a) that 2002 has the most anomalous stratospheric polar vortex in the GloSea5 hindcasts, with 14 of 15 ensemble members simulating negative anomalies, and the most negative ensemble mean. It is therefore possible that the 2002 event was to some degree predictable about two months in advance, although it has not been determined whether this predictability comes from a preconditioning of the vortex, as suggested by Scaife et al. (2005), or the result of external forcing.

It should be noted that both the SSW events simulated by GloSea5 were vortex displacement events, in contrast to the vortex splitting event which occurred in 2002 (Charlton et al. 2005). This is demonstrated in Fig. 3, which shows geopotential height in the mid-stratosphere at the date of minimum $\overline{U}$ at 60°S and 10 hPa, for the two simulated events in GloSea5 and the observed event in ERA-Interim. The distinction between splitting and displacement SSW events is important because it has been observed that tropospheric anomalies are greater following vortex splitting events, at least in the Northern Hemisphere (Nakagawa and Yamazaki 2006; Mitchell et al. 2013).

The timing of the final warming of the stratospheric polar vortex has a significant effect on stratospheric temperature and ozone concentrations (Yamazaki 1987), as well as coupling of the stratosphere to the troposphere (Black and McDaniel 2007). The predictability of these events was investigated in GloSea5, but not found to be highly significant. This is probably because the mean timing of the final warming is towards the end of the four month hindcast simulation (around 20th November at 10 hPa), and the final warming does not occur before the end of the hindcast for some ensemble members, thereby introducing a bias in the mean. It is likely that shorter lead-time forecasts would be required to produce skillful predictions of the final warming date.

b. Ozone depletion

GloSea5 does not include interactive ozone chemistry, so in order to make ozone forecasts concentrations must be inferred from other meteorological variables. Total ozone quantities over the Antarctic polar cap have been found to be highly correlated with vertical EP flux poleward of 40°S (Weber et al. 2011; Salby et al. 2012). This diagnostic is not likely to be produced directly by operational seasonal forecast systems and requires high frequency output at high spatial resolution to calculate. However, vertical EP flux dominates variability of the stratospheric polar vortex, so it may be possible to use the strength of the vortex to infer ozone quantities.

SON mean total column ozone quantities area-weighted averaged over the polar cap (60–90°S) are shown in Fig. 4(a) for ERA-Interim and the Total Ozone Mapping Spectrometer

(TOMS) satellite instrument (Kroon et al. 2008). ERA-Interim data are highly correlated with TOMS, verifying the accuracy of ERA-Interim against direct satellite measurements (TOMS values are slightly higher than ERA-Interim; this is probably because TOMS cannot make observations during the polar night). The long-term trend in polar cap total column ozone is calculated by fitting a second-order polynomial to the data. This long-term trend is due to changes in concentrations of CFCs and other ozone-depleting substances, and largely unrelated to dynamical variability. On the other hand, shorter-term interannual changes are strongly related to dynamical variability. In Fig. 4(b) anomalies of polar cap total column ozone from the long-term trend are plotted against anomalies of the SON mean $\overline{U}$ at 60°S and 10 hPa. It can be seen that these two quantities are highly correlated ($r = -0.92$), meaning polar vortex variability explains approximately 85% of the variance of polar cap total column ozone anomalies.

This strong correlation makes it possible to use GloSea5 forecasts of polar vortex winds to produce inferred predictions of polar cap total column ozone quantities. Figure 2(b) shows the GloSea5 hindcasts along with the assimilated values from ERA-Interim. The correlation between the GloSea5 ensemble mean and ERA-Interim is 0.72, which is statistically significant at the 99% level. Errors from the regression in Fig. 2(b) for the inferred ozone quantities for each ensemble member are small compared to the spread between ensemble members, and so not plotted in this figure.

c. *Southern Annular Mode*

The SAM index in both GloSea5 and ERA-Interim is depicted as the difference between the normalized anomalies of zonally averaged mean sea-level pressure at 40°S and 65°S (Gong and Wang 1999). These anomalies are calculated from the respective climatologies of GloSea5 and ERA-Interim. The ERA-Interim SAM index calculated in this way is also highly correlated with other measures of the SAM, such as the station-based index of Marshall (2003). The GloSea5 hindcast skill for the prediction of the seasonal (SON) mean SAM index is shown in Fig. 5. The correlation of the GloSea5 ensemble mean and ERA-Interim is 0.64, which is statistically significant at the 99% level, confirming skillful prediction of the SAM at 1–4 month lead times. This is similar to the value for the NAO correlation skill of 0.62 found by Scaife et al. (2014) in the same seasonal forecast system.

Figure 6(a) shows the correlation of ERA-Interim and GloSea5 SON averaged mean sea-level pressure anomalies at each grid point in the Southern Hemisphere. As would be expected from the low frequency variability of ENSO, correlations are greatest over the tropical Pacific. However, the correlations are also as high as 0.7 across southern Australia and parts of Antarctica. On the other hand, correlations over southern Africa and South America are not found to be significant. It is perhaps unsurprising that there is little skill over the Andes region, since this is significantly above sea-level, so mean sea-level pressure is not well defined.

Correlations of GloSea5 SON average near-surface temperature with the gridded station-based data set HadCRUT4 (Morice et al. 2012) are shown in Fig. 6(b). This dataset is chosen because of the scarcity of temperature observations in the Southern Hemisphere, which introduces significant biases into reanalysis data. Again, the highest correlations are found near the tropical Pacific, but significant correlations of about 0.5 are found across eastern Australia, New Zealand and Antarctica. There are also significant correlations in southern Africa and South America. The extratropical regions where the greatest forecast skill is found are similar to those which are observed to be most affected by variations in the SAM (Gillett et al. 2006).

d. Stratosphere-troposphere coupling

Given that statistically significant skill in hindcasts of the stratospheric polar vortex is found at the same time of year as skill in predictions of the SAM, the question arises as to whether skill in one may affect the other. In order to investigate this, forecast skill as a function of lead-time and height is studied for polar cap (60-90°S) mean geopotential height anomalies (Z')[1]. Figure 7(a) shows the correlation of Z' in ERA-Interim with the GloSea5 ensemble mean hindcast values. Values are smoothed with a 30-day running mean before correlations are calculated, and plotted such that values for 15th September represent the correlation of the ERA-Interim and GloSea5 ensemble mean September mean values (without this smoothing, there are noisier but still significant correlations in a similar pattern).

