JGR Atmospheres

RESEARCH ARTICLE
10.1029/2018JD029626

Special Section:
Bridging Weather and Climate: Subseasonal-to-Seasonal (S2S) Prediction

Prediction of Northern Hemisphere Regional Surface Temperatures Using Stratospheric Ozone Information

Kane A. Stone[1], Susan Solomon[1], Douglas E. Kinnison[2], Cory F. Baggett[3], and Elizabeth A. Barnes[3]

[1]Department of Earth, Atmospheric, and Planetary Science, Massachusetts Institute of Technology, Cambridge, MA, USA, [2]Atmospheric Chemistry Observations and Modeling Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [3]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA

Abstract Correlations between springtime stratospheric ozone extremes and subsequent surface temperatures have been previously reported for both models and observations at particular locations in the Northern Hemisphere. Here we quantify for the first time the potential use of ozone information for Northern Hemisphere seasonal forecasts, using observations and a nine-member chemistry climate model ensemble. The ensemble composite correlations between March total column ozone (TCO) and April surface temperatures display a similar structure to observations, but with slightly lower correlation magnitudes. This is likely due to the larger number of cases smoothing out sampling error in the pattern, which is visible in the difference between correlations calculated from individual ensemble members. Using a linear regression model with March TCO as the predictor, predictions of the following April surface temperatures in regions that show large correlations are possible up to 4 years following the regression model end date in individual ensemble members, and up to 6 years in observations. We create an empirical forecast model to predict the sign of the observed as well as the modeled surface temperature anomalies using March TCO. Through a leave-three-years-out cross-validation method, we show that March TCO can forecast the sign of the April surface temperature anomalies well in parts of Eurasia that show the lowest model internal variability.

Plain Language Summary There are known linkages between springtime Arctic ozone in the upper atmosphere and Northern Hemisphere regional surface temperatures in the following months. Here, using both a state-of-the-art climate model and observations, we investigate whether these linkages can be useful for seasonal forecasting of regional surface temperatures in the Northern Hemisphere. Not only do we find good agreement between the model and the observations, using a predictive model, we show that March ozone can forecast the sign of the April surface temperature anomaly well in parts of Eurasia.

Key Points:
- Excellent agreement between modeled and observed correlations of March Arctic ozone and April surface temperatures
- Variability in local spatial locations of correlations between ensemble members, but large scale agreement
- Using extreme values of March ozone as a predictor, skillful forecasts of April Northern Hemisphere surface temperatures are attainable

Supporting Information:
- Supporting Information S1

Correspondence to:
K. A. Stone,
stonek@mit.edu

Citation:
Stone, K. A., Solomon, S., Kinnison, D. E., Baggett, C. F., & Barnes, E. A. (2019). Prediction of Northern Hemisphere regional surface temperatures using stratospheric ozone information. *Journal of Geophysical Research: Atmospheres*, *124*. https://doi.org/10.1029/2018JD029626

Received 7 SEP 2018
Accepted 11 MAR 2019
Accepted article online 29 MAR 2019

1. Introduction

Potential connections of polar stratospheric ozone changes to surface climate have been an area of interest over recent decades, first sparked by the observation that stratospheric circulation anomalies appear to propagate down to the surface over a time scale of weeks (Baldwin & Dunkerton, 2001). Springtime Antarctic ozone depletion has resulted in a stronger and colder polar vortex, which in turn has driven changes in Southern Hemisphere summer weather patterns (Kang et al., 2011; Thompson & Solomon, 2002). In the Northern Hemisphere (NH), springtime polar stratospheric ozone extremes display large but localized surface temperature differences in both models (Calvo et al., 2015) and observations (Ivy et al., 2017), particularly in Siberia. In contrast to the Southern Hemisphere, whether Arctic stratospheric ozone changes are a harbinger or a cause of subsequent tropospheric climate change has not been robustly established. One recent study has suggested that cold winter conditions in Siberia are linked to stratospheric anomalies stemming from abnormal planetary wave activity from Barents-Kara sea ice loss in the previous autumn (Zhang et al., 2018). Additionally, it has been suggested that the Arctic ozone anomalies that coincide with the stratospheric temperature circulation anomalies have a positive feedback effect, enhancing the stratospheric-tropospheric link (Haase & Matthes, 2018). Full analysis of the role of Arctic ozone on spring surface climate is a complex problem and is not the subject of the present paper. However, while the science regarding stratosphere-troposphere coupling is becoming more robust, there has been recent increased interest regarding subseasonal to seasonal forecasting (Robertson et al., 2015). Therefore, our focus is on evaluation of

whether knowledge of early spring ozone concentrations could be beneficial for seasonal forecasting. Total column ozone (TCO) abundances are routinely measured on a global basis to very high precision by satellite methods (WMO, 2014), motivating our effort to evaluate their potential for seasonal prediction.

Previous studies have investigated ozone's role in seasonal forecasting during the 2011 springtime extreme low Arctic ozone (Manney et al., 2011). One study found no model-based improvement in extended range forecasting of up to 31 days when using a representation of the 2011 ozone anomaly based on satellite data within the Met Office Global and Regional Ensemble Prediction System (Cheung et al., 2014). In contrast, another study found that both a cold Arctic stratosphere and tropospheric temperature anomalies associated with sea surface temperatures contributed to the state of the 2011 tropospheric spring climate within the European Centre/Hamburg version 5 atmospheric general circulation model (Karpechko et al., 2014). It has also been shown that springtime extratropical prediction of land temperature is improved using realistic stratospheric initialization. Further, this predictability arises through the stratosphere's influence on the Northern Annular Mode (Jia et al., 2017). To the best of our knowledge, a robust assessment of surface temperature forecast skill using NH ozone information for multiple years has not been done before. We exploit a nine-member ensemble of simulations with a state-of-the-art chemistry climate model for this purpose. In particular, we use the Whole Atmosphere Community Climate Model (WACCM4) module of the Community Earth System Model, version 1 (CESM1; see section 3). The fully coupled free running atmosphere-ocean chemistry climate general circulation setup is used to produce the nine ensemble members over 1995–2024 (a total of 270 years), a time period representative of high chlorine (Engel et al., 2018), chosen to maximize the chance of larger Arctic springtime ozone depletion and indicate potential utility for using stratospheric Arctic ozone for seasonal forecasting in coming years. Model results are compared to the observed TCO and ERA-interim reanalysis surface temperature data (hereafter referred to as observations; see section 2). The El Niño Southern Oscillation (ENSO) has been accounted for in this work by the removal of the temperature anomalies correlated with the NINO 3.4 index from the surface temperature; see section 3 , this accounts for the ENSO influenced variability in surface air temperatures, especially in southern Asia (Thirumalai et al., 2017). For the observations, we use the time period of 1980–2016 to obtain the maximum available observed high and low extreme ozone years (Ivy et al., 2017).

