

Improving Antarctic Total Ozone Projections by a Process-Oriented Multiple Diagnostic Ensemble Regression

ALEXEY YU. KARPECHKO

Arctic Research, Finnish Meteorological Institute, Helsinki, Finland

DOUGLAS MARAUN

GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

VERONIKA EYRING

Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

(Manuscript received 25 February 2013, in final form 26 July 2013)

ABSTRACT

Accurate projections of stratospheric ozone are required because ozone changes affect exposure to ultraviolet radiation and tropospheric climate. Unweighted multimodel ensemble-mean (uMMM) projections from chemistry–climate models (CCMs) are commonly used to project ozone in the twenty-first century, when ozone-depleting substances are expected to decline and greenhouse gases are expected to rise. Here, the authors address the question of whether Antarctic total column ozone projections in October given by the uMMM of CCM simulations can be improved by using a process-oriented multiple diagnostic ensemble regression (MDER) method. This method is based on the correlation between simulated future ozone and selected key processes relevant for stratospheric ozone under present-day conditions. The regression model is built using an algorithm that selects those process-oriented diagnostics that explain a significant fraction of the spread in the projected ozone among the CCMs. The regression model with observed diagnostics is then used to predict future ozone and associated uncertainty. The precision of the authors' method is tested in a pseudoreality; that is, the prediction is validated against an independent CCM projection used to replace unavailable future observations. The tests show that MDER has higher precision than uMMM, suggesting an improvement in the estimate of future Antarctic ozone. The authors' method projects that Antarctic total ozone will return to 1980 values at around 2055 with the 95% prediction interval ranging from 2035 to 2080. This reduces the range of return dates across the ensemble of CCMs by about a decade and suggests that the earliest simulated return dates are unlikely.

1. Introduction

There is a large spread among chemistry–climate models (CCMs) in their projected evolution of stratospheric ozone during the twenty-first century (Eyring et al. 2007, 2010b; WMO 2011). Providing reliable stratospheric ozone projections is important for a variety of different reasons, including its importance for UV radiation (Hegglin and Shepherd 2009) and impacts on tropospheric climate. In the Southern Hemisphere, the

recovery of Antarctic stratospheric ozone is expected to influence tropospheric circulation and, hence, climate change (Perlwitz et al. 2008; Karpechko et al. 2010a; Son et al. 2010). The importance of stratospheric ozone changes as a climate factor has been widely recognized. Climate models participating in the fifth phase of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012) have been recommended to prescribe time-varying ozone forcing in case ozone is not calculated interactively (Eyring et al. 2013). The projected part of the ozone-forcing dataset provided for the CMIP5 simulations without interactive chemistry is based on an “ensemble of opportunity” (Tebaldi and Knutti 2007) of CCM simulations from the second round of the Chemistry–Climate Model Validation (CCMVal-2) activity. It consists

Corresponding author address: A.Yu. Karpechko, Arctic Research, Finnish Meteorological Institute, P.O.B. 503, Helsinki FIN-00101, Finland.

E-mail: alexey.karpechko@fmi.fi

of a time series averaged across all available CCMVal-2 models, which is merged with observational data to provide a continuous time series from 1850 to 2100 (Cionni et al. 2011).

The question remains whether a “one model–one vote” multimodel mean in which all available models are equally weighted represents the best estimate of future ozone or any other quantity of interest (Knutti 2010; Knutti et al. 2010b; Weigel et al. 2010). Efforts have been undertaken to grade CCMs based on their ability to simulate observed stratospheric ozone climatology and trends (Karpechko et al. 2010b), or key processes relevant for stratospheric ozone (Vaugh and Eyring 2008, hereafter WE08), and to use these grades to explore the value of weighting of ozone projections. The important caveat of these and similar studies, such as those focused on tropospheric projections (e.g., Murphy et al. 2004; Connolley and Bracegirdle 2007; Reichler and Kim 2008), is that they rely on ad hoc selected model-grading metrics that are difficult to justify. Part of the problem is the selection and weighting of the diagnostics. WE08 adopted the process-oriented approach advocated by Eyring et al. (2005, 2006) and based model grading on diagnostics selected to represent specific chemical and transport processes important for stratospheric ozone. However, it has not been investigated in a quantitative way whether these diagnostics contain any information about projected ozone changes. The importance of considering the relation between diagnostics and projected changes has recently been demonstrated in several studies (Whetton et al. 2007; Boe et al. 2009; Hall and Qu 2006; Räisänen et al. 2010; Abe et al. 2011; Bracegirdle and Stephenson 2012, hereafter BS12). In particular, BS12 proposed the ensemble regression approach based on a linear regression between surface temperature biases in present climate and projected surface temperature changes, and demonstrated an improved precision in the estimates of projected temperature change over regions adjacent to climatological sea ice edges. The selection of diagnostics remains, however, a challenge and is done differently in these studies. So far, the majority of the studies seek to link the change in the quantity of interest either to its biases in the present-day climatology (Whetton et al. 2007; Abe et al. 2011; BS12) and trends (Boe et al. 2009) or to a variety of ad hoc diagnostics (Räisänen et al. 2010). It has been argued that ultimately it is the realistic representation of processes that is most linked to the credibility of model projections (Eyring et al. 2005; Knutti et al. 2010a), thus providing a means to successful weighting of model projections based on model performance.

The issues raised above—namely, assessing the relevance of diagnostics for projection weighting and providing

an objective measure for the selection of these diagnostics—are addressed in this paper. Specifically, we extend the ensemble regression approach by BS12 to the case of multiple diagnostics and apply it to estimate future stratospheric ozone changes. Our method takes into account the dependence of projected ozone changes on the simulation of some selected key processes that are known to be relevant for stratospheric ozone. In principle, the method can be applied for any other quantity of interest, provided that processes driving its long-term changes are sufficiently well known and can be diagnosed from model outputs. A process-oriented approach has long been adopted for CCM evaluation as part of the CCMVal activity (Eyring et al. 2005, 2006, 2010b). Its application led to a substantial improvement of our knowledge of model biases and deficiencies as well as better understanding of the spread in ozone projections among CCMs (WMO 2007, 2011). In the present study we apply process-oriented diagnostics that have been shown to be important for stratospheric ozone in CCMVal and that have been applied before by WE08 and Eyring et al. (2010b). Our approach provides an objective way of selecting those diagnostics that significantly improve the statistical prediction of ozone change from present-day climate across the model ensemble for estimating future ozone change. The approach essentially consists of building a multiple regression model based on CCM simulations and using the regression model for statistical prediction of future ozone.

In section 2 we provide the theoretical background. In section 3 we describe the method that we refer to as multiple diagnostic ensemble regression (MDER). In section 4 we present the model simulations and diagnostics. In section 5 we use the MDER method to predict future ozone change. We also carry out a cross validation to demonstrate that the MDER approach leads to an improved estimate of future ozone compared to the weighting scheme suggested by WE08 and to the unweighted multimodel mean approach (uMMM) (i.e., approach where all models are equally weighted). In section 6 we conclude with a summary and discussion of the results.

2. Theory

BS12 described a statistical model that relates climate model projections to a single diagnostic of present-day climate [see their Eq. (3)]. Here, we extend their approach to the case of multiple diagnostics. Let $\mathbf{Y} = \{y_1, y_2, \dots, y_n\}^T$ be the vector of projected model values for the quantity of interest (here, projected ozone change), where n is the number of models. Let us assume that there is a relationship between simulated

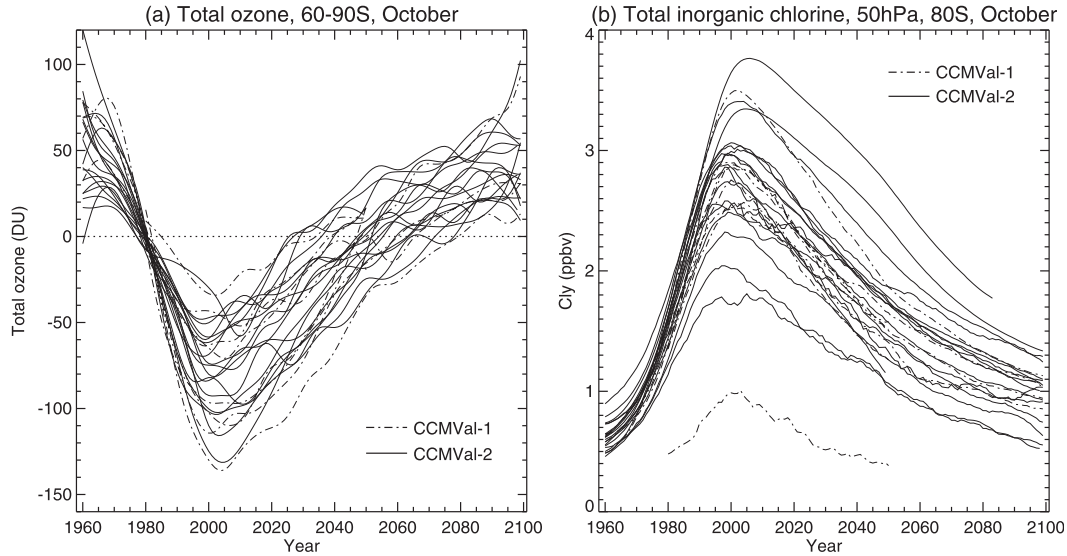


FIG. 1. Time series of (a) Antarctic October total ozone anomalies with respect to 1980 and (b) October total Cl_y at 50 hPa, 80°S from the CCMVal-1 (REF-2) and CCMVal-2 (REF-B2) models. All model time series are smoothed with a 1–2–1 filter repeated iteratively 30 times.

present-day diagnostics \mathbf{X} and the projected model values of the quantity of interest \mathbf{Y} , which can be written as

$$\mathbf{Y} = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1)$$

where $\mathbf{1} = \{1, 1, \dots, 1\}^T$ is a column vector of size n ,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \dots & \mathbf{x}_{1,m} \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \dots & \mathbf{x}_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}_{n,1} & \dots & \dots & \mathbf{x}_{n,m} \end{pmatrix}, \quad m \text{ is the number of}$$

diagnostics, $\boldsymbol{\varepsilon}$ is the vector of independent random variables representing the uncertainty in the projections, and β_0 and $\boldsymbol{\beta}$ are the regression model parameters to be estimated, with $\boldsymbol{\beta}$ being a column vector of size m . The vector $\boldsymbol{\varepsilon}$ can be understood to represent the influence on the projections of all factors not accounted for by \mathbf{X} , the nonlinear interactions between the diagnostics in \mathbf{X} and the climate noise.