As would be expected from the initialization of GloSea5 from ERA-Interim data, correlations are high in both the troposphere and the stratosphere for the August mean, due to predictability on weather timescales. However, tropospheric and lower-stratospheric skill rapidly decays and becomes statistically insignificant throughout September. In contrast, stratospheric correlations remain statistically significant throughout the hindcast simulation, and as high as 0.8 through to mid-October (corresponding to a 2–3 month lead time). This observed greater stratospheric than tropospheric skill might be expected from the longer 'memory' of the stratosphere; SAM decorrelation timescales are about 60–70 days in the stratosphere but only about 10 days in the troposphere during SON (Simpson et al. 2011).

Importantly, the region of high levels of stratospheric skill descends with time and is present at the tropopause at the same time as a re-emergence of significant tropospheric skill in mid-October. This re-emergence cannot be accounted for by the persistence of tropospheric anomalies, so must be the result of the effect of another predictable signal on the tropospheric circulation. An obvious candidate for such a signal is the polar stratosphere, since this remains predictable throughout the hindcast period. The re-emergence of tropospheric skill also occurs at the same time as the strongest observed coupling between the stratosphere and troposphere found in other studies (e.g., Thompson et al. 2005; Simpson et al. 2011).

In order to determine the stratospheric influence on tropospheric skill, a simple statistical forecast model is formed, which has as its only input the initial conditions of the Antarctic stratosphere. ERA-Interim values are used to produce this model based on the linear re-

[1] At 1000 hPa monthly mean Z' is highly correlated ($r = 0.98$) with the SAM index

gression of Z' at 10 hPa on 1st August with Z' at all other times and heights for 31 of the 32 years from 1979–2009. This model is then used to produce a hindcast of the 32nd year based on its Z' at 10 hPa on 1st August. The method ensures that no information from the hindcast year enters the model. The process is then repeated to make hindcasts of all 32 years; a procedure known as leave-one-out cross validation (Wilks 2006).

Figure 7(b) shows the average correlation of 30-day running means of these statistical forecasts with ERA-Interim values. As might be expected, skill is initially high in the mid-stratosphere but not the troposphere. As with the GloSea5 hindcasts, the region of high skill descends with time, and statistically significant correlations re-emerge in the troposphere throughout October. This demonstrates that skilful forecasts of the Antarctic troposphere during October can be produced based only on knowledge of Z' in the mid-stratosphere on 1st August. It also suggests that the re-emergence of tropospheric skill in the GloSea5 hindcasts in October is likely to be caused by persistent stratospheric anomalies which descend with time.

The statistical hindcasts in Fig. 7(b) show lower skill than the GloSea5 hindcasts at all times, and do not show statistically significant tropospheric correlations, nor the increase in upper-stratospheric skill during November. These observations could potentially be explained by the importance of non-linearities or the influence of external factors, such as ENSO, on the Antarctic stratosphere-troposphere system. Indeed, statistical hindcasts similar to those shown in Fig. 7(b) were produced based on the Niño-3 index, and found to have statistically significant tropospheric skill during November, but none at other times or heights. This is consistent with the results of Lim et al. (2013), who find the greatest correlation between tropical Pacific sea-surface temperatures and the SAM during November–January.

4. Discussion

We have demonstrated that Antarctic total column ozone amounts are predictable up to four months in advance during the austral spring, even with a model which lacks interactive chemistry. While using such a model has the advantage of being less computationally expensive than a chemistry-climate model, there are also some drawbacks. Primarily, the model will not be able to simulate zonal asymmetries in ozone concentrations or the feedback between ozone concentrations and stratospheric temperatures. Both these factors have been shown to be important in driving long-term trends in the SAM as a result of ozone depletion (Thompson and Solomon 2002; Crook et al. 2008; Waugh et al. 2009).

Perhaps more relevant for seasonal forecasts is the fact that we have not been able to determine whether the observed strong correlation between the stratospheric circulation and Antarctic ozone concentrations is dominated by a chemical or dynamical mechanism. If the relationship is dominated by a chemical mechanism, whereby enhanced descent over the pole inhibits the activation of ozone-depleting substances, we would expect the correlation to weaken as concentrations of these substances return to pre-industrial levels. Accurate forecasts of ozone with models lacking interactive chemistry would then not be possible. On the other hand, if the mechanism is largely dynamical, whereby transport of ozone-rich air from the tropics is the important factor, we would not expect the relationship to change in time. Although a study to distinguish these mechanisms has been carried out for chemistry-climate models (Garny et al. 2011), it has not been possible to do so in observations. In either case, we do not expect the relationship to break down soon, as concentrations of ozone-depleting substances are not projected to return to 1980 levels until the late 21st century (WMO 2011).

The correlation skill of 0.64 for the SON mean SAM in the GloSea5 hindcasts is greater than that of other contemporary seasonal forecast systems at similar lead times. For instance, Lim et al. (2013) report a correlation of 0.3–0.4 for the SON mean SAM from 1st August initialized forecasts using the Predictive Ocean and Atmosphere Model for Australia, version

2 (POAMA2). Significantly, this system has only two model levels in the stratosphere, and so is unable to simulate the stratosphere-troposphere coupling described here. Lim et al. (2013) suggest that the significant SAM predictability found from October–January in their system is the result of the influence of ENSO through a tropospheric teleconnection. These findings are not inconsistent with our results, since this time period is beyond the extent of the GloSea5 hindcasts, and largely after the stratospheric final warming, when the stratosphere is much less variable. Lim et al. (2013) were also mostly concerned with longer range forecasts (up to 6-month lead time) which are beyond the persistence time scales of the Antarctic stratosphere, but within those of the tropical Pacific.

Despite this significant correlation skill in hindcasts of the SAM, it is clear from Figure 6 that the amplitude of the ensemble mean hindcast is much less than that of observations. The signal-to-noise ratio (ratio of the standard deviation of the ensemble mean to that of all ensemble members) is just 0.4. For a 'perfect' forecast system (one in which observations are indistinguishable from an ensemble member), the signal-to-noise ratio and correlation are directly related (Kumar 2009), so that the expected correlation would be just 0.3. The fact that it is greater than this is because the average correlation between ensemble members and observations is much greater than that between pairs of ensemble members. A similar but smaller difference is also found for the stratospheric polar vortex forecasts. These results mean that individual ensemble members have a smaller predictable signal than observations.