Here, we compare observed and modeled linear correlations and surface temperature differences corresponding to the upper and lower twentieth percentiles of March TCO averaged over 63–90°N and April surface temperatures. This is followed by using a linear regression model to predict April surface temperatures over consecutive years using March TCO. A conditional empirical model is then used through a leave-three-out cross-validation method to predict the sign of the surface temperature anomaly using March ozone (see section 3).

2. Data

2.1. Chemistry-Climate Model

The model runs presented here are from The CESM1. It is a fully coupled climate model that incorporates atmosphere, ocean, land, and sea ice modules (Garcia et al., 2017; Marsh et al., 2013). The atmosphere module is the WACCM4 that has a horizontal resolution of 1.9° latitude by 2.5° longitude and up to 66 vertical levels with a high top at 5.1×10^{-6} hPa (~140 km). The chemical scheme used is the Model of OZone And Related Tracers (Kinnison et al., 2007) and includes 183 different species, 341 gas phase reactions, 114 photolytic processes, and 17 heterogeneous reactions on multiple aerosol types. This chemical scheme has been shown to accurately simulate polar ozone depletion and levels of chlorine reservoir species (Solomon et al., 2015, 2016). This setup has a repeated cyclic 28 month quasi-biennial oscillation, no solar cycle or solar proton events, and updated sulfate area densities including 21st century volcanic eruptions (Mills et al., 2016; Neely & Schmidt, 2016). The model was spun up from a REF-C2 Chemistry Climate Model Initialization simulation commencing in 1955. Initialization of the model ensemble members commences from slightly differing initial air temperatures in 1995 (Kay et al., 2015; Solomon et al., 2017).

2.2. Ozone Data

TCO observations are sourced from the National Institute for Water and Atmosphere-Bodeker Scientific TCO database V3.2 patched data set. This data set combines multiple satellite observations and accounts

for drifts and offsets using ground-based Brewer and Dobson instruments and spans 1980–2016 (Struthers et al., 2009).

2.3. ERA-Interim Reanalysis

ERA-Interim reanalysis data, from the European Centre for Medium-Range Weather Forecasts, are used for the "observed" surface temperature data. Observations in conjunction with a forecast model are used to create the data set (Dee et al., 2011), which spans the period of 1979 to present. We use monthly mean of daily mean ERA-Interim 2-m temperature data with a resolution of 1.5° latitude by 1.5° longitude over the period of 1980–2016.

3. Methods

3.1. Linear Correlation and Surface Temperature Differences

Linear correlations of observed March TCO averaged over 63–90°N and April surface temperatures are constructed and shown in sections 4.1 and 4.2. The data sets were preconditioned by removing the linear trends from TCO, and the linear trends and ENSO from the surface temperatures. ENSO is defined as the NINO 3.4 index, which is described as the sea surface temperature anomaly in the region of 5–5°S and 120–170°W. Removal of the linear trends and ENSO is performed by removing the linear least squares regression fits of a linear and ENSO regression function separately over the entire time periods of the data sets following

$$\boldsymbol{T}^{*}(\boldsymbol{t})=\boldsymbol{T}(\boldsymbol{t})\text{-}\boldsymbol{\beta}\boldsymbol{x}(\boldsymbol{t}), \tag{1}$$

where $\boldsymbol{T}(\boldsymbol{t})$ is the surface temperature at time $\boldsymbol{t}$, $\boldsymbol{T}^{*}(\boldsymbol{t})$ is the modified surface temperature, and β is the regression coefficient associated with regression function $\boldsymbol{x}(\boldsymbol{t})$. Additionally, a lag is likely associated with the ENSO influence on surface temperature variability. Therefore, the ENSO regression coefficients are calculated with a lag of up to 2 months from the surface temperature month being evaluated at each grid point. The lag that is associated with the largest regression coefficient at each grid point is used to remove the ENSO variability from the surface temperatures. This results in $\boldsymbol{x}(\boldsymbol{t})$ being defined as $\boldsymbol{x}(\boldsymbol{t}-L)$ for ENSO, where L is the lag time chosen. Removal of the linear trend and ENSO acts to reduce the influence from greenhouse gas increases and ENSO variability. It has been shown previously that there is both a tropospheric and stratospheric (through influence of SSWs) ENSO pathway on surface temperature, especially over the Eurasian region (Butler et al., 2014). Therefore, through this method, the removal of ENSO is likely different depending on how much each region is influenced by the tropospheric or stratospheric pathways. However, we note that the linear correlations only change slightly depending on whether ENSO is removed or not (not shown). After preconditioning, the Pearson linear correlation is calculated between March 63–90°N mean TCO and April surface temperatures.

The modeled composite of surface temperature differences is constructed from the data points that correspond to the March 63–90°N TCO upper twentieth percentiles minus the lower twentieth percentiles for each model individually. This ensures a robust assessment of the internal variability across all ensemble members. The linear trend and ENSO are removed similarly to the method for the linear correlations.

3.2. Regression Model

We use a linear least squares regression model with mean March 63–90°N TCO as the sole predictor of the April surface temperature predictands in section 4.3. Preconditioning of the TCO predictor and surface temperature predictands is similar to that performed in the linear correlations and surface temperature differences. However, only data from the training periods, 1980–2010 for observations and 1995–2020 for the model, are used to remove the linear trend and ENSO, following equation (1). This ensures that no forecasted data is used in construction of the regression model. The surface temperature predictands are also normalized to the mean of the training periods to produce the surface temperature anomaly. Using the regression model, modeled surface temperatures are fit to the training periods and forecasts are produced for the remaining years. After the forecast has been made, ENSO is recalculated over the entire time period and is used in conjunction with the training period ENSO regression coefficients to remove the ENSO variability in the forecast time period. The means and linear trends of the training period are also removed from the forecast time period. It is noted that Arctic anthropogenic ozone depletion and recovery, even though small compared to the Southern Hemisphere, is likely not linear in nature. Therefore, preconditioning by

removing the linear trend may not be the best approach and could introduce biases into the future predictions.

3.3. Conditional Empirical Model

To predict the sign of the surface temperature anomaly, we construct a conditional empirical model based on the sign of the surface temperature difference from the March TCO upper twentieth percentiles minus the March TCO lower twentieth percentiles. The twentieth percentile data were extracted in a similar fashion to that of the surface temperature differences. If the sign of the surface temperature difference is positive for a given location, we state that a positive ozone anomaly will forecast a positive surface temperature anomaly. If the sign of the surface temperature difference is negative, we state that a positive ozone anomaly will forecast a negative surface temperature anomaly. This results in a simple binary prediction for all surface temperature grid points.