The first two terms of Eq. (1) can be combined by combining the vector $\mathbf{1}$ with the matrix \mathbf{X} into the design matrix (von Storch and Zwiers 1999). However, we keep the terms separated in order to explicitly demonstrate that in the case $\boldsymbol{\beta} = 0$ our approach reduces to the uMMM. The model parameters β_0 and $\boldsymbol{\beta}$ can be estimated using the maximum likelihood approach, which in the case of normally distributed residuals is the simple least squares estimator:

$$\hat{\boldsymbol{\beta}}_0 = (\mathbf{1}^T \mathbf{1})^{-1} (\mathbf{1}^T \mathbf{Y} - \mathbf{1}^T \mathbf{X} \boldsymbol{\beta}) \equiv \mathbf{N}^T \mathbf{Y} - \mathbf{N}^T \mathbf{X} \boldsymbol{\beta}, \quad (2a)$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X} - \mathbf{X}^T \mathbf{1} \mathbf{N}^T \mathbf{X})^{-1} (\mathbf{X}^T - \mathbf{X}^T \mathbf{1} \mathbf{N}^T) \mathbf{Y}, \quad (2b)$$

where $\mathbf{N} \equiv (\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}$ is a column vector of size n with all elements equal to $1/n$. We next assume that the relationship defined by the regression model in Eq. (1), with parameters estimated in the model world, holds for the true climate. Under this assumption, Eq. (1) can be used to predict the expected future change y_0 , given the vector of observed diagnostics \mathbf{X}_0 :

$$\begin{aligned} \hat{y}_0 &= \hat{\beta}_0 + \mathbf{X}_0^T \hat{\boldsymbol{\beta}} \\ &= [\mathbf{N}^T + (\mathbf{X}_0^T - \mathbf{N}^T \mathbf{X})(\mathbf{X}^T \mathbf{X} - \mathbf{X}^T \mathbf{1} \mathbf{N}^T \mathbf{X})^{-1} \\ &\quad \times (\mathbf{X}^T - \mathbf{X}^T \mathbf{1} \mathbf{N}^T)] \mathbf{Y}, \end{aligned} \quad (3)$$

where \hat{y}_0 is the estimate of y_0 . The error vector $\boldsymbol{\varepsilon}$ is not present in Eq. (3) because the equation is written for an estimate rather than for the random variable y_0 . Note that $\mathbf{N}^T \mathbf{Y} \equiv \bar{y}$ is the uMMM. Therefore, if there is no link between simulated present-day diagnostics and $\mathbf{Y}(\boldsymbol{\beta} = 0)$, then the second term in the right-hand part of Eq. (3) vanishes and the best estimate of true climate change is equal to the uMMM:

$$\hat{y}_0 = \hat{\beta}_0 = \mathbf{N}^T \mathbf{Y} \equiv \bar{y}.$$

For illustrative purposes we present here an example with $m = 1$ using data from CCMs participating in the

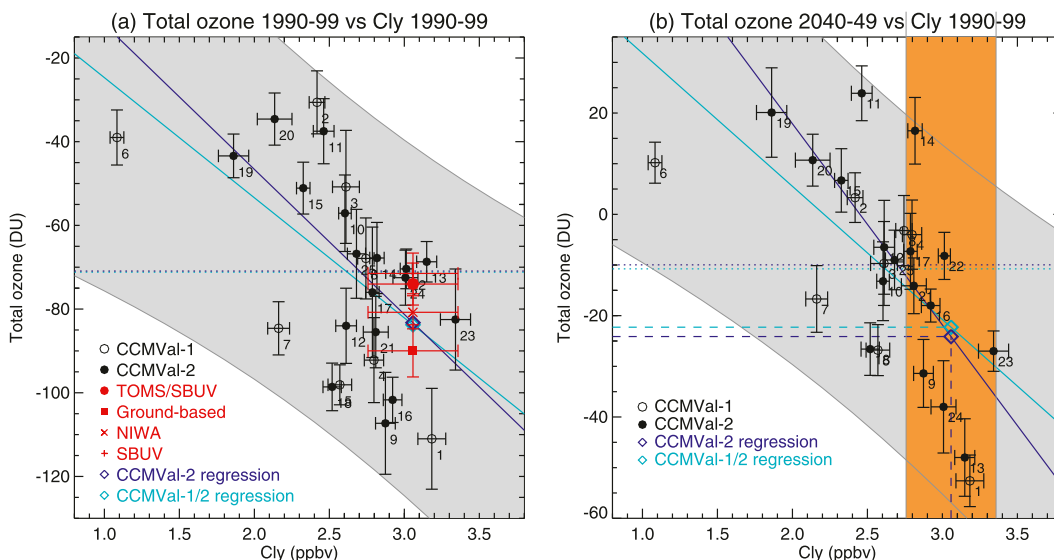


FIG. 2. Scatterplots of climatological-mean (a) present-day Antarctic October total ozone anomalies (1990–99) and (b) future Antarctic October total ozone change (2040–49) vs present-day Antarctic October Cl_y (1990–99), Cl_y -SP, for the CCMVal models. The models are numbered according to Table 1. REF-1 and REF-B1 simulations are used for the present-day total ozone and Cl_y -SP. REF-2 and REF-B2 simulations are used for the future total ozone. The error bars show the simulated $1\text{-}\sigma$ error of the mean values. The red symbols and error bars in (a) correspond to the observed values based on the four total ozone datasets (1990–99) and Cl_y -SP inferred from the HALOE observations. For total ozone, the error bars show the observed $1\text{-}\sigma$ error of the mean value. For Cl_y -SP, the error bars in (a) and the orange shading in (b) show measurement uncertainty. The solid blue (turquoise) lines show the least squares fit to the CCMVal-2 (CCMVal-1/2) data. The gray shading around the least squares fit marks the 95% prediction interval for the total ozone response based on CCMVal-2. The dotted lines show the ensemble-mean ozone anomaly ($u\text{MMM}$). The blue (turquoise) diamond and dashed lines in (b) indicate the regression-predicted ozone anomalies corresponding to the observed Cl_y -SP value based on the CCMVal-2 (CCMVal-1/2).

coordinated model intercomparison organized by CCMVal. There were two rounds of the intercomparison, referred to as CCMVal-1 and CCMVal-2. These simulations are described in section 4. Figure 1 shows the time series of simulated Antarctic total column ozone anomalies for October calculated with respect to 1980 and October polar-vortex inorganic chlorine (Cl_y) at 50 hPa. Models simulate the minimum in total ozone around the year 2000, approximately at the same time as they simulate the maximum in Cl_y (WMO 2007; Eyring et al. 2007); see section 5 for discussion of the mechanism behind this relationship. Figure 2a shows that there is a significant anticorrelation across the models between simulated present-day (1990–99) Antarctic total column ozone and Cl_y at 50 hPa just before the time of the Cl_y maximum. The correlation r does not depend much on whether we consider only the simulations from the more recent ensemble CCMVal-2 ($r = -0.60$, $p = 0.01$) or combine the CCMVal-2 with the simulations from the preceding CCMVal-1 ensemble ($r = -0.57$, $p = 0.003$). Figure 2b shows that simulated present-day Cl_y concentrations (1990–99) also correlate well with projected total column ozone in 2040–49. This

result is similar regardless of whether we only consider the simulations from the CCMVal-2 ensemble ($r = -0.71$, $p = 0.001$) or combine CCMVal-2 and CCMVal-1 datasets ($r = -0.62$, $p = 0.001$). Note that a relationship similar to that shown in Fig. 2b was shown in Strahan et al. (2011). The regression between simulated present-day Cl_y and present-day total column ozone agrees well with observations (Fig. 2a). Assuming that the regression between simulated Cl_y and projected mid-twenty-first-century total ozone change will hold for future true climate, one can use the regression model from Fig. 2b to predict future total ozone. The regression model based on the CCMVal-2 models with observed Cl_y predicts an ozone change of -24 Dobson units (DU) for the period 2040–49 with the 95% prediction intervals of $(-57, 9 \text{ DU})$. The prediction is somewhat lower than, but consistent with, the $u\text{MMM}$ estimate of -10 DU within the 95% confidence interval. Reassuringly, the prediction based on the combined CCMVal-1 and -2 dataset gives an ozone change of -22 DU with the 95% prediction intervals of $(-57, 12 \text{ DU})$, which is very close to that based on CCMVal-2 only. We emphasize, however, that at this stage the relationship between Cl_y and future ozone change

is only used to illustrate the concept, rather than to make any inferences about future ozone change.

The assumption that the regression between the present-day diagnostic and a future-climatic variable—derived from climate models—holds also for true climate might seem to be weak at first sight, especially since it explicitly requires imperfect models away from observed climate to span the desired relationship. However, a much weaker assumption is done in traditional model-weighting approaches: a model that simulates present-day climate better than another model necessarily better simulates future climate. Whereas traditional approaches do not show this relationship, ensemble regression establishes exactly such a relationship across a broad model ensemble.

BS12 pointed out that the vector multiplier of \mathbf{Y} in Eq. (3) can be associated with model weights \mathbf{W} ; that is,

$$\mathbf{W} = [\mathbf{N}^T + (\mathbf{X}_0^T - \bar{\mathbf{X}}^T)(\mathbf{X}^T\mathbf{X} - \mathbf{X}^T\mathbf{1}\bar{\mathbf{X}}^T)^{-1}(\mathbf{X}^T - \mathbf{X}^T\mathbf{1}\mathbf{N}^T)]^T. \tag{4}$$

Here, we note that $\bar{\mathbf{X}} \equiv (\mathbf{N}^T\mathbf{X})^T$ is the vector of multi-model mean diagnostics. As discussed by BS12, the weights provided by Eq. (4) do not involve terms proportional to differences between observed and simulated diagnostics; that is, they are not proportional to model biases. The weights can also be negative; thus, the estimate \hat{y}_0 can fall outside the range of model projections. These properties differentiate the weights defined by Eq. (4) from the traditional model weights because the latter depend on model biases and cannot be negative. However, if the statistical model given by Eq. (1) is adequate ($\beta \neq 0$), then projections by models having smaller biases in diagnostics will tend to stay closer to the predicted change. In other words, these models would have higher weights in the traditional sense.

In a similar manner one can rewrite the equation for confidence intervals for the mean of the response variable, or for the prediction interval,

$$cf = \sqrt{\mathbf{W}_{cf}^T \mathbf{Y}}, \tag{5}$$

where \mathbf{W}_{cf} , the model weights for calculating $p \times 100\%$ confidence intervals, are calculated as follows (von Storch and Zwiers 1999):

$$\begin{aligned} \mathbf{W}_{cf} &= (n - m - 1)^{-1} t_{(1-p)/2}^2 \\ &\times [b + \mathbf{X}_0^T (\mathbf{X}_D^T \mathbf{X}_D)^{-1} \mathbf{X}_0] \mathbf{Y}^T (\mathbf{I} - \mathbf{X}_D (\mathbf{X}_D^T \mathbf{X}_D)^{-1} \mathbf{X}_D^T). \end{aligned} \tag{6}$$

Here, $\mathbf{X}_D = \begin{pmatrix} 1 & \mathbf{x}_{1,1} & \dots & \mathbf{x}_{1,m} \\ 1 & \mathbf{x}_{2,1} & \dots & \mathbf{x}_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \mathbf{x}_{n,1} & \dots & \mathbf{x}_{n,m} \end{pmatrix}$ is the design matrix,

\mathbf{I} is the $n \times n$ identity matrix, and $t_{(1-p)/2}$ is the critical value from the t distribution with $(n - m - 1)$ degrees of freedom. The parameter b is either equal to 0, in which case Eq. (6) gives the confidence intervals for the mean of the response variable, or 1, in which case it gives the prediction interval. In this paper we always provide the 95% prediction interval; that is, Eq. (6) is used with $b = 1$.