Given this result, it might be expected that more skillful predictions could be obtained with a larger ensemble size. To illustrate the variation of hindcast skill with ensemble size we systematically sample smaller sets of forecasts from the full 15 members for each year, following the method of Scaife et al. (2014). This is repeated many times and an average value for a given sample size calculated. This variation of correlation skill with ensemble size for both the SON mean SAM and stratospheric polar vortex winds is shown in Fig. 8. These curves closely follow the theoretical relationship of Murphy (1990), which relies only on the mean correlation between pairs of ensemble members, $\langle r_{mm} \rangle$, and the mean correlation between individual ensemble members and observations, $\langle r_{mo} \rangle$, given by

$$r = \frac{\langle r_{mo} \rangle \sqrt{n}}{\sqrt{1 + (n-1)\langle r_{mm} \rangle}} \tag{1}$$

where r is the ensemble mean correlation, and n is the ensemble size. These curves are shown in Fig. 8, along with their asymptote for an infinite sized ensemble. Although the stratospheric forecasts cannot be greatly improved with a larger ensemble size in the current system, correlation scores of about 0.7 of the SAM could be achieved with an ensemble size near 30. Both have an asymptote near 0.8, similar to that found by Scaife et al. (2014) for the NAO.

5. Conclusions

Using a set of seasonal hindcasts initialized at the start of the austral spring, we have demonstrated skillful prediction of the interannual variability of the Antarctic stratospheric polar vortex at 1-4 month lead times. This includes extreme events such as the 2002 SSW, which is the most extreme year in the ensemble mean, and has one ensemble member which simulates a SSW. Because this variability is observed to be closely correlated with Antarctic column ozone amounts, we are able to infer skillful prediction of interannual variability in Antarctic ozone depletion.

We also find significant skill, which exceeds that of other contemporary seasonal forecast systems, in hindcasts of the spring mean SAM index. By studying the variation of this skill with time and height, we suggest that this skill is influenced by stratospheric anomalies which descend with time and are coupled with the troposphere in October and November. In fact, the influence of the stratosphere is such that skillful statistical predictions of the October SAM can be made using only information from 1st August in the mid-stratosphere.

Assuming that the 14 year period studied here is representative of future years, these results suggest that it may now be possible to make skillful seasonal forecasts of interannual variations in ozone depletion and large scale weather patterns across the Southern Hemisphere. They also demonstrate the importance of the inclusion of a well-resolved stratosphere in seasonal forecast models.

Acknowledgments.

WJMS was funded by a 3-year CASE studentship from the Natural Environment Research Council (NERC), supported in part by the Met Office. LJG was supported by the NERC National Centre for Atmospheric Science (NCAS). The work of SCH, NB, CM and AAS was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). ERA-Interim data used in this study were obtained from the European Centre for Medium-range Weather Forecasts data server and TOMS data were obtained from the British Atmospheric Data Centre.

REFERENCES

Baldwin, M. P. and T. J. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294 (dd)**, 581–584.

Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric memory and skill of extended-range weather forecasts. *Science*, **301**, 636–640.

Bell, C. J., L. J. Gray, A. J. Charlton-Perez, M. M. Joshi, and A. A. Scaife, 2009: Stratospheric Communication of El Niño Teleconnections to European Winter. *J. Climate*, **22**, 4083–4096, doi:10.1175/2009JCLI2717.1.

Black, R. X. and B. A. McDaniel, 2007: Interannual Variability in the Southern Hemisphere Circulation Organized by Stratospheric Final Warming Events. *J. Atmos. Sci.*, **64**, 2968–2974.

Blockley, E. W., et al., 2013: Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts. *Geosci. Model Dev. Discuss.*, **6**, 6219–6278, doi:10.5194/gmdd-6-6219-2013.

Bodeker, G., H. Shiona, and H. Eskes, 2005: Indicators of Antarctic ozone depletion. *Atmos. Chem. Phys.*, **5**, 2603–2615.

Charlton, A. J., A. O'Neill, W. A. Lahoz, A. C. Massacand, and P. Berrisford, 2005: The impact of the stratosphere on the troposphere during the southern hemisphere stratospheric sudden warming, September 2002. *Q. J. R. Meteorol. Soc.*, **131**, 2171–2188, doi:10.1256/qj.04.43.

Charlton, A. J., A. O'Neill, D. B. Stephenson, W. A. Lahoz, and M. P. Baldwin, 2003: Can knowledge of the state of the stratosphere be used to improve statistical forecasts of the troposphere? *Q. J. R. Meteorol. Soc.*, **129**, 3205–3224, doi:10.1256/qj.02.232.

Cionni, I., et al., 2011: Ozone database in support of CMIP5 simulations: results and corresponding radiative forcing. *Atmos. Chem. Phys.*, **11**, 11 267–11 292, doi:10.5194/acp-11-11267-2011.

Crook, J. A., N. P. Gillett, and S. P. E. Keeley, 2008: Sensitivity of Southern Hemisphere climate to zonal asymmetry in ozone. *Geophys. Res. Lett.*, **35**, L07 806, doi:10.1029/2007GL032698.

Dee, D. P., et al., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

Dragani, R., 2011: On the quality of the ERA-Interim ozone reanalyses: comparisons with satellite data. *Q. J. R. Meteorol. Soc.*, **137**, 1312–1326, doi:10.1002/qj.821.

Garfinkel, C. I. and D. L. Hartmann, 2011: The Influence of the Quasi-Biennial Oscillation on the Troposphere in Winter in a Hierarchy of Models. Part I: Simplified Dry GCMs. *J. Atmos. Sci.*, **68**, 1273–1289, doi:10.1175/2011JAS3665.1.

Garny, H., V. Grewe, M. Dameris, G. E. Bodeker, and A. Stenke, 2011: Attribution of ozone changes to dynamical and chemical processes in CCMs and CTMs. *Geosci. Model Dev.*, **4**, 271–286, doi:10.5194/gmd-4-271-2011.

Gillett, N. P., T. D. Kell, and P. D. Jones, 2006: Regional climate impacts of the Southern Annular Mode. *Geophys. Res. Lett.*, **33**, L23 704, doi:10.1029/2006GL027721.

Gong, D. and S. Wang, 1999: Definition of Antarctic Oscillation index. *Geophys. Res. Lett.*, **26**, 459–462, doi:10.1029/1999GL900003.

Gray, L. J., et al., 2013: A lagged response to the 11 year solar cycle in observed winter Atlantic/European weather patterns. *J. Geophys. Res.*, **118**, 13,405–13,420, doi:10.1002/2013JD020062.