3.4. Leave-Three-Years-Out Heidke Skill Scores

Using the conditional empirical model, we construct Heidke skill scores (HSSs) based on a leave-three-years-out method (section 4.4). For the nine model ensemble members, 30 consecutive empirical models are produced after each three consecutive data points are left out. The remaining 27 data points are the training data sets. This results in 30 slightly different empirical models. The training data sets are preconditioned by removing the linear least-square trends and the influence of ENSO through linear regression following equation (1). This is performed in a similarly manner to the regression model methods, and ensures that no forecasted data are used in construction of the empirical models, and is repeated for each consecutive leave-three-year-out case. Considering we are aiming to predict the variability of surface temperatures due to ozone and not the linear trend or ENSO, by removing these quantities, we predict the sign of the surface temperature anomaly about the climatological linear trend line, and in the absence of ENSO. After preconditioning of the training data sets, the empirical model is constructed and used to forecast only the middle year point from the three that are left out. This is to account for autocorrelation. After the forecast has been made, ENSO is recalculated for the forecast year, and is removed from the surface temperature data that was forecasted using the training period regression coefficients. Similarly, for the linear trend. The same approach is done for the observations. For the nine-member ensemble, when all data are used, a HSS is calculated from the 270 forecasts. This is defined as

$$\boldsymbol{HSS} = \frac{\boldsymbol{H-E}}{\boldsymbol{T-E}} \times \mathbf{100}, \tag{2}$$

where H is the number of correct predictions, T is the number of forecasts (in this case, 270), and E is the number of correct forecasts by chance. Since we are forecasting the median anomaly, E is equivalent to T/2. The HSS values range from −100 to 100. A HSS of 0 can be interpreted as an equal number of correct and incorrect forecasts. A positive HSS means the model has provided more correct forecasts than incorrect forecasts, for example, a HSS of 50 indicates 3 correct forecasts out of 4. Whereas, a negative HSS indicates more incorrect forecasts than correct.

3.5. Calculation of Significance

Significance calculated in the linear correlations is based on rejection of the null hypothesis that there is nonzero correlation compared to a zero correlation and is shown at the 95th percentile. Significance in the surface temperature difference between the March TCO upper and lower 20th percentiles is based on a two-tailed student *t* test at the 95th percentile. Significance of the HSSs are constructed after the model predictions are bootstrapped 500 times, the HSSs are recalculated, and the 95th percentile of the bootstrapped samples is determined. Significance is shown where the HSSs are greater than the 95th percentile or as 95th percentile error bars.

4. Results

4.1. Polar Ozone and Surface Temperature Correlations

Linear correlations of observed March TCO averaged over 63–90°N and April surface temperatures are shown in Figure 1a (see section 3). The correlations between TCO and subsequent surface temperatures approach −0.45 in central Russia and +0.45 across southern Asia, especially in the Himalayas and

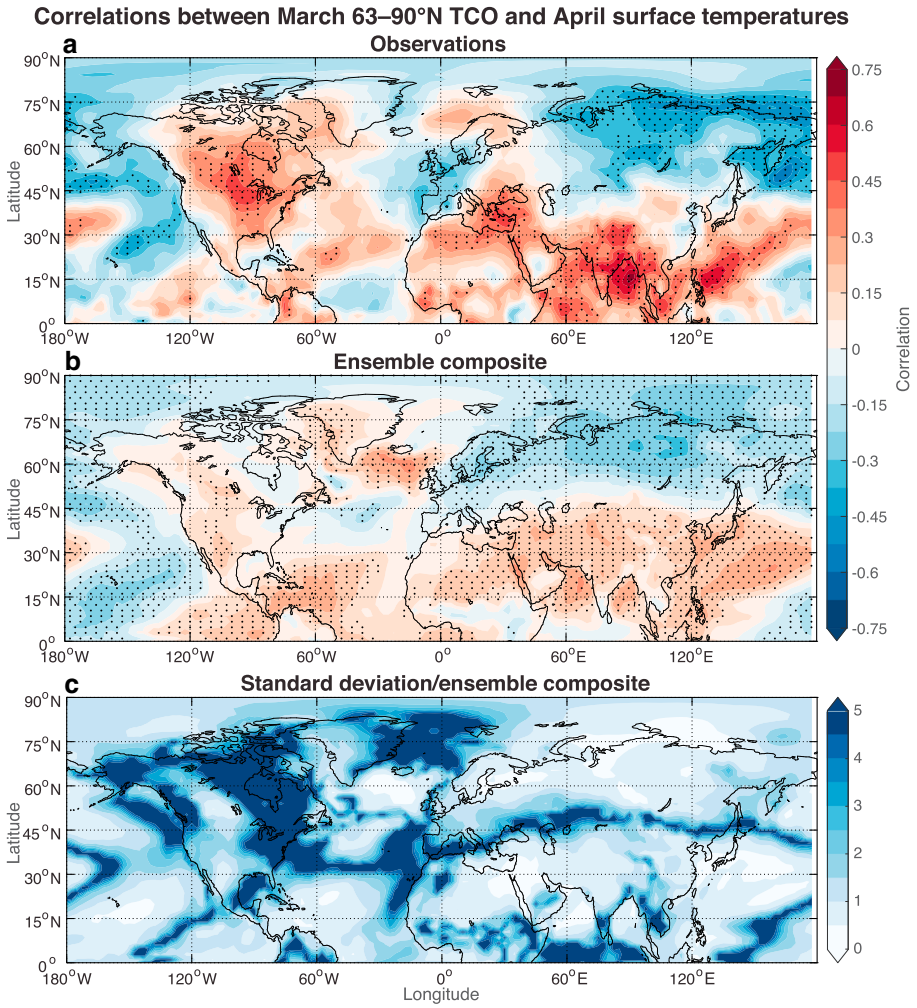

Figure 1. Correlations of March 63–90°N TCO and April surface temperatures for observations over 1980–2016 (a), and for the ensemble composite over 1995–2024 (b). Stippling shows significance to the 95th percentile. The ensemble standard deviation divided by the ensemble mean, with lower values indicating more consistent correlations among members are also shown (c). TCO = total column ozone.

southeastern Asia. High correlations are also seen in parts of the eastern North Pacific, western Europe, and North America. A dipole pattern is seen across Eurasia, with negative correlations between ozone and surface temperature in Russia and western Europe, and positive correlations from northern Africa to southeastern Asia; this indicates a shift consistent with Northern Annular Mode variability (Thompson & Wallace, 2001), associated with TCO values, in agreement with Jia et al. (2017). March TCO-May surface temperature correlations are shown in Figure S1, and indicate a broadly similar structure.

The model ensemble composite correlations are shown in Figure 1b. The spatial structure of the correlations obtained in the observations is well reproduced in the model, including the sharp dividing line across Eurasia. Figure 1 thus shows that the ensemble reproduces the observed connection between TCO and the tropospheric circulation. However, the magnitude of the correlations is lower in the model composite than in the observations. As explained in detail the next paragraph, this is likely because each individual member displays slightly different spatial patterns, smearing the correlation magnitudes somewhat.