3. Multiple diagnostic ensemble regression

We use the above theory to develop the MDER method. The method consists of the two principal steps: (i) building the statistical model based on climate model simulations and (ii) using the statistical model with observed diagnostics to predict future change. The first step consists of selecting the diagnostics for the matrix \mathbf{X} in Eq. (1) and estimating the parameters of the regression. Ideally, the process-oriented diagnostics should account for all processes that are expected to affect the future evolution of ozone, and the MDER should identify those that are important for a particular application. The initial selection of the present-day diagnostics is unavoidably based on expert judgment (i.e., on the knowledge of which processes drive future changes in the variable of interest). However, a priori it is not clear which of the selected diagnostics help to confine the estimate of the future variable: the linear predictive power might be low or different diagnostics might correlate with each other. Choosing a suitable subset of m useful diagnostics from the set of initially selected l diagnostics ($l \geq m$) is a classical statistical-model-selection problem (e.g., Davison 2003). A standard approach to model selection is the stepwise algorithm, which starts with one diagnostic and continues until a certain stopping criterion is met (Wilks 2006; von Storch and Zwiers 1999, 166–167). The problem of choosing the stopping criterion is discussed in section 5.

The outcomes of the model selection algorithm are the matrix \mathbf{X} and the model parameters β_0 and β . In the second step, Eq. (3) with the vector of observed diagnostics \mathbf{X}_0 is used to predict future change, and Eq. (5) is used to estimate the uncertainty of the prediction.

The above algorithm provides a prediction for a certain time in the future. It is often desirable to predict time evolution of the quantity of interest, in which case Eq. (3) can be rewritten as

TABLE 1. CCMs used in this study. Further details on CCMVal-1 models can be found in Eyring et al. (2006) and on CCMVal-2 models in Morgenstern et al. (2010) and Eyring et al. (2010b) as well as in the references given below.

No.	Model name	Model expansion	Group and location	Simulation period	Reference
CCMVal-1					
1	AMTRAC	Atmospheric Model with Transport and Chemistry	Geophysical Fluid Dynamics Laboratory (GFDL), United States	1960–2099	Austin et al. (2007)
2	CCSRNIES	Center for Climate System Research/National Institute for Environmental Studies	National Institute of Environmental Studies (NIES), Tsukuba, Japan	1980–2050	Akiyoshi et al. (2004)
3	CMAM	Canadian Middle Atmosphere Model	Canadian Centre for Climate Modelling and Analysis (CCCma), University of Toronto, Canada	1960–2099	Fomichev et al. (2007)
4	GEOSCCM	Goddard Earth Observing System Chemistry–Climate Model	National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC), United States	1960–2099	Pawson et al. (2008)
5	MRI	Meteorological Research Institute	Meteorological Research Institute (MRI), Japan	1980–2099	Shibata and Deushi (2005)
6	SOCOL	Solar–Climate–Ozone Links	Davos Physical Meteorological Observatory and World Radiation Centre (PMOD/WRC) and Institute for Atmospheric and Climate Science, Swiss Federal Institute of Technology Zurich (IAC ETHZ), Switzerland	1980–2050	Egorova et al. (2005)
7	ULAQ	Università degli Studi dell'Aquila	University of L'Aquila, Italy	1960–2050	Pitari et al. (2002)
8	WACCM	Whole Atmosphere Chemistry–Climate Model	National Center for Atmospheric Research (NCAR), United States	1980–2050	Garcia et al. (2007)
CCMVal-2					
9	AMTRAC3	Atmospheric Model with Transport and Chemistry 3	GFDL, United States	1960–2099	Austin and Wilson (2010)
10	CAM3.5	Community Atmosphere Model, version 3.5	NCAR, United States	1960–2099	Lamarque et al. (2008)
11	CCSRNIES	Center for Climate System Research/National Institute for Environmental Studies	NIES, Tsukuba, Japan	1960–2100	Akiyoshi et al. (2009)
12	CMAM	Canadian Middle Atmosphere Model	MSC, University of Toronto, York University, Canada	1960–2099	Scinocca et al. (2008); de Grandpré et al. (2000)
13	CNRM-ACM	Centre National de Recherches Météorologiques–ARPEGE Climat Coupled MOCAGE	Meteo-France, France	1960–2099	Déqué (2007); Teysédre et al. (2007)
14	E39CA	ECHAM4.L39(DLR)/CHEM/ATTILA	German Aerospace Center (DLR), Germany	1960–2049	Stenke et al. (2009); Garny et al. (2009)
15	EMAC	ECHAM/MESSy Atmospheric Chemistry model	Free University of Berlin (FU Berlin), Germany	1960–2100	Jöckel et al. (2006); Nissen et al. (2007)
16	GEOSCCM	Goddard Earth Observing System Chemistry–Climate Model	NASA GSFC, United States	1960–2099	Pawson et al. (2008)

TABLE 1. (Continued)

No.	Model name	Model expansion	Group and location	Simulation period	Reference
17	LMDZrepro	Laboratoire de Météorologie Dynamique Zoom-REPROBUS	L'Institut Pierre-Simon Laplace (IPSL), France	1960–2098	Jourdain et al. (2008)
18	MRI	Meteorological Research Institute	MRI, Japan	1960–2099	Shibata and Deushi (2008a; 2008b)
19	NIWA-SOCOL	National Institute of Water and Atmospheric Research Solar–Climate–Ozone Links	NIWA, New Zealand	1960–2098	Schraner et al. (2008)
20	SOCOL	Solar–Climate–Ozone Links	PMOD/WRC and IAC ETHZ, Switzerland	1960–2100	Schraner et al. (2008)
21	ULAQ	Università degli Studi dell'Aquila	University of L'Aquila, Italy	1960–2100	Pitari et al. (2002)
22	UMSLIMCAT	Unified Model–SLIMCAT	University of Leeds, United Kingdom	1960–2099	Tian and Chipperfield (2005); Tian et al. (2006)
23	UMUKCA-METO	Unified Model/U.K. Chemistry Aerosol Community Model–Met Office	Met Office, United Kingdom	1960–2083	Morgenstern et al. (2009)
24	UMUKCA-UCAM	Unified Model/U.K. Chemistry Aerosol Community Model–University of Cambridge	University of Cambridge, United Kingdom, and NIWA, New Zealand	1960–2099	Morgenstern et al. (2009)
25	WACCM	Whole Atmosphere Chemistry–Climate Model	NCAR, United States	1960–2098	Garcia et al. (2007)

$$\hat{y}_0(t) = \mathbf{W}^T \mathbf{Y}(t), \quad (7)$$

where $\hat{y}_0(t)$ and $\mathbf{Y}(t)$ refer to the time-dependent variables and \mathbf{W} is defined by Eq. (4). Similarly, the equation for the confidence intervals [Eq. (5)] can be rewritten in a time-dependent form.

Time independence of \mathbf{W} in Eq. (7) can only be assumed if the matrix \mathbf{X} is time independent (i.e., if the same processes drive the changes through the whole period of time). Also, estimation of \mathbf{W} based on a limited set of \mathbf{Y} is prone to sampling errors that may vary in time. This can be inferred from Eq. (6), which shows that the uncertainty of the prediction depends on \mathbf{Y} and is therefore time dependent in general. Therefore, for a time-dependent prediction, the MDER algorithm needs to be repeated for all time steps, unless the assumption of time independence of \mathbf{X} is valid. In section 5c we test the validity of the assumption that \mathbf{X} , and thus \mathbf{W} , is time independent.

4. Model simulations and observations

We apply the MDER method to the data from CCMs participating in the CCMVal-1 and -2 intercomparisons organized by CCMVal. Compared to CCMVal-1, more models participated and more diagnostics have been evaluated in CCMVal-2 (Eyring et al. 2010b). The CCMVal-2 simulations have also covered longer periods,

but the simulations in both rounds were performed with very similar external forcing.

The ensembles of these simulations were extensively analyzed and discussed elsewhere (Eyring et al. 2006, 2007, 2010a,b; Austin et al. 2010; Oman et al. 2010). Table 1 lists the models used. Only models that extended their projections until at least 2050 were included in the analysis—a criterion that left four CCMVal-1 models (E39C, LMDZrepro, MAECHAM4CHEM, and UMSLIMCAT) out. Altogether, projections from 25 different models are used in the analysis: 8 CCMVal-1 and 17 CCMVal-2 models.

We base our results mainly on the analysis of the CCMVal-2 simulations. We use the combined CCMVal-1 and -2 data, referred to as CCMVal-1/-2, to test the sensitivity of the results to CCM sampling. Combining the two datasets increases the sample size, which may be expected to provide a more robust estimate of the relationship between process-oriented diagnostics and future ozone change. On the other hand, combining the datasets may be questioned since it increases the issue of model independence (Knutti et al. 2010b; Masson and Knutti 2011), because some of the models participated in both CCMVal-1 and -2 and had only minor updates implemented between the two rounds and thus can hardly be considered as independent. However, one might expect that model dependencies would only decrease the effective number of degrees of freedom (i.e.,

TABLE 2. Description of the diagnostics and the number of CCMs for which the diagnostics are available. The last column shows the observed values with uncertainty, but only for those diagnostics that have been selected by the MDER method.

Short Name	Diagnostic	Averaging period	<i>N</i> models	Observed value
Temp-SP	Temperature, September–November (SON), 60°–90°S, 30–50 hPa	1980–99	25	—
Temp-NP	Temperature, December–February (DJF), 60°–90°N, 30–50 hPa	1980–99	25	-208.9 ± 4.9 K
U-SP	Date of transition to easterlies, 60°S, 20 hPa	1980–99	25	—
HFlux-SH	Eddy heat flux, July–August, 40°–80°S, 100 hPa	1980–99	25	-7.3 ± 2.0 K m s ⁻¹
HFlux-NH	Eddy heat flux, January–February, 40°–80°N, 100 hPa	1980–99	25	—
Temp-Trop	Temperature, annual, 5°S–5°N, 80 hPa	1980–99	25	—
H ₂ O-Trop	Water vapor, annual, 5°S–5°N, 80 hPa	1990–99	25	3.3 ± 0.5 ppm
CH ₄ -Subt	Methane gradient between 0°–30°N/S, March/October, 50 hPa	1990–99	25	—
CH ₄ -SP	Methane, October, 80°S, 30–50 hPa	1990–99	25	0.6 ± 0.1 ppm
CH ₄ -EQ	Methane, March, 10°S–10°N, 30–50 hPa	1990–99	25	1.6 ± 0.1 ppm
Tape-R	Tape recorder amplitude attenuation	1990–99	23	—
Tape-c	Tape recorder phase speed	1990–99	23	—
Age-50	Age of air, annual, 10°S–10°N and 35°–55°N, 50 hPa	1990–99	18	—
Age-10	Age of air, annual, 10°S–10°N and 35°–55°N, 10 hPa	1990–99	18	—
Cl _y -SP	Cl _y , October, 80°S, 50 hPa	1990–99	25	3.1 ± 0.3 ppm
Cl _y -Mid	Cl _y , annual, 30°–60°N, 50 hPa	1990–99	25	—
HCl-SP	HCl decrease from May to August, 70°–90°S, 30–50 hPa	1990–99	25	—
Cl _x -SP	Cl _x , August–October, 70°–90°S, 50 hPa	1990–99	14	—
HNO ₃ -SP	HNO ₃ decrease from May to August, 70°–90°S, 30–50 hPa	1990–99	17	—

increase the uncertainty) but not bias the regression coefficients, nor the best-estimate prediction.