Haigh, J. D., 2003: The effects of solar variability on the Earth's climate. *Philos. Trans. R. Soc. A*, **361**, 95–111, doi:10.1098/rsta.2002.1111.

Hardiman, S. C., et al., 2011: Improved predictability of the troposphere using stratospheric final warmings. *J. Geophys. Res*, **116**, D18 113, doi:10.1029/2011JD015914.

Haynes, P., M. McIntyre, T. G. Shepherd, and K. P. Shine, 1991: On the "downward control" of extratropical diabatic circulations by eddy- induced mean zonal forces. *J. Atmos. Sci.*, **48**, 651–678.

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. Clim.*, **20**, 2452–2467, doi:10.1175/JCLI4134.1.

Hewitt, H. T., D. Copsey, I. D. Culverwell, C. M. Harris, R. S. R. Hill, a. B. Keen, A. J. McLaren, and E. C. Hunke, 2011: Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system. *Geosci. Model Dev.*, **4**, 223–253, doi:10.5194/gmd-4-223-2011.

Hurwitz, M. M., P. A. Newman, L. D. Oman, and A. M. Molod, 2011: Response of the Antarctic Stratosphere to Two Types of El Niño Events. *J. Atmos. Sci*, **68**, 812–822, doi:10.1175/2011JAS3606.1.

Ineson, S. and A. A. Scaife, 2009: The role of the stratosphere in the European climate response to El Niño. *Nat. Geosci.*, **2**, 32–36, doi:10.1038/NGEO381.

Kroon, M., J. P. Veefkind, M. Sneep, R. D. McPeters, P. K. Bhartia, and P. F. Levelt, 2008:

Comparing OMI-TOMS and OMI-DOAS total ozone column data. *J. Geophys. Res.*, **113**, D16S28, doi:10.1029/2007JD008798.

Kumar, A., 2009: Finite Samples and Uncertainty Estimates for Skill Measures for Seasonal Prediction. *Mon. Weather Rev.*, **137**, 2622–2631, doi:10.1175/2009MWR2814.1.

Kuroda, Y., 2008: Role of the stratosphere on the predictability of medium-range weather forecast: A case study of winter 20032004. *Geophys. Res. Lett.*, **35**, L19 701, doi:10.1029/2008GL034902.

Lim, E.-P., H. H. Hendon, and H. Rashid, 2013: Seasonal Predictability of the Southern Annular Mode due to its Association with ENSO. *J. Climate*, **26**, 8037–8045, doi:10.1175/JCLI-D-13-00006.1.

MacLachlan, C., et al., 2014: Global Seasonal Forecast System version 5 (GloSea5): a high resolution seasonal forecast system. *Q. J. R. Meteorol. Soc.*, **submitted**.

Marshall, A. G. and A. A. Scaife, 2009: Impact of the QBO on surface winter climate. *J. Geophys. Res.*, **114**, D18 110, doi:10.1029/2009JD011737.

Marshall, A. G. and A. A. Scaife, 2010: Improved predictability of stratospheric sudden warming events in an atmospheric general circulation model with enhanced stratospheric resolution. *J. Geophys. Res.*, **115**, D16 114, doi:10.1029/2009JD012643.

Marshall, G. J., 2003: Trends in the southern annular mode from observations and reanalyses. *J. Clim.*, **16**, 4134–4143.

Maycock, A. C., S. P. E. Keeley, A. J. Charlton-Perez, and F. J. Doblas-Reyes, 2011: Stratospheric circulation in seasonal forecasting models: implications for seasonal prediction. *Clim. Dyn.*, **36**, 309–321, doi:10.1007/s00382-009-0665-x.

Mitchell, D. M., L. J. Gray, J. Anstey, M. P. Baldwin, and A. J. Charlton-Perez, 2013:

The Influence of Stratospheric Vortex Displacements and Splits on Surface Climate. *J. Climate*, **26**, 2668–2682, doi:10.1175/JCLI-D-12-00030.1.

Monge-Sanz, B. M., M. P. Chipperfield, D. P. Dee, A. J. Simmons, and S. M. Uppala, 2013: Improvements in the stratospheric transport achieved by a chemistry transport model with ECMWF (re)analyses: identifying effects and remaining challenges. *Q. J. R. Meteorol. Soc.*, **139**, 654–673, doi:10.1002/qj.1996.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res.*, **117**, D08 101, doi:10.1029/2011JD017187.

Murphy, J., 1990: Assessment of the practical utility of extended range ensemble forecasts. *Q. J. R. Meteorol. Soc.*, **116**, 89–125.

Nakagawa, K. I. and K. Yamazaki, 2006: What kind of stratospheric sudden warming propagates to the troposphere? *Geophys. Res. Lett.*, **33**, L04 801, doi:10.1029/2005GL024784.

Reason, C. J. C. and M. Rouault, 2005: Links between the Antarctic Oscillation and winter rainfall over western South Africa. *Geophys. Res. Lett.*, **32**, L07 705, doi:10.1029/2005GL022419.

Roff, G., D. W. J. Thompson, and H. Hendon, 2011: Does increasing model stratospheric resolution improve extended-range forecast skill? *Geophys. Res. Lett.*, **38**, L05 809, doi:10.1029/2010GL046515.

Roscoe, H. K., J. D. Shanklin, and S. R. Colwell, 2005: Has the Antarctic vortex split before 2002? *J. Atmos. Sci.*, **62**, 581–588, doi:10.1175/JAS-3331.1.

Salby, M. L., E. A. Titova, and L. Deschamps, 2012: Changes of the Antarctic ozone hole: Controlling mechanisms, seasonal predictability, and evolution. *J. Geophys. Res.*, **117**, D10 111, doi:10.1029/2011JD016285.

Scaife, A. A., D. R. Jackson, R. Swinbank, N. Butchart, H. E. Thornton, M. Keil, and L. Henderson, 2005: Stratospheric vacillations and the major warming over Antarctica in 2002. *J. Atmos. Sci.*, **62**, 629–639, doi:10.1175/JAS-3334.1.

Scaife, A. A., et al., 2011: Improved Atlantic winter blocking in a climate model. *Geophys. Res. Lett.*, **38 (23)**, doi:10.1029/2011GL049573.

Scaife, A. A., et al., 2014: Skillful Long Range Prediction of European and North American Winters. *Geophys. Res. Lett.*, **in press**, doi:10.1002/2014GL059637.