To examine this in more detail, Figure 1c shows a metric of consistency in the correlations across the ensemble members. This metric is defined as the standard deviation of the ensemble member correlations divided by the mean of the ensemble member correlations and gives low values for consistent correlations across members, and high values for large differences across members. This figure highlights the large variability between individual ensemble member circulation patterns and their correlations with TCO changes,

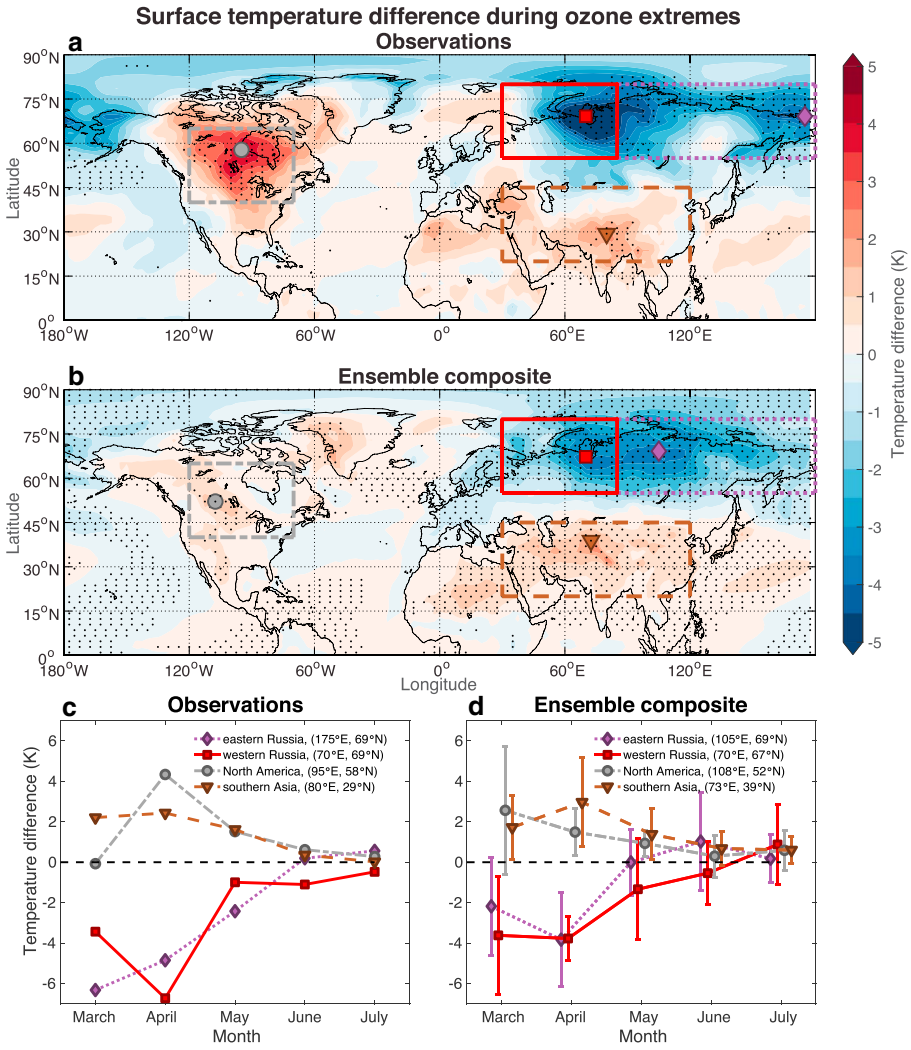

Figure 2. April surface temperature differences corresponding to the upper minus the lower twentieth percentile 63–90°N total column ozone March values over 1980–2016 (a). The equivalent model ensemble surface temperature differences over 1995–2024 (b). Temperature differences corresponding to the March total column ozone extremes at selected locations within each box are also shown as a function of months March through July (c and d). The error bars show plus and minus one ensemble standard deviation.

especially across North America; this is somewhat due to internal decadal variability among the individual ensemble members. Where some of the variability is linked to the transition between positive and negative correlation, so that slight shifts in circulation lead to low correlations in these transition regions. However, we also expect to see some change between ensemble members due to sampling uncertainty, and these regions of high correlation variability indicate where the ensemble signal is too weak to emerge outside of this uncertainty. This is further highlighted by Figure S3, which shows that the correlation magnitudes of individual ensemble members are similar to that of the observations, and that the resolved model signal exceeding the sampling uncertainty emerges in the ensemble mean with its far greater sampling (270 total years). This reveals that the most robust regions (less influenced by internal variability and large enough to be detected outside of sampling uncertainty) are within central Russia and southern Asia. Using this information, we can identify consistent regions across ensemble members and investigate their behavior in more detail.

4.2. Surface Temperature Differences During Ozone Extremes

Figure 2 shows April surface temperature differences corresponding to the March TCO upper and lower twentieth percentiles (hereafter referred to as ozone extremes) for both (a) observations and the (b) model

ensemble composite (see section 3). The values of the mean upper and lower twentieth percentiles in TCO are 452 DU and 366 DU, respectively, in the observations and 466 and 390, respectively, in the model composite, showing that the values between the model and observations agree reasonably well. The spatial pattern of the temperature differences for the extreme ozone cases is very similar to the correlation pattern shown in Figures 1a and 1b. The largest differences occur in central Russia, southern Asia, and North America in the observations. This is replicated in the model ensemble composite, with the exception of North America. In sum, the model captures the behavior as observed with a high degree of fidelity in regions where the surface temperature differences are greatest and where the model has a well-resolved signal. (The global March TCO and May surface temperature differences for observations and the model ensemble are also shown in Figure S2 and are consistent with Figure 2).

Figures 2c and 2d show the surface temperature difference from March ozone extremes over the March–July months independently for the locations marked in Figures 2a and 2b for observations and the model ensemble composite, respectively. The ensemble lines also show the ensemble standard deviation as error bars. Figure 2 demonstrates that for both the observations and the model ensemble composite, April surface temperature differences associated with March ozone extremes can exceed those of March surface temperature differences, indicating predictability. The temperature differences remain elevated through May, then begin to approach zero in both the observations and the ensemble composite. As mentioned previously, North America also displays large observed temperature differences in April, greater than 4 K, which are only weakly replicated in the model ensemble composite. However, this is also a region where the model does not have a well-defined signal (Figure 1c).

4.3. Predicting the Last Few Years Based on Earlier Records

In this section, we investigate the predictive power of March TCO for localized surface temperature anomalies in April. Using a linear regression model (see section 3), we present an analysis of forecasting the last few years in a time series.

A key finding from the previous sections is that the detailed correlations between March TCO and April surface temperature are spatially displaced somewhat between ensemble members due to model internal variability and sampling uncertainty (Figure 1c). However, regions in central Russia and southern Asia show much more consistent correlations among members, and also correspond to locations of very large observed correlations. We therefore examine how well the relationship between March TCO and April surface temperatures holds in these regions.