It is expected that different processes drive ozone changes in different regions and altitudes (Eyring et al. 2005, 2007, 2010b; Oman et al. 2010; Strahan et al. 2011). For example, the removal of halogens is expected to play a key role in the recovery of Antarctic lower-stratospheric ozone, while stratospheric cooling is expected to dominate the upper-stratospheric ozone changes globally (Oman et al. 2010). This suggests that the regression model would yield different results if applied to different regions and altitudes. Here, we focus on the twenty-first-century projections of Antarctic 60°–90°S total ozone in October taken from REF-2 and REF-B2 simulations by the CCMVal-1 and -2 models, respectively. The greenhouse gas (GHG) concentrations in these simulations follow the Intergovernmental Panel on Climate Change (IPCC) Special Report on Emission Scenarios (SRES) A1B scenario. Sea surface temperatures and sea ice concentrations are taken from coupled atmosphere–ocean model projections using the same GHG scenario. Surface halogen concentrations in REF-2 follow the WMO (2003) Ab scenario, while those in REF-B2 follow the WMO (2007) adjusted A1 scenario. The return of halogen concentrations to the background level in the A1 scenario is delayed compared to the Ab scenario, but this difference is modest (Newman et al. 2007) and neglected here. For the CCMVal-2 models EMAC and E39CA, REF-B2 simulations are not available and thus are replaced by the SCN-B2d simulations,

which are identical to REF-B2 but additionally include observed solar variability, volcanic activity, and the quasi-biennial oscillation (QBO) until 2006, as well as repeating solar cycle and QBO in the future.

Many models have large offset biases in total ozone (WMO 2007; Eyring et al. 2006, 2010b) that persist throughout the simulation time and complicate direct intermodel comparison. Therefore, it is a common practice to consider ozone anomalies adjusted to total ozone in some reference year. Here, we choose to adjust the time series to total ozone in the year 1980, which is the starting year for many CCMVal-1 simulations. To reduce the uncertainty due to interannual variability, the value of total ozone in 1980 was estimated by fitting a third-order polynomial to the original time series for each model between the starting year of the simulation (1960 or 1980) and 1999. The ozone anomaly in the twenty-first century is thereafter referred to as ozone change. The main findings of this study do not depend on the details of calculating the anomaly because similar results are obtained when the anomaly is calculated following the procedure described in Eyring et al. (2006).

The starting set of *l* process-oriented diagnostics is taken from WE08 and selected diagnostics from chapter 6 of Eyring et al. (2010b) (Table 2). More discussion on the relevance of these diagnostics for stratospheric ozone can be found in Eyring et al. (2005, 2006, 2010b). The diagnostics from WE08 cover transport and dynamical processes relevant for stratospheric ozone, but not polar chemistry. They are complemented with

diagnostics for two processes important for Antarctic ozone: denitrification and chlorine activation (chapter 6 of Eyring et al. 2010b). As a diagnostic for denitrification, we use the decrease of nitric acid (HNO_3) from May to August in the lower-stratospheric Antarctic polar vortex ($\text{HNO}_3\text{-SP}$). For chlorine activation, we use two diagnostics: the decrease of lower-stratospheric hydrogen chloride (HCl) from May to August (HCl-SP) and active chlorine Cl_x ($\equiv \text{ClO} + 2 \times \text{Cl}_2\text{O}_2$) averaged over the August–October period within the polar vortex ($\text{Cl}_x\text{-SP}$). The first diagnostic is constrained by observations, but it is not an ideal proxy for chlorine activation, because it is soluble in liquid aerosols and therefore can be removed from the gas phase and thus not converted into active chlorine (D. Kinnison 2012, personal communication). The second diagnostic directly quantifies the amount of active chlorine, but it is not well constrained by observations and, in principle, is not suitable for the MDER method.

Because the simulated diagnostics have to be comparable to the observed ones, they are calculated using the historical climate simulations REF-1 and REF-B1 from CCMVal-1 and -2, respectively. These simulations are forced with observed sea surface temperature, sea ice concentrations, surface concentrations of well-mixed GHGs and halogens, solar variability, and aerosol from major volcanic eruptions (Eyring et al. 2006, 2010b). Some of the CCMVal-1 simulations are only available for the period 1980–99. Therefore, the diagnostics are calculated either over the period 1980–99 or 1990–99, depending on available observations (see Table 2). Only 13 diagnostics are available for all CCMal-2 (and also for CCMVal-1) models and used for the MDER analysis. The other diagnostics are used only for correlation calculation (section 5).

Since not all the diagnostics are selected by the MDER method, the observations [\mathbf{X}_0 in Eq. (3)] are only needed for a subset of diagnostics. The observed values for this subset are shown in the last column of Table 2. We use the same observational datasets as in WE08: HALOE observations (Grooß and Russell 2005) are used for $\text{CH}_4\text{-SP}$, $\text{CH}_4\text{-EQ}$, $\text{Cl}_y\text{-SP}$, and $\text{H}_2\text{O-Trop}$, and the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) (Uppala et al. 2005) is used for stratospheric temperatures (Temp-NP) and eddy heat flux (HFlux-SH). We have also considered the National Centers for Environmental Prediction (NCEP) and Met Office (UKMO) stratospheric analyses as alternative sources of observations for Temp-NP and found that this diagnostic is very similar among these three datasets. The difference between the coldest (UKMO) and the warmest (NCEP) datasets is only 0.2 K, which makes a negligible difference to our results. More observations for the

chemical parameters are available for the period after 1999; however, we do not use them because they may not be directly comparable to the simulated values. The exception is $\text{Cl}_y\text{-SP}$ for which the observations are only available during years 1992 and 2005. To avoid a high bias in simulations, which would otherwise be introduced if the observed value from 1992 were directly compared with the 1990–99 simulated values, we used both 1992 and 2005 observations and interpolated/extrapolated these two values for each year over the period 1990–99 by assuming a linear increase. Finally, four observational datasets are used for total ozone: merged satellite dataset constructed from individual Total Ozone Mapping Spectrometer (TOMS) version 8 and Solar Backscatter Ultraviolet SBUV/2 version 8 satellite datasets (Stolarski and Frith 2006), ground-based measurements [updated from Fioletov et al. (2002)], the National Institute of Water and Atmospheric Research (NIWA) combined total column ozone database (Bodeker et al. 2005), and SBUV/SBUV/2 retrievals [updated from Miller et al. (2002)].

5. Estimates of future ozone

a. The period 2040–49

We now apply the MDER method to predict future ozone change. We first focus on the period 2040–49, which is the latest decade for which the simulations by all 25 models are available (Table 1). The decadal averaging is done in order to remove the interannual variability considered here as noise. Figure 3 shows the absolute correlation coefficients between the present-day diagnostics (averaged over 1980–99 or 1990–99; see Table 2) and future ozone change by 2040–49. In the CCMVal-2 ensemble the highest correlation ($r = -0.76$, $p = 0.006$) is found between future ozone change and present-day active chlorine at 50 hPa averaged over 70°–90°S and between August and October ($\text{Cl}_x\text{-SP}$). Active chlorine is directly responsible for ozone depletion; therefore, a strong anticorrelation with total ozone is expected. However, $\text{Cl}_x\text{-SP}$ is available for a limited number of models and is not constrained by observations. Therefore, this diagnostic is not used in our further analysis. The second highest correlation ($r = 0.73$, $p = 0.001$) is found with October methane concentration at 80°S averaged between 30 and 50 hPa ($\text{CH}_4\text{-SP}$). In the combined CCMVal-1/-2 ensemble $\text{CH}_4\text{-SP}$ has the highest correlation with future ozone change ($r = 0.67$, $p = 0.001$). $\text{CH}_4\text{-SP}$ is targeted to diagnose the spread among the models caused by differences in transport. CH_4 is photodissociated in the upper stratosphere; thus, lower values of $\text{CH}_4\text{-SP}$ indicate a larger fraction of upper-stratospheric air mass in the polar vortex, and normally

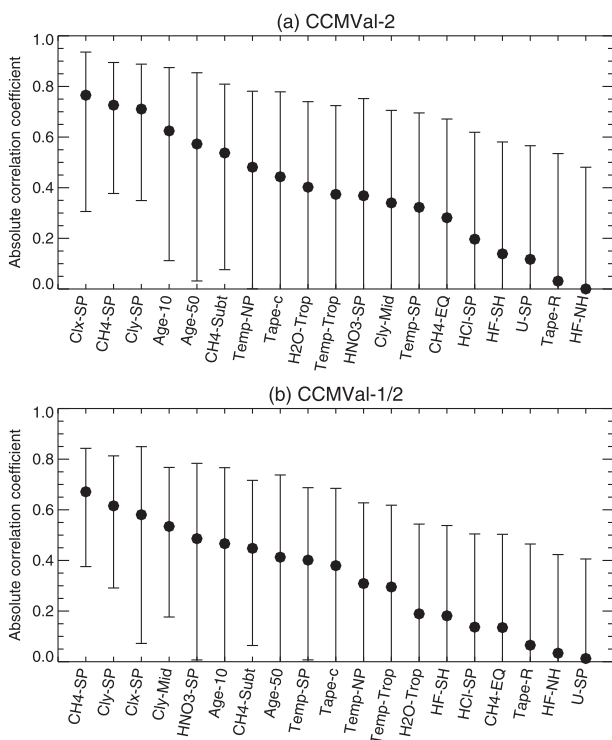


FIG. 3. Absolute correlation coefficient between Antarctic October total ozone anomalies for 2040–49 and present-day diagnostics in Table 2 among the (a) CCMVal-2 and (b) CCMVal-1/2 models. The error bars show the 95% confidence intervals for the correlation coefficients.

suggest a more isolated polar vortex and/or slower stratospheric circulation. The strong positive correlation between $\text{CH}_4\text{-SP}$ and future ozone indicates that models with a less isolated polar vortex and/or faster stratospheric circulation project earlier stratospheric ozone return dates. In the CCMVal-2 ensemble the correlation is also statistically significant ($p = 0.05$) for polar Cl_y

diagnostic ($\text{Cl}_y\text{-SP}$), the age of air diagnostics (Age-10, Age-50), and methane in the subtropics ($\text{CH}_4\text{-Subt}$). Cl_y is produced in the upper stratosphere and reflects the time spent there by an air mass before it descends to the lower stratosphere; therefore, $\text{Cl}_y\text{-SP}$ is also a transport diagnostic. In the polar vortex Cl_y is converted to ozone-depleting active chlorine compounds (Cl_x) during austral winter–spring and is therefore related to chemical ozone destruction. Note that the transport diagnostics ($\text{CH}_4\text{-SP}$, $\text{Cl}_y\text{-SP}$, $\text{Cl}_y\text{-Mid}$, Age-10, Age-50) and $\text{Cl}_x\text{-SP}$ correlate with each other (not shown).

The results of the MDER calculations for 2040–49 are presented in Table 3. The model selection relies on a subjectively chosen stopping criterion. A standard approach is to add diagnostics until the regression sum of squares is not significantly increased according to an F test. Applying this test with the $p = 0.05$ significance level to the CCMVal-2 dataset and the 13 diagnostics available for all models results in a regression model that includes four terms: $\text{CH}_4\text{-SP}$, $\text{H}_2\text{O-Trop}$, $\text{CH}_4\text{-EQ}$, and HF-SH . Although a physical relation between all these diagnostics and future ozone change may be hypothesized, we note that the contribution of the last three terms is rather small and hardly improves the model. Moreover, cross validation (see section 5b) indicates that this model is overfitted and thus should not be used. In general, another stopping criterion for model selection could be applied. We have tried the ones based on the increase in the coefficient of determination R^2 and the decrease in the mean squared error (MSE). For example, adding diagnostics to the model until none of the remaining diagnostics increase R^2 by more than 0.1 results in a regression model with three terms: $\text{CH}_4\text{-SP}$, $\text{H}_2\text{O-Trop}$, and $\text{CH}_4\text{-EQ}$. However cross-validation tests indicate that a regression model that performs best on independent data is the one that contains the first

TABLE 3. The regression models for the 2040–49-mean total ozone discussed in the text. The following symbols are used: $y \equiv \text{O}_3$; $x_1 \equiv \text{CH}_4\text{-SP}$; $x_2 \equiv \text{H}_2\text{O-Trop}$; $x_3 \equiv \text{CH}_4\text{-EQ}$; $x_4 \equiv \text{HF-SH}$; $x_5 \equiv \text{Cl}_y\text{-SP}$. Also shown is the uMMM.