Seviour, W. J. M., N. Butchart, and S. C. Hardiman, 2012: The Brewer-Dobson circulation inferred from ERA-Interim. *Q. J. R. Meteorol. Soc.*, **138**, 878–888, doi:10.1002/qj.966.

Sigmond, M., J. F. Scinocca, V. V. Kharin, and T. G. Shepherd, 2013: Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nat. Geosci.*, **6**, 98–102, doi:10.1038/NGEO1698.

Silvestri, G. E. and C. Vera, 2003: Antarctic Oscillation signal on precipitation anomalies over southeastern South America. *Geophys. Res. Lett.*, **30**, 2115, doi:10.1029/2003GL018277.

Simpson, I. R., P. Hitchcock, T. G. Shepherd, and J. F. Scinocca, 2011: Stratospheric variability and tropospheric annular-mode timescales. *Geophys. Res. Lett.*, **38**, L20 806, doi:10.1029/2011GL049304.

Smith, D. M., A. A. Scaife, and B. P. Kirtman, 2012: What is the current state of scientific knowledge with regard to seasonal and decadal forecasting? *Environ. Res. Lett.*, **7**, 015 602, doi:10.1088/1748-9326/7/1/015602.

Son, S.-W., A. Purich, H. H. Hendon, B.-M. Kim, and L. M. Polvani, 2013: Improved seasonal forecast using ozone hole variability? *Geophys. Res. Lett.*, **40**, 6231–6235, doi:10.1002/2013GL057731.

Tennant, W. J., G. J. Shutts, A. Arribas, and S. A. Thompson, 2011: Using a Stochastic Kinetic Energy Backscatter Scheme to Improve MOGREPS Probabilistic Forecast Skill. *Mon. Weather Rev.*, **139**, 1190–1206, doi:10.1175/2010MWR3430.1.

Thompson, D. W. J., M. P. Baldwin, and S. Solomon, 2005: Stratosphere-troposphere coupling in the Southern Hemisphere. *J. Atmos. Sci.*, 708–715, doi:10.1175/JAS-3321.1.

Thompson, D. W. J. and S. Solomon, 2002: Interpretation of recent Southern Hemisphere climate change. *Science*, **296**, 895–899, doi:10.1126/science.1069270.

Waugh, D. W., L. Oman, P. A. Newman, R. S. Stolarski, S. Pawson, J. E. Nielsen, and J. Perlwitz, 2009: Effect of zonal asymmetries in stratospheric ozone on simulated Southern Hemisphere climate trends. *Geophys. Res. Lett.*, **36**, L18 701, doi:10.1029/2009GL040419.

Waugh, D. W., et al., 1994: Transport out of the lower stratospheric Arctic vortex by Rossby wave breaking. *J. Geophys. Res.*, **99**, 1071–1088.

Weber, M., S. Dikty, J. P. Burrows, H. Garny, M. Dameris, A. Kubin, J. Abalichin, and U. Langematz, 2011: The Brewer-Dobson circulation and total ozone from seasonal to decadal time scales. *Atmos. Chem. Phys.*, **11**, 11 221–11 235, doi:10.5194/acp-11-11221-2011.

Wilcox, L. J., B. J. Hoskins, and K. P. Shine, 2012: A global blended tropopause based on ERA data. Part I: Climatology. *Q. J. R. Meteorol. Soc.*, **138**, 561–575, doi:10.1002/qj.951.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2d ed., Academic Press, 627 pp.

WMO, 2011: Scientific Assessment of Ozone Depletion: 2010. Tech. rep., World Meteorological Organization, Global Ozone Reasearch and Monitoring Project, Report 52., 438 pp., Geneva.

Yamazaki, K., 1987: Observations of the Stratospheric Final Warmings in the Two Hemispheres. *J. Meteor. Soc. Japan*, **65**, 51–66.

List of Figures

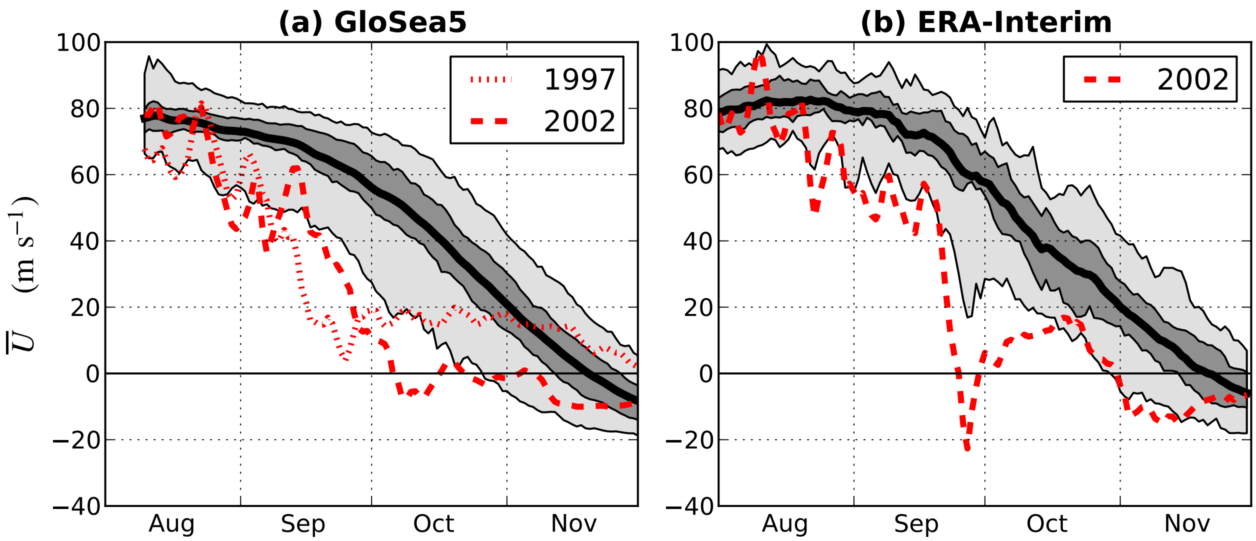

FIG. 1. Time series of daily 10 hPa zonal-mean zonal wind ($\overline{U}$) at 60°S for all GloSea5 ensemble members from 1996–2009 (a) and ERA-Interim from 1979–2010 (b). The thick black line indicates the mean, dark gray shading the interquartile range and light gray the 95th percentile range. Individual time series of the two ensemble members of GloSea5 which simulate an SSW (one for 1997 and one for 2002), and the year with an observed SSW (2002) are shown in red.