Figure 3 shows observations (a and b) and two individual ensemble member (c–f) regression model fits and predictions using the March TCO predictor for April surface temperatures. For locations of high correlation in central Russia and southern Asia, the regression model is fit to 1980–2010 for observations, and the remaining 2011–2016 years are predicted; similarly, the regression model is fit to 1995–2020 for the model, and the remaining 2021–2024 years are predicted. Correlation maps over 1980–2010 and 1995–2020 for observations and the model (respectively) are shown in Figure S3, highlighting the selected regions of high correlations

Figure 3a shows that the largest observed correlations in central Russia and southern Asia are −0.41, 0.50, respectively. For the years of 2011–2016, we see that the regression forecast tracks the subsequent observed data well for both central Russia and southern Asia in both periods shown here. Note that the large amount of ozone depletion in 2011, a very cold year in the stratosphere, has indeed resulted in very accurate local predictions for that year in these locations. The forecasts in central Russia and southern Asia suggest that the relationship is locally consistent in these regions 6 years beyond the training data set.

Figures 3c–3f show a similar consistency for model ensemble members 4 and 7. Additional ensemble members are shown in Figure S4. In ensemble member 4, the model forecast over years of 2021–2024 using the regression model training period of 1995–2020 tracks the sign of the changes in the surface temperature very well in both central Russia and southern Asia, although the absolute values are less well replicated in some cases. This could be due to preconditioning of the predictands (see section 3). The correlation values over 1995–2020 are −0.59 and 0.68 for central Russia and southern Asia, respectively. Ensemble member 7 presents similar results to ensemble member 4. We see high correlations in central Russia of −0.61 with

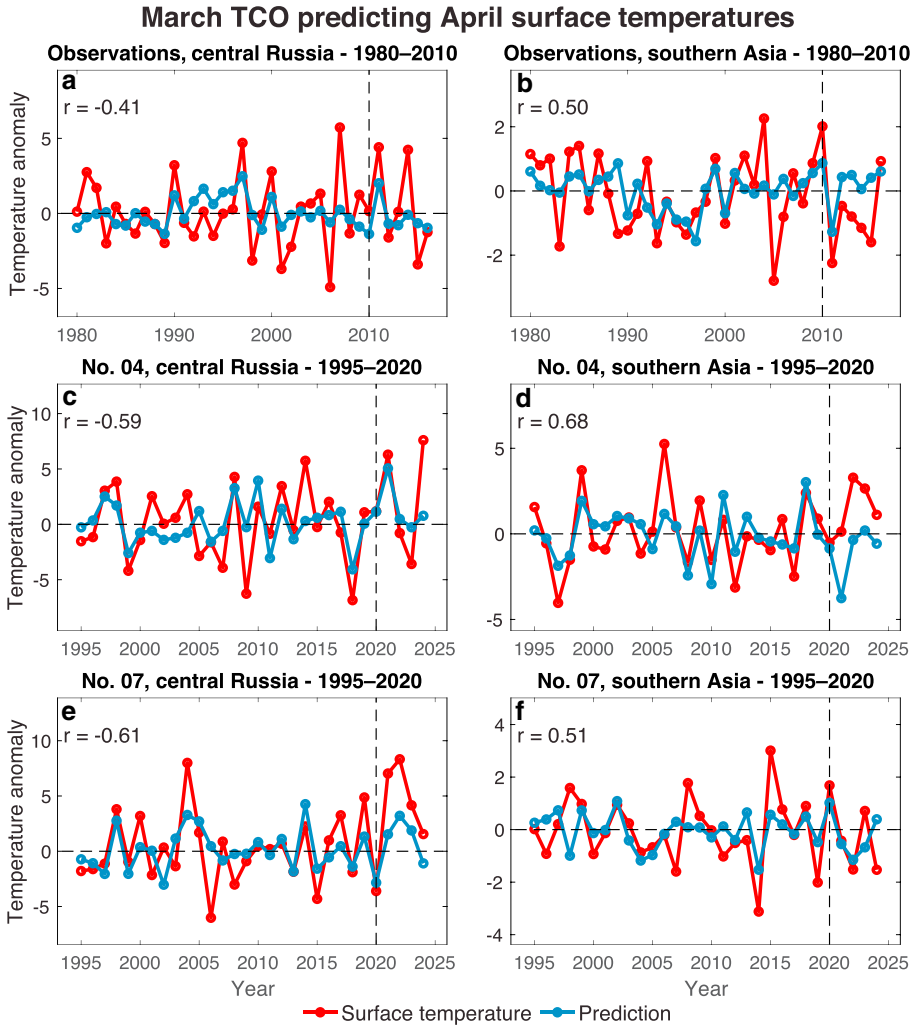

Figure 3. March 63–90°N TCO and regression model fits of April surface temperature at specific locations within central Russia and southern Asia for observations over 1980–2010 and the regression model forecasts over 2011–2016 (a and b). Similarly, for ensemble members 4 (c and d) and ensemble member 7 (e and f) with a regression model time period of 1995–2020 and forecasts over 2021–2024. The specific locations chosen are shown in Figure S3. TCO = total column ozone.

excellent forecasts over 2021–2024. In southern Asia, a correlation of 0.51 is seen with excellent forecasts for 2021 and 2022. However, there is less agreement in the forecasts in 2023 and 2024.

Figure 3 highlights within the locations of statistically useful relationships (e.g., high regression beta values), the predictions can remain consistent over a few years past the regression model end date within the regions that have a well-resolved signal in the model. Further, although prediction of the absolute values appears feasible in some cases (Figures 3c and 3f over 2021–2022), prediction of the sign of the anomaly may be more broadly applicable (Figure 3e), and represents one approach to testing skill (Mundhenk et al., 2017). The next section analyzes to what degree a simple empirical model based on March TCO can predict the sign of the surface temperature anomaly over regions where the model is not well resolved.

4.4. Leave-Three-Years-Out Anomaly Prediction

To analyze the use of March TCO for predicting the sign of localized surface temperature anomalies, we use a conditional empirical model based on the sign of the temperature change relating to the difference between the March ozone extremes. This model is used in a leave-three-years-out cross-validation method to construct a HSS (see section 3), which ranges from −100 (no correct forecasts) to 100 (all correct forecasts),