Model name	Model ensemble	Model equation	Predicted mean ozone change and the 95% prediction intervals	R^2
uMMM	CCMVal-2	$\frac{1}{N} \sum_{i=1, \dots, N} y_i$	-10 ± 40 DU	—
MDER: F test	CCMVal-2	$217.0 + 81.7 \times x_1 - 11.9 \times x_2 - 143.3 \times x_3 + 2.8 \times x_4$	-21 ± 18 DU	0.90
MDER: R^2 test	CCMVal-2	$208.4 + 81.1 \times x_1 - 11.4 \times x_2 - 152.5 \times x_3$	-22 ± 21 DU	0.84
MDER: $\text{CH}_4\text{-SP}$ only	CCMVal-2	$-52.2 + 57.1 \times x_1$	-17 ± 32 DU	0.53
MDER: $\text{CH}_4\text{-SP}$ excluded	CCMVal-2	$47.4 - 34.0 \times x_5$	-24 ± 34 DU	0.51
uMMM	CCMVal-1/2	$\frac{1}{N} \sum_{i=1, \dots, N} y_i$	-11 ± 39 DU	—
MDER: F test	CCMVal-1/2	$1.3 + 44.4 \times x_1 - 17.0 \times x_5$	-23 ± 29 DU	0.58
MDER: $\text{CH}_4\text{-SP}$ only	CCMVal-1/2	$-54.5 + 60.0 \times x_1$	-18 ± 32 DU	0.45
MDER: $\text{CH}_4\text{-SP}$ excluded	CCMVal-1/2	$58.1 - 26.3 \times x_5$	-22 ± 35 DU	0.38

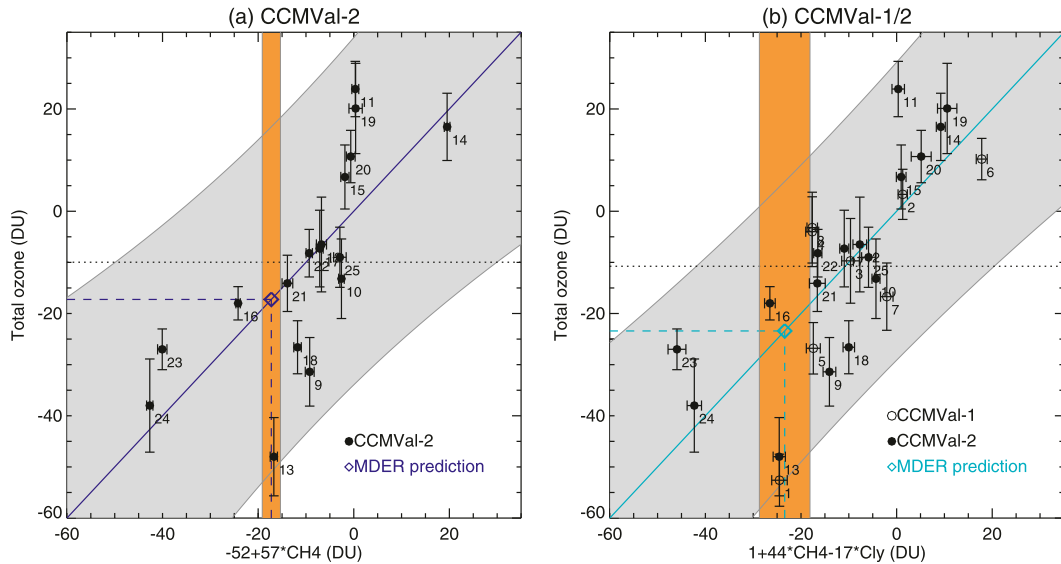


FIG. 4. Scatterplots of climatological-mean Antarctic October total ozone change (2040–49) vs (a) the quantity $(-52 + 57 \times \text{CH}_4\text{-SP})$ for the CCMVal-2 models and (b) the quantity $(1 + 44 \times \text{CH}_4\text{-SP} - 17 \times \text{Cl}_y\text{-SP})$ for the CCMVal-1/-2 models. $\text{CH}_4\text{-SP}$ and $\text{Cl}_y\text{-SP}$ are climatological-mean Antarctic October CH_4 and Cl_y , respectively, for the present day (1990–99). The models are numbered according to Table 1. The error bars show the simulated 1σ error of the mean values. The solid blue and turquoise lines show the least squares fit to the model data. The gray shading around the least squares fit marks the 95% confidence interval for the total ozone response. The dotted lines show the ensemble-mean ozone anomaly (uMMM). The blue/turquoise diamond and dashed lines indicate the regression-predicted ozone anomalies. The orange shading shows measurement uncertainty.

term (i.e., $\text{CH}_4\text{-SP}$) only. Table 3 shows that predictions based on these three regression models are close to each other. Thus, the final choice of the model impacts mainly the confidence intervals of the prediction. We choose the model with the $\text{CH}_4\text{-SP}$ term only because this model performs best in the cross-validation tests (section 5b), suggesting that the other models may be overfitted. This choice is also the most conservative one because this model has the largest uncertainty and its prediction interval covers the prediction intervals based on the other models. This model explains 53% of the spread among CCMs in projected ozone change. Using Eq. (3) with $\text{CH}_4\text{-SP}$ being the only term in \mathbf{X} and the observed $\text{CH}_4\text{-SP}$ value (\mathbf{X}_0) taken from Table 2 we predict a future ozone change by 2040–49 of -17 ± 32 DU (Fig. 4).

It is instructive to test the sensitivity of the MDER method to the absence of $\text{CH}_4\text{-SP}$ diagnostic. When $\text{CH}_4\text{-SP}$ is excluded, the algorithm based on an F test selects the regression model including the $\text{Cl}_y\text{-SP}$ term only; that is, it is the same model as that shown in Fig. 2b. $\text{Cl}_y\text{-SP}$ strongly anticorrelates with the excluded $\text{CH}_4\text{-SP}$ diagnostic ($r = -0.59$), and the fact that the automated selection algorithm stops after selecting $\text{Cl}_y\text{-SP}$ provides further evidence that the additional terms in the regression model are unnecessary. Reassuringly, regression models based on either $\text{CH}_4\text{-SP}$ or $\text{Cl}_y\text{-SP}$ predict

future ozone change values that are close to each other and consistent with each other within the uncertainty (Table 3).

We next test the sensitivity of the result to CCM sampling by applying MDER to the CCMVal-1/-2 dataset. The algorithm based on an F test selects a regression model with both $\text{CH}_4\text{-SP}$ and $\text{Cl}_y\text{-SP}$ terms. This model predicts future ozone change of -23 DU, which is consistent with the predictions based on CCMVal-2 (Table 3). Like in CCMVal-2, the contribution of the second selected term is small; thus, it may be redundant. On the other hand, this model performs well in cross validation (section 5b) and therefore the inclusion of both terms might be justified. The regression model including $\text{CH}_4\text{-SP}$ only predicts future ozone change of -17 DU, which is close to the previous model. Thus, for predicting purposes, the choice of the model is not crucial. Results for the model containing both $\text{CH}_4\text{-SP}$ and $\text{Cl}_y\text{-SP}$ are shown in Fig. 4b. When $\text{CH}_4\text{-SP}$ is excluded, the resulting regression model contains only the $\text{Cl}_y\text{-SP}$ term; that is, it is the same model as that shown in Fig. 2b. All three statistical models predict future ozone change values that are close to each other and consistent with each other within the uncertainty (Table 3).

The 95% prediction intervals for the future ozone change by the MDER method are comparable to the spread across individual model projections. However,

the spread does not provide any information about the probability of a particular ozone change value. For example, the probability that the future true ozone change will fall outside the spread cannot be inferred from the spread alone. On the contrary, the MDER method provides the information about the probabilities of future true ozone change values, given that the assumptions discussed in section 2 are correct. The prediction interval for the uMMM can be calculated by assuming that the future ozone changes simulated by individual models are random samples from the “true change” estimated by uMMM. In this case the 95% prediction interval can be estimated by multiplying the standard deviation across the individual model projections by 1.96 (Table 3). However, given the dependence of model projections on the ability to simulate transport in present-day climate, this assumption is unlikely to be valid.

The prediction intervals cited above only include the uncertainty of the statistical model. Accounting for observational uncertainty would increase the intervals. Observational uncertainty shown in Fig. 4 includes the standard error of the mean CH₄-SP estimated from the available HALOE CH₄ data south of 78°S. For the case of the CCMVal-1/-2 dataset it is combined by the law of combination of errors with the observational uncertainty of Cl_y-SP. It is seen that the uncertainty of the statistical model is considerably larger than the uncertainty due to observations. However, more studies are needed to assess the impact of the observational uncertainty on the MDER predictions.

Based on the above results we conclude that the MDER prediction of ozone change by 2040–49 does not depend on the specific choice of diagnostic as long as the applied set of diagnostics contains a diagnostic representing the key process responsible for the future change. We also conclude that results are consistent for both CCMVal-2 and the CCMVal-1/-2 datasets. The MDER method captures the dependence of projected ozone on simulation of the specific process, stratospheric meridional circulation, which controls the age of air and thus the concentration of ozone-depleting substances within the Antarctic polar vortex. On the other hand, the uncertainty of the prediction, as diagnosed by the coefficient of determination (Table 3), varies from one statistical model to another. This points out that a successful choice of diagnostic is required in order to reduce the uncertainty of prediction. Moreover, when there are too many potential predictors, an automated model selection algorithm can fail, resulting in a model containing unnecessary terms (i.e., in an overfitted model). The fact that the overfitting issue is more apparent when MDER is applied to the CCMVal-2 ensemble suggests that it may be

related to the smaller sample size ($N = 17$) in this ensemble compared to CCMVal-1/-2.

b. Cross validation of the MDER method in pseudoreality

Since, trivially, future observations of ozone are not available, it is not clear whether the MDER method really provides more precise estimates of future ozone than the uMMM or than the approach described in WE08, who selected different sets of diagnostics (e.g., average transport or only Cl_y grades) to explore the value of weighting. Therefore, we cross validate the methods in a pseudoreality (e.g., Maraun 2012) to compare model-averaging approaches at least in a model world. To this end, we select one model of our ensemble as pseudoreality, and consider the remaining models as the model ensemble that is used to predict future ozone changes of the pseudoreality model. As there is no decisive preference to which model should be regarded as pseudoreality, we choose each model as pseudoreality in turns. For each pseudoreality, the predictions made by the tested methods are compared to the future pseudoobservations.