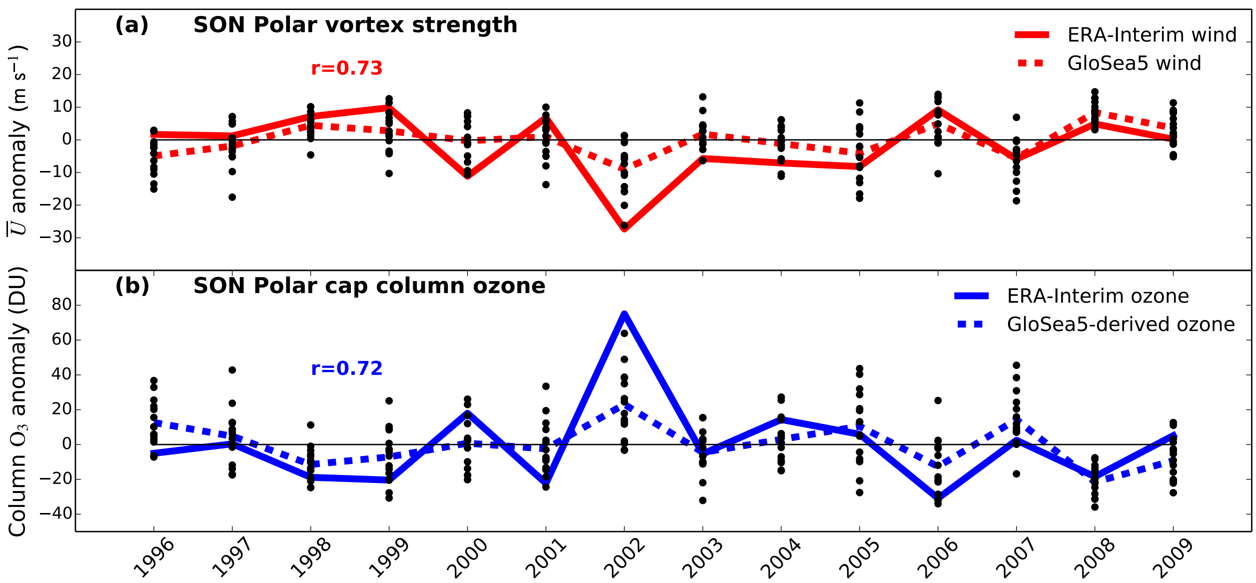

Fig. 2. (a) SON mean anomalies at 10 hPa and 60°S in ERA-Interim and the GloSea5 hindcast ensemble mean. (b) SON mean polar cap averaged (60–90°S) total column ozone anomalies from ERA-Interim and those derived from the GloSea5 anomalies as described in the text. Individual ensemble members are shown as black dots. Hindcasts are initialized near 1st August.

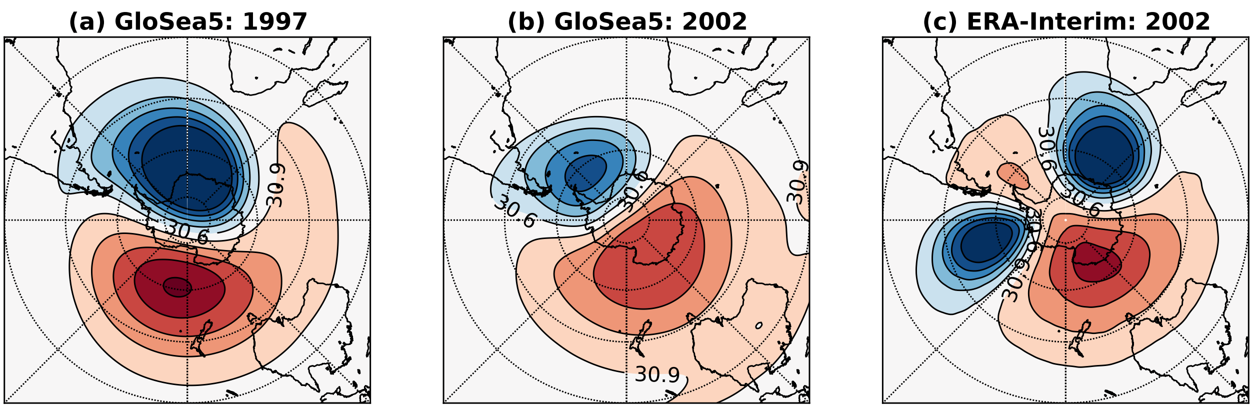

FIG. 3. Geopotential height at 10 hPa on the date at which $\overline{U}$ at 60°S and 10 hPa is at its minimum value, for the two GloSea5 ensemble members which simulate a SSW (a,b), and for ERA-Interim at the central date of the 2002 SSW (c). Units are km and the contour interval is 0.3 km.