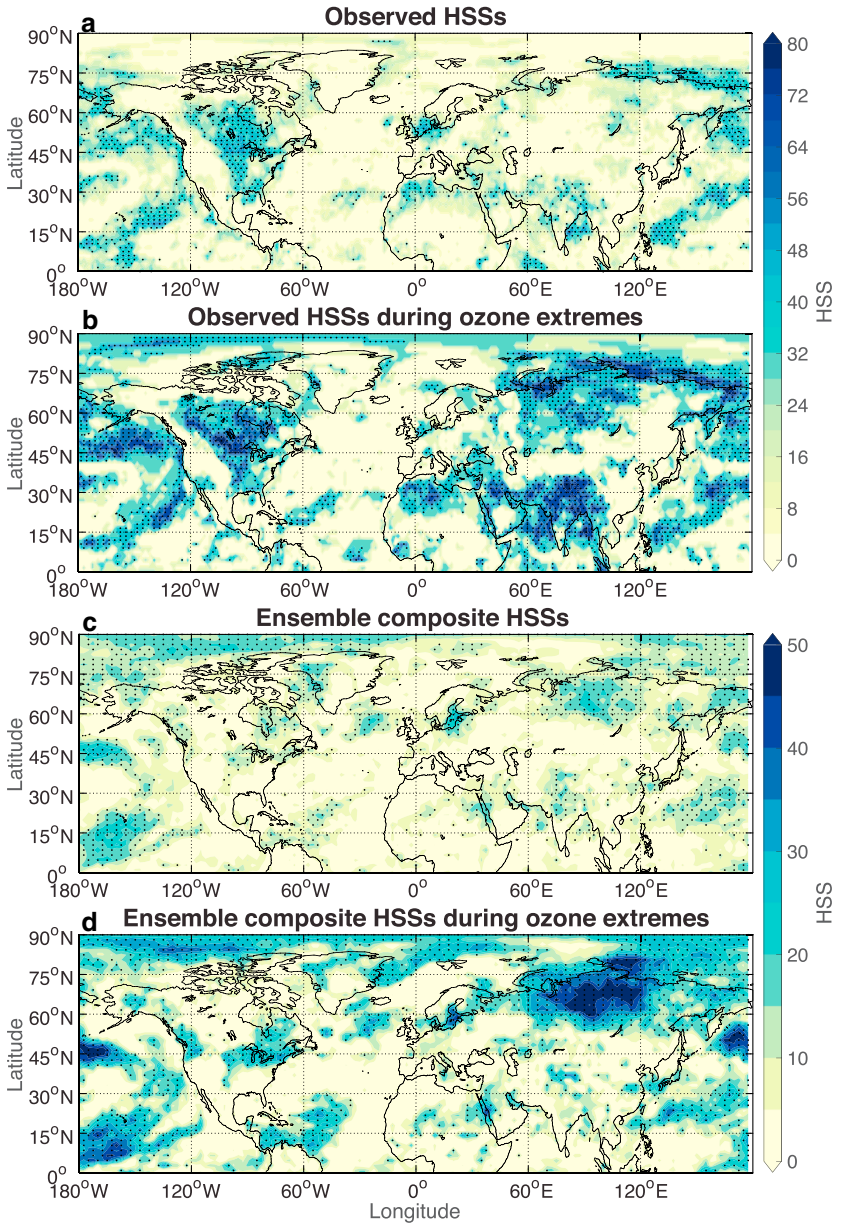

Figure 4. Heidke skill scores for prediction of the sign of the April surface temperature anomaly using March Arctic total column ozone after the leave-three-years-out cross-validation method. Observations are shown when using all data (a), and after extraction of the values corresponding to March ozone extremes (b). Similarly, for the ensemble composite (c), and ensemble composite during March ozone extremes (d). Significance, as stippling, is shown where the HSSs are greater than the 95th percentile based on 500 bootstrap samples. Note the different colorbar scales between the observations and the model. HSSs below zero are saturated to zero for clarity. HSS = Heidke skill score.

with 0 being equivalent to the reference forecast (50% correct forecasts; Johnson et al., 2014; Mundhenk et al., 2017).

Figure 4 shows the HSSs for March TCO predictors and April surface temperature predictands for both observations and the model data. The observed HSSs are shown in Figure 4a when using all data (37 years). Figure 4b shows results for the HSSs using only predictions corresponding to ozone extremes (14 years). The HSS pattern resembles that of the correlations and surface temperature differences in Figures 1 and 2. Skill scores of up to 40 are obtained in North America, central Russia, and southern Asia, and up to 60 in central Russia and 80 in southern Asia when only using the time periods associated with

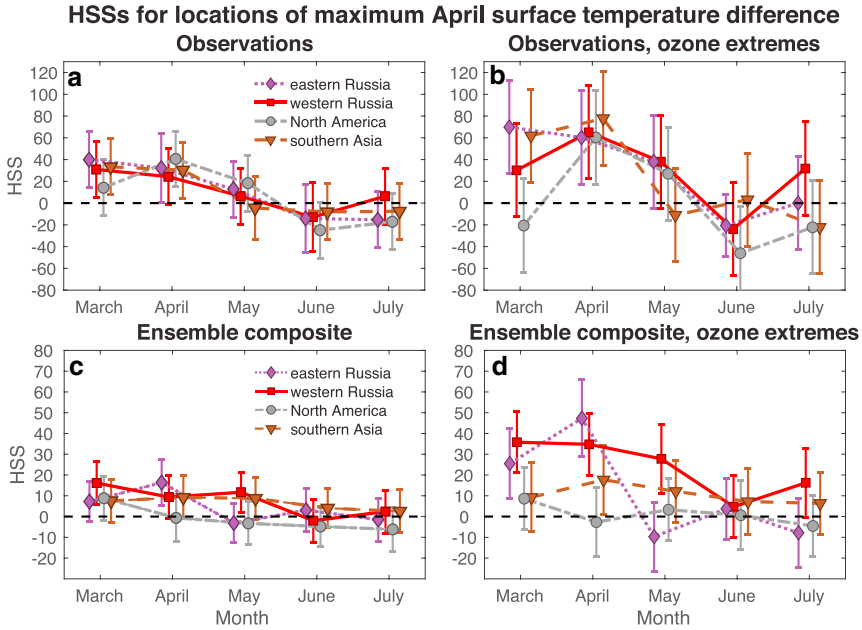

Figure 5. Observed HSSs for the locations that show the largest April surface temperatures differences based on March ozone extremes. This is shown for all observed data (37 data points—a) and data that corresponds to the March ozone extremes (14 data points—b). The April surface temperature differences corresponding to the ozone extremes are shown in Figure 2a. Similarly, for the model composite (c and d) with the locations used shown in Figure 2b. After each grid point was selected, the corresponding HSS score was calculated by averaging the grid point chosen and all surrounding grid boxes. Note the different *y* axis limits between observations and the model ensemble. HSS = Heidke skill score.

ozone extremes. This corresponds to up to 80% and 90% correct predictions of the sign of the surface temperature anomaly in these locations, respectively.

The model ensemble data are shown in a similar manner in Figures 4c and 4d, again displaying the largest values in the regions of highest correlations and surface temperature differences. For the ensemble composite when all data are considered (270 years), HSS values of over 20 are obtained in central Russia, and up to 20 in southern Asia and North America. For the ozone extreme years (108 years), these values increase to over 50 in central Russia (indicating that over 75% of the predictions are correct). There are also other regions that show large HSSs in both the observations and the model, for example, in southern Asia, in North America, the Arctic, East Pacific, North Pacific, and Caribbean, but these are less clear.