We compare the quality of future ozone estimates by the MDER method against both the uMMM and the weighting method that was used in Fig. 8 of WE08 (i.e., average transport grade) in a pseudoreality. Only the 13 diagnostics available for all models are used. In each pseudoreality the quality of the prediction is quantified by the squared difference between the future Antarctic total ozone change averaged over 2040–49 in a pseudoreality and the ensemble-mean total ozone change (2040–49) estimated from the model ensemble projections using the uMMM and the MDER. The results are shown in Fig. 5. When the MDER with stopping algorithm based on an F test is applied to the CCMVal-2 ensemble, the first selected term is either CH₄-SP (14 cases) or Cl_y-SP (3 cases) depending on pseudoreality. In all but three pseudorealities the statistical models contain three to four terms and often the same set (CH₄-SP, H₂O-Trop, CH₄-EQ, HF-SH) as that in the full dataset (see section 5a).

We quantify the predictive skill of the MDER method by the Brier skill score (Wilks 2006):

$$\text{BSS} = 100 \times \left(1 - \frac{\sum_{i=1,n} e_i^2}{\sum_{i=1,n} o_i^2} \right),$$

where e_i and o_i are the errors of prediction in i th pseudoreality by the MDER and uMMM, respectively, and n is the number of pseudorealities. BSS is a relative quantity and shows a percentage improvement of the

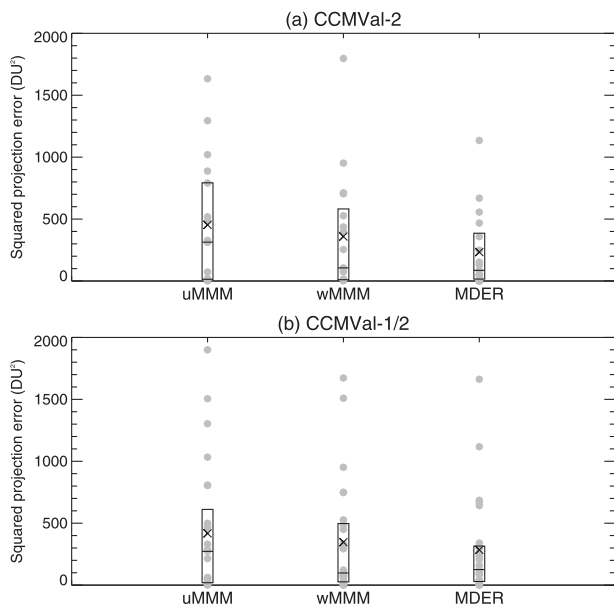


FIG. 5. The squared differences (projection errors) between the ensemble-mean Antarctic October total ozone change in 2040–49 and the total ozone change in pseudoreality for each pseudoreality considered (gray circles) based on (a) CCMVal-2 and (b) CCMVal-1/2. The ensemble mean is calculated from equally weighted (i.e., unweighted) projections (uMMM), projections weighted following the WE08 method and employing their transport diagnostics (wMMM), and the MDER method. Crosses show the mean projection error for each method; the boxes show 25th–75th percentiles across the error ensembles and the bars inside the boxes show the medians.

MDER method compared to the uMMM. For the 2040–49 period, the percentage improvement with respect to uMMM of the MDER method with the stopping algorithm based on an F test is less than 1%. The poor performance of the model on independent data suggests that it is overfitted. However, when the model has $\text{CH}_4\text{-SP}$ as the only term in all pseudorealities, the percentage improvement compared to the uMMM is 47%, indicating that such a model performs better than uMMM. The predictions errors based on this model are shown in Fig. 5a along with those based on uMMM. Since this result is obtained in a pseudoreality it cannot be interpreted as an indication of improved accuracy because the observations for prediction verification are not available. Rather, the results indicate an improved precision of the prediction compared to the uMMM.

The cross validation of the MDER method applied to the combined CCMVal-1/2 dataset is shown in Fig. 5b. The results are shown for the MDER method with the stopping algorithm based on an F test. In all but two pseudorealities the statistical models contain the $\text{CH}_4\text{-SP}$ and $\text{Cl}_y\text{-SP}$ terms—that is, the same as in the full dataset (section 5a). The percentage improvement compared to

the uMMM is 32%. As in the case of CCMVal-2 dataset, restricting the model to only the $\text{CH}_4\text{-SP}$ term leads to even a larger skill improvement of 42%.

Figure 5 also reveals that, despite an overall improved performance, in some cases MDER predictions lead to large errors. These cases correspond to models that are far away from the regression lines. Large deviations from the regression lines may be related, among other things, to unphysical characteristics of these models. Such a possibility implies that a model screening based on their performance may be required before admitting models to MDER; however, this has not been done in this study.

For comparison, we also validate the weighting method by WE08 in pseudoreality. WE08 used the simple weighting metric applied to individual process-oriented diagnostics:

$$g = \max\left(0; 1 - \frac{|\mu_{\text{mod}} - \mu_{\text{obs}}|}{3\sigma_{\text{obs}}}\right),$$

where μ_{mod} and μ_{obs} are a climatological-mean diagnostic in a model and in observations, respectively, and σ_{obs} is a measure of observation uncertainty, normally represented by the interannual standard deviation. To explore the value of model weighting and to test the robustness of the ensemble, model weights were assigned according to g averaged over all diagnostics, or over different subsets of the diagnostics (more details are given in WE08). Figure 5 shows the projection errors for the weighting based on the average transport grades ($\text{CH}_4\text{-Subt}$, $\text{CH}_4\text{-SP}$, $\text{CH}_4\text{-EQ}$, $\text{Cl}_y\text{-SP}$, $\text{Cl}_y\text{-Mid}$) because the transport diagnostics were used for projection weighting shown in Fig. 8 of WE08. BSS for this method is 21% when applied to CCMVal-2 data and 17% when applied to CCMVal-1/2 data (i.e., the WE08 weighting method based on the transport diagnostics also improves the precision of the prediction compared to the uMMM).

c. The MDER prediction for the twenty-first century

We now apply the MDER method to predict the ozone evolution for the twenty-first century. First, we use the assumption of time independency of weights \mathbf{W} and apply Eq. (7) with the weights calculated for the period 2040–49 based on the model containing only the $\text{CH}_4\text{-SP}$ term to obtain ensemble-mean simulated total ozone for the period 1960–2099. The weighted ozone time series is shown in Fig. 6 along with the uMMM. The individual model projections have been smoothed with the 1–2–1 filter iteratively repeated 30 times (e.g., Eyring et al. 2007) before applying the weights. During the twentieth century the ensemble-mean total ozone

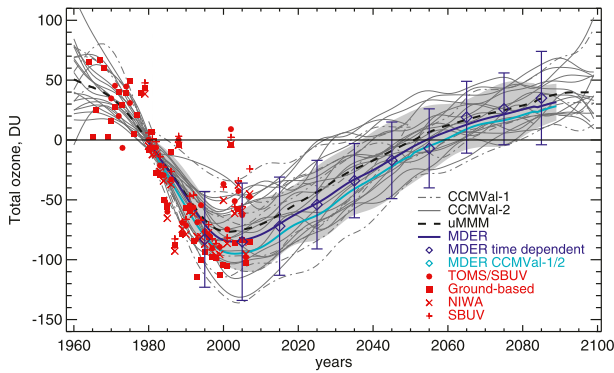


FIG. 6. Time series of Antarctic October total ozone anomalies from the CCMVal simulations and the four observational datasets. The model time series are smoothed with a 1–2–1 filter iteratively repeated 30 times. The dashed black line shows the uMMM CCMVal-2 time series. The blue line and the shading show the CCMVal-2 ensemble-mean time series and the 95% confidence interval for the MDER prediction with the model weights estimated for the 2040–49 period. The turquoise line show the CCMVal-1/-2 ensemble-mean time series for the MDER prediction with the model weights estimated for the 2040–49 period. The blue diamonds and the error bars show the ensemble-mean time series and the 95% confidence interval for the MDER prediction with model weights estimated for each decade separately.

time series weighted with the MDER weights are in a good agreement with observations and show ozone values slightly smaller than the uMMM. During the twenty-first century the weighted time series consistently shows ozone values smaller than the uMMM by about 5–10 DU. The constant- \mathbf{W} model projects that Antarctic ozone column returns to 1980 values by 2055, 3 years later than projected by the uMMM. The 95% prediction intervals associated with the constant- \mathbf{W} model prediction give the return date ranging from 2037 to 2079. For comparison, the return dates inferred from the smoothed individual CCMVal-2 model projections

range from 2026 to 2078 (Fig. 6). Thus, the MDER estimated range is about a decade smaller than that inferred from the individual model projections. A similar result is obtained based on the CCMVal-1/-2 dataset. Here, the MDER projects the return date by 2059 with the 95% prediction intervals ranging from 2041 to 2079—that is, close to that based on the CCMVal-2 result (Fig. 6).

We next remove the assumption of time independency of \mathbf{W} and repeat the MDER calculations for each decade of the twenty-first century, as well as for the 1990s, performing the decadal averaging in order to remove the interannual variability. The MDER algorithm with stopping criteria based on an F test selects models with one to four terms depending on a decade. Following the results of section 5a, only the first selected term is retained in the final models. Since the contribution of the second and higher terms is small, the introduced difference to the decadal-mean ozone change value is less than 5 DU in all but one case (in the decade 2060–69 the difference is 10 DU) and therefore may be neglected. The results are shown in Table 4 and in Fig. 6. Between 1990 and 2049 total ozone agrees well with that predicted by the constant- \mathbf{W} model. During this period the MDER algorithm selects $\text{CH}_4\text{-SP}$ or $\text{Cl}_y\text{-SP}$ diagnostics as the explanatory variable for total ozone, suggesting that the same processes drive the total ozone change during this period. Note that the regression model for 1990–99 contains only the $\text{Cl}_y\text{-SP}$ diagnostic (i.e., it is the same model as that shown in Fig. 2a). After 2060, the MDER predicted total ozone exceeds the total ozone values predicted by the constant- \mathbf{W} model. During this period the regression model selects the Arctic stratospheric temperature, Temp-NP, as the explanatory variable (Table 4). One possible interpretation of this relation is that this diagnostic, which is partly related to simulated planetary wave activity, may correlate with projected changes in planetary wave activity,

TABLE 4. The regression models for the decadal-mean total ozone during the period 1990–2089. The following symbols are used: $y \equiv \text{O}_3$; $x_1 \equiv \text{CH}_4\text{-SP}$; $x_2 \equiv \text{Cl}_y\text{-SP}$; $x_3 \equiv \text{Temp-NP}$. Also shown is the uMMM.

Period	uMMM $\left(\frac{1}{N} \sum_{i=1, \dots, N} y_i\right)$	MDER equation	MDER predicted mean ozone change and the 95% confidence intervals	R^2
1990–99	-71 ± 42 DU	$22.7 - 34.6 \times x_2$	-83 ± 40 DU	0.35
2000–09	-76 ± 56 DU	$-128.4 + 70.9 \times x_1$	-85 ± 49 DU	0.43
2010–19	-64 ± 48 DU	$-108.9 + 61.2 \times x_1$	-72 ± 41 DU	0.44
2020–29	-45 ± 50 DU	$-100.3 + 75.4 \times x_1$	-54 ± 37 DU	0.60
2030–39	-25 ± 44 DU	$-74.8 + 67.2 \times x_1$	-34 ± 31 DU	0.62
2040–49	-10 ± 40 DU	$-52.2 + 57.1 \times x_1$	-17 ± 32 DU	0.53
2050–59	7 ± 39 DU	$106.8 - 37.2 \times x_2$	-7 ± 33 DU	0.50
2060–69	17 ± 34 DU	$897.0 - 4.2 \times x_3$	19 ± 30 DU	0.42
2070–79	24 ± 33 DU	$869.4 - 4.0 \times x_3$	26 ± 30 DU	0.41
2080–89	34 ± 30 DU	$752.3 - 3.4 \times x_3$	35 ± 29 DU	0.35

and thus be related to projected total ozone change. However, this interpretation remains to be checked.