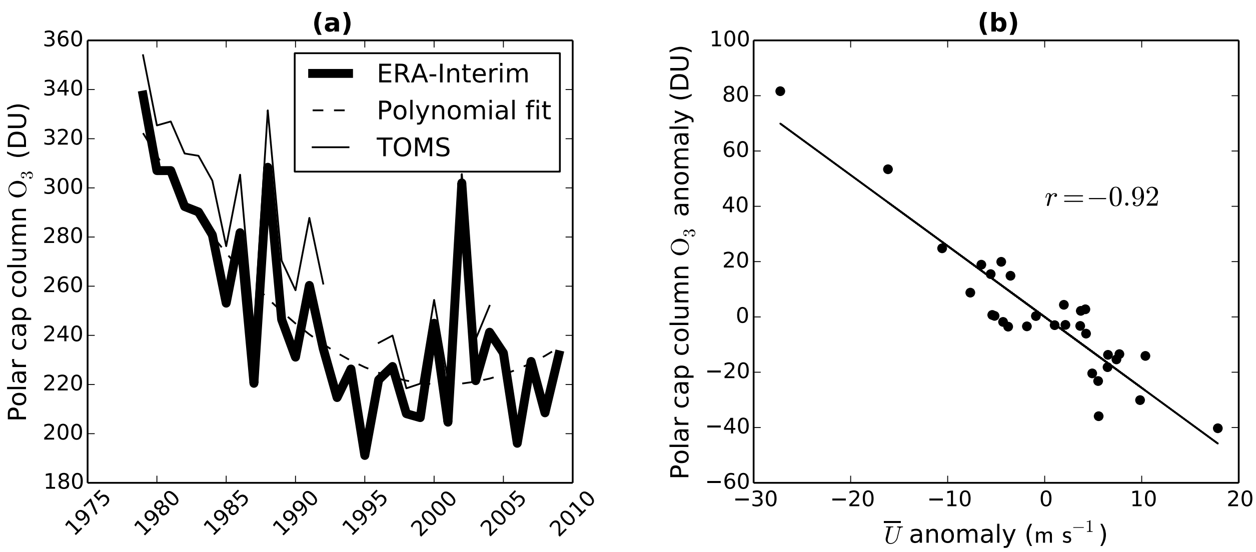

FIG. 4. (a) Time series of SON mean polar cap averaged (60-90°S) total column ozone in ERA-Interim and the TOMS satellite instrument. The ERA-Interim data are fitted with a 2nd-order polynomial. (b) Anomalies of ERA-Interim column ozone from the polynomial fit plotted against SON mean anomalies at 10 hPa and 60°S for each year from 1979–2009.

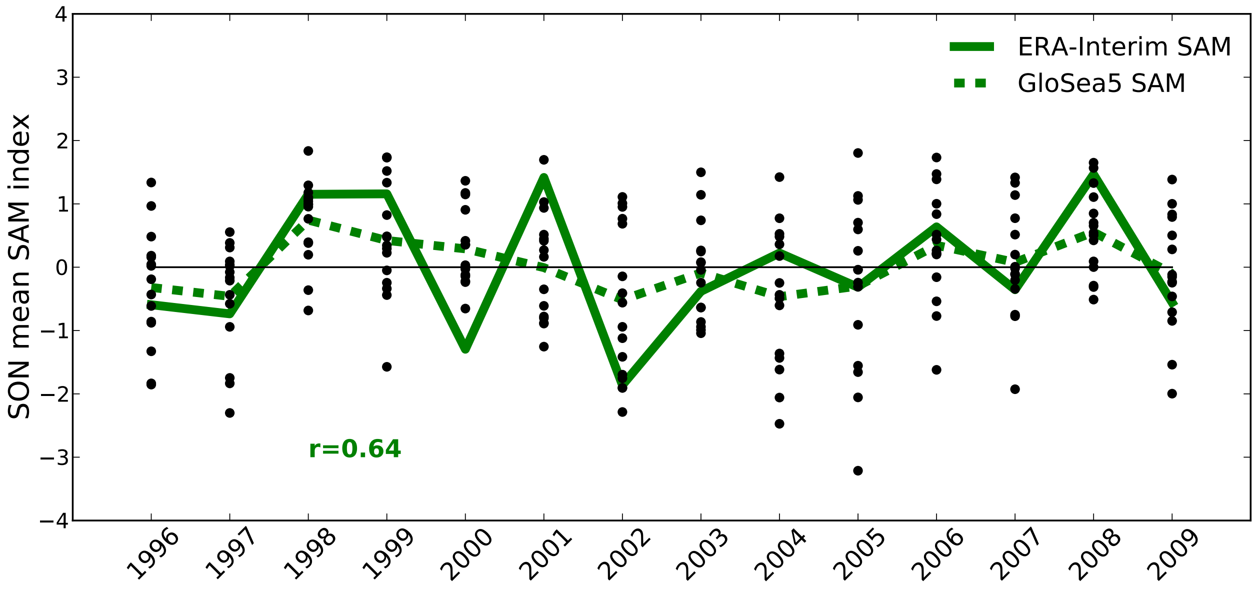

FIG. 5. SON mean Southern Annular Mode (SAM) index in individual GloSea5 hindcast ensemble members (dots), ensemble mean (dashed green curve) and ERA-Interim (solid green curve). The SAM is calculated from mean sea-level pressure data, and hindcasts initialized near 1st August. The correlation of the ensemble mean and ERA-Interim values is 0.64, which is statistically significant at the 99% level.

