

Constraining Uncertainties in CMIP5 Projections of September Arctic Sea Ice Extent with Observations

DANIEL SENFLEBEN AND AXEL LAUER

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

ALEXEY KARPECHKO

Finnish Meteorological Institute, Helsinki, Finland

(Manuscript received 28 January 2019, in final form 29 October 2019)

ABSTRACT

In agreement with observations, Earth system models participating in phase 5 of the Coupled Model Intercomparison Project (CMIP5) simulate a decline in September Arctic sea ice extent (SIE) over the past decades. However, the spread in their twenty-first-century SIE projections is large and the timing of the first ice-free Arctic summer ranges from 2020 to beyond 2100. The uncertainties arise from three sources (internal variability, model uncertainty, and scenario uncertainty), which are quantified in this study for projections of SIE. The goal is to narrow uncertainties by applying multiple diagnostic ensemble regression (MDER). MDER links future projections of sea ice extent to processes relevant to its simulation under present-day conditions using data covering the past 40 years. With this method, we can reduce model uncertainty in projections of SIE for the period 2020–44 by 30%–50% (0.8–1.3 million km²). Compared to the unweighted multimodel mean, the MDER-weighted mean projects an about 20% smaller SIE and an earlier near-disappearance of Arctic sea ice by more than a decade for a high-greenhouse gas scenario. We also show that two different methods estimating internal variability in SIE differ by 1 million km². Regardless, the total uncertainties in the SIE projections remain large (up to 3.5 million km², with irreducible internal variability contributing 30%) so that a precise time estimate of an ice-free Arctic proves impossible. We conclude that unweighted CMIP5 multimodel-mean projections of Arctic SIE are too optimistic and mitigation strategies to reduce Arctic warming need to be intensified.

1. Introduction

Observations show that the ongoing warming of Earth caused the September Arctic sea ice extent (SIE) to shrink by almost 50% since the 1970s (Stroeve et al. 2012a). But not only has the ice area decreased, the sea ice has also become thinner and younger (i.e., the amount of multiyear ice has decreased rapidly; Fowler et al. 2004; Maslanik et al. 2011); about 70% of the

winter sea ice is now seasonal ice (i.e., first-year ice) (Kwok 2018). Thinner ice melts out more easily in summer, opening more ice-free areas and thus accelerating the Arctic warming (Holland and Bitz 2003; Stroeve et al. 2012a). Earth system models (ESMs) participating in phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012) simulate a further decrease in sea ice throughout the twenty-first century in all future scenarios that keep atmospheric greenhouse gas concentrations at current or higher levels. In the RCP8.5 scenario, where atmospheric CO₂ concentrations more than double by the end of the twenty-first century compared to 2000 (Riahi et al. 2011), almost all models project the Arctic to become ice-free in summer before 2100.

The year of near-disappearance of summer Arctic sea ice (YOD) is defined as the first year of a series of five consecutive years in which the minimum SIE drops below 1 million km² (Wang and Overland 2009). The large

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-19-0075.s1>.

Corresponding author: Axel Lauer, axel.lauer@dlr.de

DOI: 10.1175/JCLI-D-19-0075.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

spread in the projections stems from internal variability, different model structures (model uncertainty) and the uncertainty in future greenhouse gas scenarios (Kay et al. 2011; Stroeve et al. 2012b; Liu et al. 2013; Swart et al. 2015). Different methods have been applied to reduce uncertainties in sea ice projections, like large ensemble simulations, model selection, weighting, or constraining projections with observations (e.g., Massonnet et al. 2012; Liu et al. 2013; Melia et al. 2015; Jahn et al. 2016; Knutti et al. 2017; Screen and Williamson 2017; Sigmond et al. 2018). Recent studies highlight the positive effects on Arctic sea ice of restricting the global warming to 1.5°C (Screen and Williamson 2017; Jahn 2018; Screen 2018; Sigmond et al. 2018). For example, Jahn (2018) and Sigmond et al. (2018) agree that the probability of an ice-free Arctic summer under 1.5°C stabilized warming is around 2%, as opposed to 20%–35% at 2.0°C warming.

Following, for example, Melia et al. (2015), Lique et al. (2016), and Hodson et al. (2012), we use the method of Hawkins and Sutton (2009, 2011) to separate total multimodel projection uncertainty into its three components: internal variability, model uncertainty, and scenario uncertainty. The goal of this study is to reduce model uncertainty in projections of Arctic SIE and to give a more precise estimate of YOD. For this we use a set of 29 ESMs from 17 different institutes participating in CMIP5. Additionally, we compare the internal variability component from the Hawkins and Sutton (2009) method to an estimate that is based on the spread among a single-model large ensemble, similar to Jahn et al. (2016).

To narrow model uncertainty, we use the multiple diagnostic ensemble regression (MDER) method developed by Karpechko et al. (2013). This method uses statistical relationships between the projected future target variable (here September Arctic SIE) and historical model performance in terms of different process-oriented diagnostics. The results obtained are cross-validated to test for spurious relationships using a pseudoreality approach. MDER has proven its potential to reduce CMIP5 multimodel projection uncertainties in two previous studies: the spread in the projected return dates of Antarctic total column ozone could be reduced by over a decade (Karpechko et al. 2013) and the near-future position of the summer austral jet stream could be bias-corrected by 1.5° southward compared to the unweighted multimodel mean (Wenzel et al. 2016).

This paper is structured in the following way: section 2 describes the MDER method and how the three types of uncertainty can be quantified. Additionally, the diagnostics used in this study are introduced and the model

simulations and observations are described. In section 3, we apply MDER to constrain SIE projections and narrow model uncertainty. We also give our estimate of YOD for a scenario with a high greenhouse gas concentration. Section 4 closes with a summary and discussion.