The uncertainty associated with the MDER prediction varies from decade to decade. The changes in the uncertainty estimates are most likely associated with the decadal climate variability which cannot be accounted for by the diagnostics and thus represents noise in our statistical model. Moreover, the source of the decadal variability may be ocean (i.e., SSTs that are prescribed in all but one analyzed CCM and therefore cannot be accounted for by present-day diagnostics in principle). Understanding sources of uncertainty in the MDER estimations is important but will not be pursued further in this study. We emphasize however that, as can be seen in Fig. 6, the differences between the uncertainty estimates have only small effects on the prediction intervals for the ozone return date in our study.

In summary, the MDER allows reducing the range of predicted return dates by about a decade when compared to the spread among the individual model projections and suggests that the earliest return dates across individual model projections are unlikely.

6. Discussion and conclusions

In this study we have introduced a method for estimating future ozone change based on multiple diagnostic ensemble regression, referred to as the MDER method. In contrast to previous studies, this method is based on selected process-oriented diagnostics that are known to be important for stratospheric ozone. The advantage of this method is that it provides an observational constraint on future projections and allows making rigorous probabilistic statements.

This method has been applied to an ensemble of CCM projections of Antarctic October total ozone. Mathematically, the method is an extension of the methods introduced by Boe et al. (2009) and BS12 to the case of multiple predictors. When there are a large number of potentially important predictors, an automated model-selection algorithm may lead to an overfitted statistical model. The overfitting problem may be associated with a too-small sample size and therefore may be important in other climate applications since climate model ensembles typically consist of a relatively small number of models (~ 20). Therefore, the choice of the final model may need to rely on cross-validation tests. Additionally, we emphasize the importance of a physically motivated choice of predictors for the regression. One advantage of the method is that it helps finding diagnostics that are important for a specific application. Specifically, our approach has recovered a tight relationship between projected Antarctic total ozone and polar vortex

methane, a transport diagnostic, simulated by CCMs in present climate.

Cross validation of the method in a pseudoreality has shown reduced errors in projected ozone in the MDER method compared to the equally weighted multimodel mean (uMMM) method, implying that estimates of future total ozone changes obtained by our method would have a higher precision than those of uMMM. The return of Antarctic total ozone to 1980 values estimated by the MDER method is projected to occur by around 2055 with the 95% confidence intervals for the return date ranging from 2037 to 2079. These estimates do not change significantly when another CCM ensemble is used. The MDER best estimate of the ozone return date differs from the uMMM-estimated return date, which is projected for 2052; however, the difference is within the uncertainty of the method. This suggests that, although the MDER method gives more precise results, the uMMM method, used for example to produce ozone time series for the CMIP5 experiments, remains a reasonable approach to calculate ensemble-projected changes, at least in the case of Antarctic October total ozone and for the simulations included here. The 95% confidence interval for the return date predicted by the MDER method is about a decade smaller than the range across values projected by all CCMs, suggesting that the earliest return dates across individual model projections are unlikely.

Our uncertainty estimates for the return dates are somewhat larger than those provided by chapter 9 of Eyring et al. (2010b). They applied the Time Series Additive Model (TSAM) approach by Scinocca et al. (2010) to the CCMVal-2 models and projected the Antarctic total ozone return date to occur around 2045–60. The TSAM method, however, does not constrain model weights by observations and instead bases model weighting on the assumption similar to that used in the uMMM method—namely, that the trends by individual models are random samples for the “true trend” [Scinocca et al. 2010, their Eq. (14)]. Given the dependence of model projections on the ability to simulate transport in present-day climate, this assumption is unlikely to be valid. Accounting for this dependence might increase the TSAM uncertainty beyond the limits suggested by chapter 9 of Eyring et al. (2010b).

The important advantage of our approach is that it provides an observational constraint on the future projections. This imposes additional requirements on the observational data quality, because the observational uncertainty may be an important source of the uncertainty of the MDER predictions. While here we found that the uncertainty of the predicted future ozone is dominated by the uncertainty of the statistical model, incorporating

observational uncertainty into the MDER predictions should be considered in future studies.

The MDER method is computationally cheap and is relatively easy to implement for estimations of ozone change in other regions, as well as of any other variable of interest, provided that strong correlations between present-day diagnostics and future change exist. For example, the method could be applied in climate change studies to variables such as global surface temperature. There, however, the search for informative process-oriented diagnostics may prove more challenging and requires a sophisticated process-oriented evaluation of climate and Earth system models, similar to the CCMVal evaluation of CCMs (Eyring et al. 2010b).

While the use of diagnostics considered here has proved successful to Antarctic total ozone, the application of the method to other regions may require consideration of additional diagnostics. For example, Strahan et al. (2011) have demonstrated an importance of tropical transport diagnostic based on the correlation between nitrous oxide (N_2O) and mean age for predicting tropical ozone. A diagnostics for nitrogen chemistry may also be useful since nitrogen chemistry is expected to become more important in the future when the concentration of halogens will decrease (Oman et al. 2010). The evolution of ozone in the upper and lower stratosphere is expected to be driven by very different processes (Oman et al. 2010) and therefore considering these two regions separately may prove useful. In addition, for ozone changes in the lower tropical stratosphere, process-oriented diagnostics that measure the tropical upwelling could be important. The challenges associated with the selections of diagnostics for ozone projections weighting in other regions and altitudes will be addressed in follow-up studies. Also, applying the MDER method to a larger set of CCM runs, after new CCM simulations become available, would allow obtaining more robust relationships between present-day diagnostics and future ozone changes.

Acknowledgments. We thank Mattia Righi and Robert Sausen (DLR, Germany), Markus Kunze (FUB, Germany), and two anonymous reviewers for their helpful comments on the manuscript. We acknowledge the Chemistry–Climate Model Validation (CCMVal) Activity for WCRP’s Stratospheric Processes and their Role in Climate (SPARC) project for organizing and coordinating the model data analysis activity and the British Atmospheric Data Center (BADC) for collecting and archiving the CCMVal model output. We thank the CCMVal model groups that are listed in Table 1 of this manuscript for providing their model data for the CCMVal Archive. We thank Greg Bodeker, the TOMS/SBUV

team, the MLS team, and the HALOE team for providing access to their data. The work of AYK is supported by Finnish Academy under Grants 132851, 140408, 259537, and 258701. DM received a travel grant from the German Academic Exchange Service (DAAD, Project 54777989). The work of VE was supported by the DLR Earth System Model Validation (ESMVal) project.

REFERENCES

- Abe, M., H. Shiogama, T. Nozawa, and S. Emori, 2011: Estimation of future surface temperature changes constrained using the future-present correlated modes in inter-model variability of CMIP3 multimodel simulations. *J. Geophys. Res.*, **116**, D18104, doi:10.1029/2010jd015111.
- Akiyoshi, H., T. Sugita, H. Kanzawa, and N. Kawamoto, 2004: Ozone perturbations in the Arctic summer lower stratosphere as a reflection of NO_x chemistry and planetary scale wave activity. *J. Geophys. Res.*, **109**, D03304, doi:10.1029/2003JD003632.
- , and Coauthors, 2009: A CCM simulation of the breakup of the Antarctic polar vortex in the years 1980–2004 under the CCMVal scenarios. *J. Geophys. Res.*, **114**, D03103, doi:10.1029/2007JD009261.
- Austin, J., and R. J. Wilson, 2010: Sensitivity of polar ozone to sea surface temperatures and halogen amounts. *J. Geophys. Res.*, **115**, D18303, doi:10.1029/2009JD013292.
- , —, F. Li, and H. Vomel, 2007: Evolution of water vapor concentrations and stratospheric age of air in coupled chemistry-climate model simulations. *J. Atmos. Sci.*, **64**, 905–921.
- , and Coauthors, 2010: The decline and recovery of total column ozone using a multi-model time series analysis. *J. Geophys. Res.*, **115**, D00M10, doi:10.1029/2010JD013857.
- Bodeker, G. E., H. Shiona, and H. Eskes, 2005: Indicators of Antarctic ozone depletion. *Atmos. Chem. Phys.*, **5**, 2603–2615.
- Boe, J. L., A. Hall, and X. Qu, 2009: September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nat. Geosci.*, **2**, 341–343, doi:10.1038/ngeo467.
- Bracegirdle, T. J., and D. B. Stephenson, 2012: Higher precision estimates of regional polar warming by ensemble regression of climate model projections. *Climate Dyn.*, **39**, 2805–2821, doi:10.1007/s00382-012-1330-3.
- Cionni, I., and Coauthors, 2011: Ozone database in support of CMIP5 simulations: Results and corresponding radiative forcing. *Atmos. Chem. Phys.*, **11**, 11267–11292, doi:10.5194/acp-11-11267-2011.
- Connolley, W. M., and T. J. Bracegirdle, 2007: An Antarctic assessment of IPCC AR4 climate models. *Geophys. Res. Lett.*, **34**, L22505, doi:10.1029/2007GL031648.
- Davison, A. C., 2003: *Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics, Vol. 11, Cambridge University Press, 738 pp.
- de Grandpré, J., S. R. Beagley, V. I. Fomichev, E. Griffioen, J. C. Mc-Connell, A. S. Medvedev, and T. G. Shepherd, 2000: Ozone climatology using interactive chemistry: Results from the Canadian Middle Atmosphere Model. *J. Geophys. Res.*, **105**, 26475–26491.
- Déqué, M., 2007: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global Planet. Change*, **57**, 16–26.