Next we further investigate the forecast lead HSSs over specific local regions. Figure 5 shows March to July observed and modeled HSSs for locations of largest April surface temperature difference corresponding to March ozone extremes, as shown in Figure 2. Further, to reduce the likelihood that results are isolated to a single grid box, the location that is selected also includes its surrounding grid boxes when calculating the average HSS. When using all 37 predictions in the observed HSS (Figure 5a), we obtain April HSSs of between 20 and 40 in the North American, central Russian, and southern Asian regions chosen. These values are also greater than the 95th percentiles constructed from a 500 sample bootstrap, with the exception of western Russia (see section 3). The HSSs are most elevated for March and April, before dropping off to 0 by May and June. When only data points that correspond to ozone extremes are used to construct the HSSs (Figure 5b), the April values for all locations are greater than 60. The Figure also indicates that the HSSs remain elevated through May before reducing to near 0 in June and July. Figures 5c and 5d similarly show the March to July HSSs for the ensemble composite when all data are used, and when only data corresponding to ozone extremes are used, respectively. Figure 5c shows that significant HSSs, approaching 20, occur in eastern Russia during April. Figure 5d also shows HSSs approaching 50 and 35 in eastern and western Russia, respectively, in the March ozone extreme years. We also see significant April HSSs in southern Asia. Additionally, in western Russia, the HSSs remain elevated through to May, similar to observations. Considering there is skill out to 2 months in advance of the ozone predictor, an interesting topic for future work would be to look at the comparison to a persistence forecast.

In sum, Figure 5 indicates that not only is there significant forecast skill in April in both observations and the model composite, but for western Russia, March ozone can be used to provide forecast skill 2 months into the future. We note that when ENSO is not removed from the surface temperature data, the modeled predictions over southern Asia reduce below the significant 95th percentile (not shown). However, there is virtually no change in the other regions. (A similar analysis for areas of largest May surface temperature differences are shown in Figures S6 and S7).

A major difference between the ensemble composite and the observations is the HSSs obtained in North America. In observations, significant HSSs of up to 40 are seen when all data are used. However, in the ensemble composite, the HSSs are 0 throughout. As shown previously (Figure 1c), this is an area where the model ensemble signal is not resolved, which ultimately results in limited predictability. This conclusion is strengthened when similar ensemble composite HSSs are constructed, but from the location of largest April surface temperature differences from each individual member. This results in the HSSs significantly increasing in North America, but less in central Russia and southern Asia (as shown in Figure S5). This suggests that either the large observed HSSs in North America may not remain consistent into the future, or within North America the model is not as realistic, unlike the other regions that show greater consistency within the ensemble composite.

5. Conclusions

Seasonal forecasting has many major applications, such as to inform public health and agriculture decisions that are based on the likelihood of future climate extremes, and to predict sea ice coverage (Smith et al., 2016). This has fueled the recent growth in development of subseasonal to seasonal forecasts (Robertson et al., 2015; Vitart et al., 2017; White et al., 2017). Here, in agreement with previous studies, we show that large correlations from observations and a nine-member ensemble from the CESM1(WACCM) model between March Arctic TCO and springtime surface temperatures are seen in central Russia, southern Asia, and North America. Similarly, surface temperature differences of over 6 K are observed between years of opposing March Arctic ozone extremes (upper and lower twentieth percentiles) in these regions. Additionally, while some locations experience large internal model variability, central Russia and southern Asia have ensemble signals detectable outside of sampling uncertainty. This gives confidence that TCO holds information that can be used for NH springtime seasonal forecasting.

Our analysis of individual ensemble members shows that a simple linear regression model that uses March TCO as its sole predictor is able to accurately predict surface temperatures in regions that exhibit large correlations, and this predictive skill is maintained over the 6 years in observations, and 4 years in the individual ensemble members. In observations, the year of 2011, which experienced an extreme low ozone anomaly was forecast very accurately.

Further, the results presented here give confidence that, due to the very large surface temperature differences between years of opposing March Arctic ozone extremes, we can with a good degree of skill use March ozone to predict the sign of the surface temperature anomalies in April in central Russia and to a lesser extent in southern Asia. These locations display large skill in the observations, consistent comparisons of the surface temperature differences between the model and observations, and the lowest internal model variability (Figure 5). The observed and ensemble composite April HSSs, when using only data corresponding to opposing March Arctic ozone extremes, approach 60 and 50 (80% and 75% correct predictions), respectively, for the locations of largest April surface temperature difference in central Russia. There are also indications that significant meaningful skill is obtained through to May in central Russia.

In summary, our results indicate that springtime NH seasonal predictions of surface temperature can be improved in some locations using stratospheric ozone information, which is consistently measured to a high degree of accuracy.

Acknowledgments
K. S. and S. S. were partly supported by National Science Foundation grants 1419667 and 1338814. D. K. was supported by National Science Foundation grant 1338814. This research has been conducted as part of the NOAA MAPP S2S Prediction Task Force and C. B. and E. B. were supported by NOAA grant NA16OAR4310064. WACCM is a component of the Community Earth System Model (CESM), which is supported by the National Science Foundation. We would like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Model results shown in this paper are available online (at https://acomstaff.acom.ucar.edu/dkin/JGR_Stone_2019). The National Institute for Water and Atmosphere-Bodeker Scientific total column ozone database V3.2 patched data set can be obtained online (at http://www.bodekerscientific.com/data/total-column-ozone). ERA-Interim data can be freely obtained online (at https://apps.ecmwf.int/datasets/).

References

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*(5542), 581–584. https://doi.org/10.1126/science.1063315

Butler, A. H., Polvani, L. M., & Deser, C. (2014). Separating the stratospheric and tropospheric pathways of El Niño–Southern Oscillation teleconnections. *Environ. Res. Lett.*, *9*(2), 024014. https://doi.org/10.1088/1748-9326/9/2/024014

Calvo, N., Polvani, L. M., & Solomon, S. (2015). On the surface impact of Arctic stratospheric ozone extremes. *Environmental Research Letters*, *10*(9). https://doi.org/10.1088/1748-9326/10/9/094003

Cheung, J. C. H., Haigh, J. D., & Jackson, D. R. (2014). Impact of EOS MLS ozone data on medium-extended range ensemble weather forecasts. *Journal of Geophysical Research: Atmospheres*, *119*, 9253–9266. https://doi.org/10.1002/2014JD021823

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Engel, A., Rigby, M., Burkholder, J. B., Fernandez, R. P., Froidevaux, L., Hall, B. D., et al. (2018). Update on ozone-depleting substances (ODSs) and other gases of interest to the Montreal Protocol, Chapter 1. In *Scientific Assessment of Ozone Depletion: 2018*, Global Ozone Research and Monitoring Project–Report No. 58. Geneva, Switzerland: World Meteorological Organization.