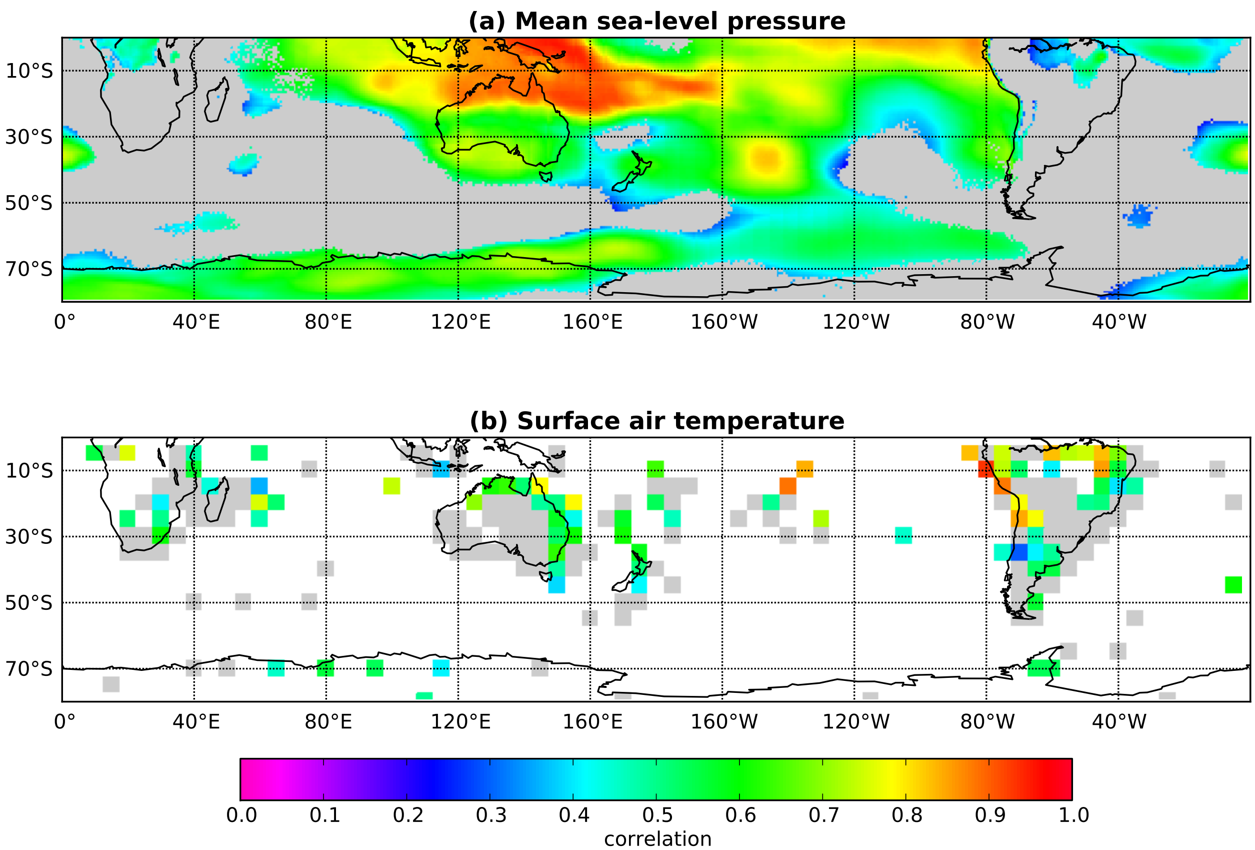

Fig. 6. (a) Correlation of SON mean sea-level pressure in the GloSea5 ensemble mean with ERA-Interim values. (b) Correlation of SON mean near-surface air temperature in the GloSea5 ensemble mean and HadCRUT4 gridded station-based temperature data. Regions with no observations are white and gray shading indicates regions where the correlation is not greater than zero at the 95% confidence level, using a bootstrap test at each gridpoint.

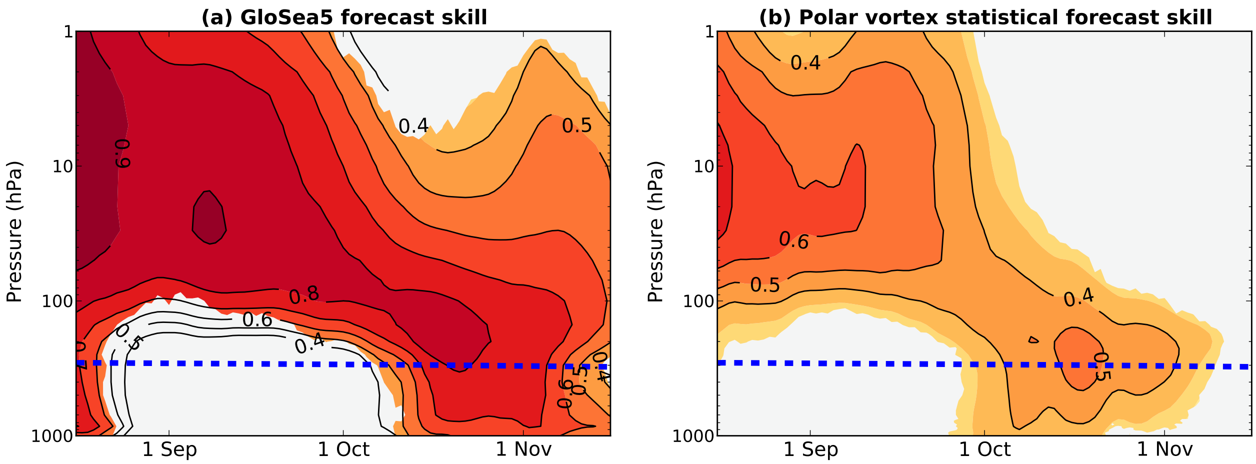

FIG. 7. (a) Correlation of GloSea5 ensemble mean polar cap (60-90°S) geopotential height anomalies (Z') with ERA-Interim values from 1996–2009, as a function of time and height. (b) Correlation of ERA-Interim from 1979–2010 values with those predicted by a linear statistical model based on Z' at 10 hPa on 1st August. All values are smoothed with a 30-day running mean before correlations are calculated. The contour interval is 0.1 and all colored regions are greater than zero at the 95% confidence interval, using a bootstrap test at each time at height. The blue dashed line indicates the approximate polar cap mean tropopause level (Wilcox et al. 2012).

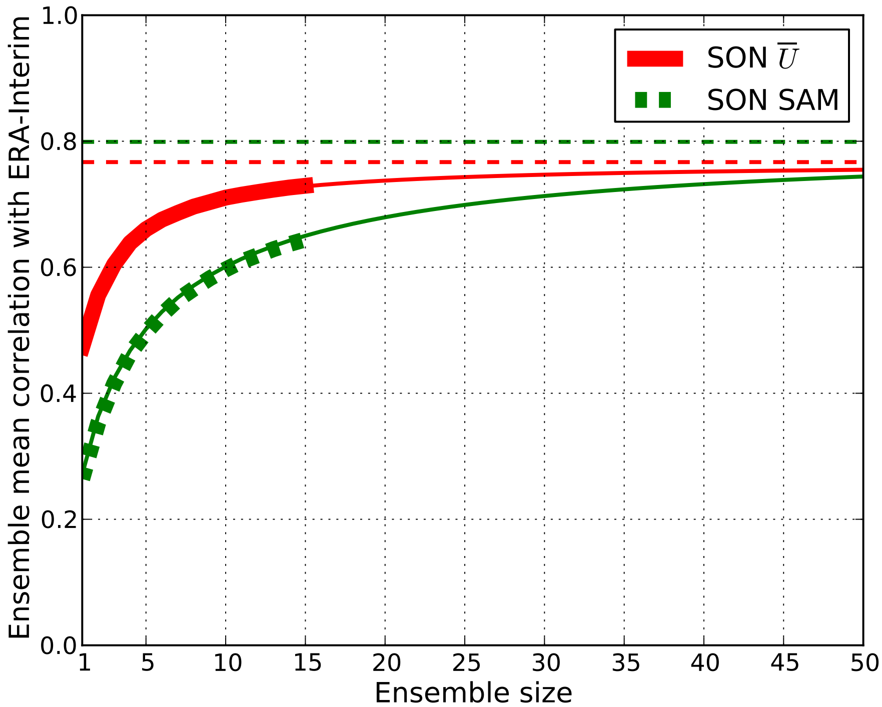

Fig. 8. GloSea5 ensemble mean correlation with ERA-Interim as a function of ensemble size for the SON mean $\overline{U}$ at 10 hPa and 60°S and SON mean SAM (thick lines). A theoretical estimate of the variation of correlation with ensemble size is shown in each case (thin solid lines), along with its asymptote for an infinite sized ensemble (dashed lines)