- Egorova, T., E. Rozanov, V. Zubov, E. Manzini, W. Schmutz, and T. Peter, 2005: Chemistry-climate model SOCOL: A validation of the present-day climatology. *Atmos. Chem. Phys.*, **5**, 1557–1576.
- Eyring, V., and Coauthors, 2005: A strategy for process-oriented validation of coupled chemistry–climate models. *Bull. Amer. Meteor. Soc.*, **86**, 1117–1133.
- , and Coauthors, 2006: Assessment of temperature, trace species and ozone in chemistry-climate model simulations of the recent past. *J. Geophys. Res.*, **111**, D22308, doi:10.1029/2006JD007327.
- , and Coauthors, 2007: Multimodel projections of stratospheric ozone in the 21st century. *J. Geophys. Res.*, **112**, D16303, doi:10.1029/2006JD008332.
- , and Coauthors, 2010a: Multi-model assessment of stratospheric ozone return dates and ozone recovery in CCMVal-2 models. *Atmos. Chem. Phys.*, **10**, 9451–9472, doi:10.5194/acp-10-9451-2010.
- , T. G. Shepherd, and D. W. Waugh, Eds., 2010b: Chemistry-climate model validation. WCRP-132, WMO/TD 1526, SPARC Rep. 5, 434 pp. [Available online at <http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no5/>.]
- , and Coauthors, 2013: Long-term ozone changes and associated climate impacts in CMIP5 simulations. *J. Geophys. Res.*, **118**, 5029–5060, doi:10.1002/jgrd.50316.
- Fioletov, V. E., G. E. Bodeker, A. J. Miller, R. D. McPeters, and R. Stolarski, 2002: Global ozone and zonal total ozone variations estimated from ground-based and satellite measurements: 1964–2000. *J. Geophys. Res.*, **107**, 4647, doi:10.1029/2001JD001350.
- Fomichev, V. I., A. I. Jonsson, J. de Grandpré, S. R. Beagley, C. McLandress, K. Semeniuk, and T. G. Shepherd, 2007: Response of the middle atmosphere to CO₂ doubling: Results from the Canadian Middle Atmosphere Model. *J. Climate*, **20**, 1121–1144.
- Garcia, R. R., D. R. Marsh, D. E. Kinnison, B. A. Boville, and F. Sassi, 2007: Simulation of secular trends in the middle atmosphere, 1950–2003. *J. Geophys. Res.*, **112**, D09301, doi:10.1029/2006JD007485.
- Garny, H., M. Dameris, and A. Stenke, 2009: Impact of prescribed SSTs on climatologies and long-term trends in CCM simulations. *Atmos. Chem. Phys.*, **9**, 6017–6031, doi:10.5194/acp-9-6017-2009.
- Grooß, J.-U., and J. M. Russell III, 2005: Technical note: A stratospheric climatology for O₃, H₂O, CH₄, NO_x, HCl and HF derived from HALOE measurements. *Atmos. Chem. Phys.*, **5**, 2797–2807.
- Hall, A., and X. Qu, 2006: Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys. Res. Lett.*, **33**, L03502, doi:10.1029/2005GL025127.
- Hegglin, M. I., and T. G. Shepherd, 2009: Large climate-induced changes in ultraviolet index and stratosphere-to-troposphere ozone flux. *Nat. Geosci.*, **2**, 687–691, doi:10.1038/ngeo604.
- Jöckel, P., and Coauthors, 2006: The atmospheric chemistry general circulation model ECHAM5/MESSy1: Consistent simulation of ozone from the surface to the mesosphere. *Atmos. Chem. Phys.*, **6**, 5067–5104, doi:10.5194/acp-6-5067-2006.
- Jourdain, L., S. Bekki, F. Lott, and F. Lefèvre, 2008: The coupled chemistry-climate model LMDz-REPROBUS: Description and evaluation of a transient simulation of the period 1980–1999. *Ann. Geophys.*, **26**, 1391–1413, doi:10.5194/angeo-26-1391-2008.
- Karpechko, A. Yu., N. P. Gillett, L. J. Gray, and M. Dall’Amico, 2010a: Influence of ozone recovery and greenhouse gas increases on Southern Hemisphere circulation. *J. Geophys. Res.*, **115**, D22117, doi:10.1029/2010JD014423.
- , —, B. Hassler, K. H. Rosenlof, and E. Rozanov, 2010b: Quantitative assessment of Southern Hemisphere ozone in chemistry-climate model simulations. *Atmos. Chem. Phys.*, **10**, 1385–1400.
- Knutti, R., 2010: The end of model democracy? *Climatic Change*, **102**, 395–404, doi:10.1007/s10584-010-9800-2.
- , G. Abramowitz, M. Collins, V. Eyring, P. J. Gleckler, B. Hewitson, and L. Mearns, 2010a: Good practice guidance paper on assessing and combining multi model climate projections. IPCC Expert Meeting on Assessing and Combining Multi Model Climate Projections Meeting Rep., 13 pp.
- , R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, 2010b: Challenges in combining projections from multiple climate models. *J. Climate*, **23**, 2739–2758.
- Lamarque, J.-F., D. E. Kinnison, P. G. Hess, and F. M. Vitt, 2008: Simulated lower stratospheric trends between 1970 and 2005: Identifying the role of climate and composition changes. *J. Geophys. Res.*, **113**, D12301, doi:10.1029/2007JD009277.
- Maraun, D., 2012: Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophys. Res. Lett.*, **39**, L06706, doi:10.1029/2012GL051210.
- Masson, D., and R. Knutti, 2011: Climate model genealogy. *Geophys. Res. Lett.*, **38**, L08703, doi:10.1029/2011GL046864.
- Miller, A. J., and Coauthors, 2002: A cohesive total ozone data set from SBUV(2) satellite system. *J. Geophys. Res.*, **107**, 4701, doi:10.1029/2001JD000853.
- Morgenstern, O., P. Braesicke, F. M. O’Connor, A. C. Bushell, C. E. Johnson, S. M. Osprey, and J. A. Pyle, 2009: Evaluation of the new UKCA climate-composition model. Part 1: The stratosphere. *Geosci. Model Dev.*, **1**, 43–57.
- , and Coauthors, 2010: Review of the formulation of present-generation stratospheric chemistry-climate models and associated external forcings. *J. Geophys. Res.*, **115**, D00M02, doi:10.1029/2009JD013728.
- Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772, doi:10.1038/nature02771.
- Newman, P. A., J. S. Daniel, D. W. Waugh, and E. R. Nash, 2007: A new formulation of equivalent effective stratospheric chlorine (EESC). *Atmos. Chem. Phys.*, **7**, 4537–4552, doi:10.5194/acp-7-4537-2007.
- Nissen, K. M., K. Matthes, U. Langematz, and B. Mayer, 2007: Towards a better representation of the solar cycle in general circulation models. *Atmos. Chem. Phys.*, **7**, 5391–5400, doi:10.5194/acp-7-5391-2007.
- Oman, L. D., and Coauthors, 2010: Multi-model assessment of the factors driving the ozone evolution over the 21st century. *J. Geophys. Res.*, **115**, D24306, doi:10.1029/2010JD014362.
- Pawson, S., R. S. Stolarski, A. R. Douglass, P. A. Newman, J. E. Nielsen, S. M. Frith, and M. L. Gupta, 2008: Goddard Earth Observing System chemistry-climate model simulations of stratospheric ozone-temperature coupling between 1950 and 2005. *J. Geophys. Res.*, **113**, D12103, doi:10.1029/2007JD009511.
- Perlwitz, J., S. Pawson, R. L. Fogt, J. E. Nielsen, and W. D. Neff, 2008: Impact of stratospheric ozone hole recovery on Antarctic climate. *Geophys. Res. Lett.*, **35**, L08714, doi:10.1029/2008GL033317.

- Pitari, G., E. Mancini, V. Rizi, and D. Shindell, 2002: Impact of future climate and sulfur emission changes on stratospheric aerosols and ozone. *J. Atmos. Sci.*, **59**, 414–440.
- Räisänen, J., L. Ruokolainen, and J. Ylhäisi, 2010: Weighting of model results for improving best estimates of climate change. *Climate Dyn.*, **35**, 407–422.
- Reichler, T., and J. Kim, 2008: How well do coupled models simulate today's climate? *Bull. Amer. Meteor. Soc.*, **89**, 303–311.
- Schraner, M., and Coauthors, 2008: Technical Note: Chemistry-climate model SOCOL: version 2.0 with improved transport and chemistry/microphysics schemes. *Atmos. Chem. Phys.*, **8**, 5957–5974, doi:10.5194/acp-8-5957-2008.
- Scinocca, J. F., N. A. McFarlane, M. Lazare, J. Li, and D. Plummer, 2008: Technical note: The CCCma third generation AGCM and its extension into the middle atmosphere. *Atmos. Chem. Phys.*, **8**, 7055–7074.
- , D. B. Stephenson, T. C. Bailey, and J. Austin, 2010: Estimates of past and future ozone trends from multi-model simulations using a flexible smoothing spline methodology. *J. Geophys. Res.*, **115**, D00M12, doi:10.1029/2009JD013622.
- Shibata, K., and M. Deushi, 2005: Partitioning between resolved wave forcing and unresolved gravity wave forcing to the quasi-biennial oscillation as revealed with a coupled chemistry-climate model. *Geophys. Res. Lett.*, **32**, L12820, doi:10.1029/2005GL022885.
- , and —, 2008a: Long-term variations and trends in the simulation of the middle atmosphere 1980–2004 by the chemistry-climate model of the Meteorological Research Institute. *Ann. Geophys.*, **26**, 1299–1326.
- , and —, 2008b: Simulation of the stratospheric circulation and ozone during the recent past (1980–2004) with the MRI Chemistry-Climate Model. CGER's Supercomputer Monograph Rep. 13, 154 pp.
- Son, S.-W., and Coauthors, 2010: Impact of stratospheric ozone on Southern Hemisphere circulation change: A multimodel assessment. *J. Geophys. Res.*, **115**, D00M07, doi:10.1029/2010JD014271.
- Stenke, A., M. Dameris, V. Grewe, and H. Garny, 2009: Implications of Lagrangian transport for coupled chemistry-climate simulations. *Atmos. Chem. Phys.*, **9**, 5489–5504.
- Stolarski, R. S., and S. Frith, 2006: Search for evidence of trend slowdown in the long-term TOMS/SBUV total ozone data record: The importance of instrument drift uncertainty. *Atmos. Chem. Phys.*, **6**, 4057–4065.
- Strahan, S. E., and Coauthors, 2011: Using transport diagnostics to understand chemistry climate model ozone simulations. *J. Geophys. Res.*, **116**, D17302, doi:10.1029/2010JD015360.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498.
- Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. Roy. Soc. London*, **A365**, 2053–2075.
- Teyssédre, H., and Coauthors, 2007: A new tropospheric and stratospheric Chemistry and Transport Model MOCAGE-Climat for multi-year studies: Evaluation of the present-day climatology and sensitivity to surface processes. *Atmos. Chem. Phys.*, **7**, 5815–5860, doi:10.5194/acp-7-5815-2007.
- Tian, W., and M. P. Chipperfield, 2005: A new coupled chemistry-climate model for the stratosphere: The importance of coupling for future O₃-climate predictions. *Quart. J. Roy. Meteor. Soc.*, **131**, 281–303.
- , —, L. J. Gray, and J. M. Zawodny, 2006: Quasi-biennial oscillation and tracer distributions in a coupled chemistry-climate model. *J. Geophys. Res.*, **111**, D20301, doi:10.1029/2005JD006871.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, doi:10.1256/qj.04.176.
- von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 494 pp.
- Waugh, D. W., and V. Eyring, 2008: Quantitative performance metrics for stratospheric-resolving chemistry-climate models. *Atmos. Chem. Phys.*, **8**, 5699–5713.
- Weigel, A. P., R. Knutti, M. A. Liniger, and C. Appenzeller, 2010: Risks of model weighting in multimodel climate projections. *J. Climate*, **23**, 4175–4191.
- Whetton, P., I. Macadam, J. Bathols, and J. O'Grady, 2007: Assessment of the use of current climate patterns to evaluate regional enhanced greenhouse response patterns of climate models. *Geophys. Res. Lett.*, **34**, L14701, doi:10.1029/2007GL030025.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 59, Academic Press, 627 pp.
- WMO, 2003: Scientific assessment of ozone depletion: 2002. World Meteorological Organization Global Ozone Research and Monitoring Project Rep. 47, 498 pp.
- , 2007: Scientific assessment of ozone depletion: 2006. World Meteorological Organization Global Ozone Research and Monitoring Project Rep. 50, 572 pp.
- , 2011: Scientific assessment of ozone depletion: 2010. World Meteorological Organization Global Ozone Research and Monitoring Project Rep. 52, 516 pp.