Garcia, R. R., Smith, A. K., Kinnison, D. E., de la Cámara, Á., & Murphy, D. J. (2017). Modification of the gravity wave parameterization in the whole atmosphere community climate model: Motivation and results. *Journal of the Atmospheric Sciences*, *74*(1), 275–291. https://doi.org/10.1175/JAS-D-16-0104.1

Haase, S., & Matthes, K. (2018). The importance of interactive chemistry for stratosphere–troposphere–coupling. *Atmospheric Chemistry and Physics*, *19*(5), 3417–3432. https://doi.org/10.5194/acp-19-3417-2019

Ivy, D. J., Solomon, S., Calvo, N., & Thompson, D. W. J. (2017). Observed connections of Arctic stratospheric ozone extremes to Northern Hemisphere surface climate. *Environmental Research Letters*, *12*(2). https://doi.org/10.1088/1748-9326/aa57a4

Jia, L., Yang, X., Vecchi, G. A., Gudgel, R., Delworth, T., Fueglistaler, S., et al. (2017). Seasonal prediction skill of northern extratropical surface temperature driven by the stratosphere. *Journal of Climate*, *30*(12), 4463–4475. https://doi.org/10.1175/JCLI-D-16-0475.1

Johnson, N. C., Collins, D. C., Feldstein, S. B., L'Heureux, M. L., & Riddle, E. E. (2014). Skillful wintertime North American temperature forecasts out to 4 weeks based on the state of ENSO and the MJO*. *Weather and Forecasting*, *29*(1), 23–38. https://doi.org/10.1175/WAF-D-13-00102.1

Kang, S. M., Polvani, L. M., Fyfe, J. C., & Sigmond, M. (2011). Impact of polar ozone depletion on subtropical precipitation. *Science*, *332*(6032), 951–954. https://doi.org/10.1126/science.1202131

Karpechko, A. Y., Perlwitz, J., & Manzini, E. (2014). A model study of tropospheric impacts of the Arctic ozone depletion 2011. *Journal of Geophysical Research: Atmospheres*, *119*, 7999–8014. https://doi.org/10.1002/2013JD021350

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Kinnison, D. E., Brasseur, G. P., Walters, S., Garcia, R. R., Marsh, D. R., Sassi, F., et al. (2007). Sensitivity of chemical tracers to meteorological parameters in the MOZART-3 chemical transport model. *Journal of Geophysical Research*, *112*, D20302. https://doi.org/10.1029/2006JD007879

Manney, G. L., Santee, M. L., Rex, M., Livesey, N. J., Pitts, M. C., Veefkind, P., et al. (2011). Unprecedented Arctic ozone loss in 2011. *Nature*, *478*(7370), 469–475. https://doi.org/10.1038/nature10556

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N., & Polvani, L. M. (2013). Climate change from 1850 to 2005 simulated in CESM1(WACCM). *Journal of Climate*, *26*(19), 7372–7391. https://doi.org/10.1175/JCLI-D-12-00558.1

Mills, M. J., Schmidt, A., Easter, R., Solomon, S., Kinnison, D. E., Ghan, S. J., et al. (2016). Global volcanic aerosol properties derived from emissions, 1990–2014, using CESM1(WACCM). *Journal of Geophysical Research: Atmospheres*, *121*, 2332–2348. https://doi.org/10.1002/2015JD024290

Mundhenk, B. D., Barnes, E. A., Maloney, E. D., & Baggett, C. F. (2017). Skillful empirical subseasonal prediction of landfalling atmospheric river activity using the Madden–Julian oscillation and quasi-biennial oscillation. *npj Climate and Atmospheric Science*, *1*(1). https://doi.org/10.1038/s41612-017-0008-2

Neely, R. R. I., & Schmidt A. (2016). VolcanEESM: Global volcanic sulphur dioxide (SO2) emissions database from 1850 to present—Version 1.0, Cent. Environ. Data Anal. https://doi.org/10.5285/76ebdc0b-0eed-4f70-b89e-55e606bcd568

Robertson, A. W., Kumar, A., Peña, M., & Vitart, F. (2015). Improving and promoting subseasonal to seasonal prediction. *Bulletin of the American Meteorological Society*, *96*(3), ES49–ES53. https://doi.org/10.1175/BAMS-D-14-00139.1

Smith, G. C., Roy, F., Reszka, M., Surcel Colan, D., He, Z., Deacu, D., et al. (2016). Sea ice forecast verification in the Canadian Global Ice Ocean Prediction System. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 659–671. https://doi.org/10.1002/qj.2555

Solomon, S., Ivy, D., Gupta, M., Bandoro, J., Santer, B., Fu, Q., et al. (2017). Mirrored changes in Antarctic ozone and stratospheric temperature in the late 20th versus early 21st centuries. *Journal of Geophysical Research: Atmospheres*, *122*, 8940–8950. https://doi.org/10.1002/2017JD026719

Solomon, S., Ivy, D. J., Kinnison, D., Mills, M. J., Neely, R. R. III, & Schmidt, A. (2016). Emergence of healing in the Antarctic ozone layer. *Science*, *353*(6296), 269–274. https://doi.org/10.1126/science.aae0061

Solomon, S., Kinnison, D., Bandoro, J., & Garcia, R. (2015). Simulation of polar ozone depletion: An update. *Journal of Geophysical Research: Atmospheres*, *120*, 7958–7974. https://doi.org/10.1002/2015JD023365

Struthers, H., Bodeker, G. E., Austin, J., Bekki, S., Cionni, I., Dameris, M., et al. (2009). The simulation of the Antarctic ozone hole by chemistry-climate models. *Atmospheric Chemistry and Physics*, *9*(17), 6363–6376. https://doi.org/10.5194/acp-9-6363-2009

Thirumalai, K., DiNezio, P. N., Okumura, Y., & Deser, C. (2017). Extreme temperatures in Southeast Asia caused by El Niño and worsened by global warming. *Nature Communications*, *8*, 15531. https://doi.org/10.1038/ncomms15531

Thompson, D. W., & Solomon, S. (2002). Interpretation of recent Southern Hemisphere climate change. *Science*, *296*(5569), 895–899. https://doi.org/10.1126/science.1069270

Thompson, D. W., & Wallace, J. M. (2001). Regional climate impacts of the Northern Hemisphere annular mode. *Science*, *293*(5527), 85–89. https://doi.org/10.1126/science.1058958

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, *98*(1), 163–173. https://doi.org/10.1175/BAMS-D-16-0017.1

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J. T., Lazo, J. K., Kumar, A., et al. (2017). Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, *24*(3), 315–325. https://doi.org/10.1002/met.1654

WMO (2014). Scientific assessment of ozone depletion: 2014, Glob. Ozone Res. Monit. Proj. No. 55, 416.

Zhang, P., Wu, Y., Simpson, I. R., Smith, K. L., Zhang, X., De1, B., & Callaghan, P. (2018). A stratospheric pathway linking a colder Siberia to Barents-Kara Sea sea ice loss. *Science Advances*, *4*(7), eaat6025. https://doi.org/10.1126/sciadv.aat6025