2. Methods

a. Multiple Diagnostic Ensemble Regression

MDER was developed by Karpechko et al. (2013) and implemented into the Earth System Model Evaluation Tool, version 1.0 (ESMValTool; Eyring et al. 2016), by Wenzel et al. (2016). The ESMValTool is a community-developed open source software package aiming to facilitate the complex evaluation of ESMs. For this study, the ESMValTool was extended by additional diagnostics related to sea ice and MDER was adapted for applications to Arctic sea ice.

MDER is based on the correlation between selected process-oriented diagnostics applied to historical or present-day periods for which observations are available, and a future target variable. An iterative step-wise regression algorithm based on von Storch and Zwiers (1999) takes a set of preselected diagnostics as an input and selects a subset of these diagnostics to build a regression model that best predicts the future variable. The algorithm iteratively adds and removes diagnostics to and from the regression model until the regression sum of squares does not increase significantly, based on an F test with a significance level $p = 0.1$. The final regression model is then the linear combination of the selected process-oriented diagnostics that best predicts the future variable and is of the form

$$y = \beta_0 + \mathbf{X}^T \boldsymbol{\beta}, \quad (1)$$

where y is the estimated climate response (SIE), β_0 and $\boldsymbol{\beta}$ are the multiple regression parameters with $\boldsymbol{\beta}$ being a column vector of the size of the number of the selected diagnostics, and \mathbf{X} is the matrix of diagnostic values of the selected diagnostics. Using observational data for the selected diagnostics with the regression model yields a multidagnostic constraint. MDER then calculates model weights based on this constraint by extending the formula derived by Bracegirdle and Stephenson (2012) for a case with a single diagnostic to multiple diagnostics:

$$\mathbf{W} = [\mathbf{N}^T + (\mathbf{X}_0^T - \bar{\mathbf{X}}^T)(\mathbf{X}^T \mathbf{X} - \mathbf{X}^T \mathbf{1} \bar{\mathbf{X}}^T)^{-1} \times (\mathbf{X}^T - \mathbf{X}^T \mathbf{1} \mathbf{N}^T)]^T, \quad (2)$$

where $\mathbf{N} \equiv (\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}$ is a vector of a size equal to the number of models n and the value of all elements equal to n^{-1} , \mathbf{X}_0 is the vector of observed diagnostics, and

$\bar{\mathbf{X}} \equiv (\mathbf{N}^T \mathbf{X})^T$ is the vector of the multimodel-mean diagnostics. The weights are used to calculate a weighted multimodel mean from the CMIP5 model ensemble with the aim of reducing model uncertainty in the multimodel projection.

The diagnostic selection is one of the key strengths of MDER: compared to simply taking all subjective diagnostics (e.g., [Snape and Forster 2014](#); [Knutti et al. 2017](#); [Sanderson et al. 2017](#)), the step-wise regression algorithm reduces redundancy in selected diagnostics, which is the danger of including multiple diagnostics that effectively describe the same process. Another potential danger in using all diagnostics instead of only the MDER-selected diagnostics lies in overfitting caused by spurious relationships between historical diagnostics and the future variable ([Bracegirdle and Stephenson 2012](#)). To test for spurious relationships, the MDER results are cross-validated in a pseudoreality approach: since observations of future variables are naturally unavailable, we select one model at a time as reference (= pseudoreality) and benchmark the other models against this reference model by calculating the root-mean-square error (RMSE) as a measure of prediction uncertainty. Since there is no preferred reference model, each model is taken as pseudoreality once all of the remaining models are tested against it. The difference in RMSE between the unweighted multimodel mean (uMMM) and the MDER results reveals the potential of MDER to reduce uncertainty in the SIE projections and is a measure of the uncertainty of the MDER result.

MDER is based on the following assumptions that are similar to other regression-based approaches ([Bracegirdle and Stephenson 2012](#)):

- There is a linear relationship between the present-day mean state and trends and future change, which is similar in climate models and observations.
- The residuals from the regression fit are independently distributed.
- Climate model results and observations are interchangeable.
- The effects due to measurement errors are assumed to be negligible compared to other sources of uncertainty.

For further details on the MDER method, see [Karpechko et al. \(2013\)](#) and [Wenzel et al. \(2016\)](#).

The target variable in this study is the 2020–44 September-mean Arctic sea ice extent. SIE is derived from the gridded variable sea ice concentration (SIC, also known as “sea ice area fraction”), which describes the area fraction of each ocean surface grid cell that is covered with sea ice. Sea ice extent is defined as the total area of all grid cells in which $SIC \geq 15\%$. Here,

we define SIE as the September sea ice extent in the Arctic (60° – 90° N). The annual minimum Arctic sea ice extent typically occurs in September and thus September mean sea ice quantities are commonly used in literature, for example in analyses of the timing of an ice-free Arctic (e.g., [Massonnet et al. 2012](#); [Jahn et al. 2016](#); [Screen 2018](#); [Sigmond et al. 2018](#)).

b. Diagnostics

The set of process-oriented diagnostics preselected by the authors as input for the MDER method is listed in [Table 1](#) and is based on published literature on processes that are known to influence sea ice concentration. The selection is not meant to be an exclusive list, but considers many of the variables that previous work suggests are important for sea ice evolution. All diagnostics are applied to (present-day) data for the same time period (1979–2012); see also [section 2d](#).

The diagnostics are based on five variables. These include two different sea ice variables, historical SIE and sea ice thickness (SIT), which affect sea ice processes and projections (e.g., [Laxon et al. 2003](#); [Massonnet et al. 2018](#)). To account for freezing and melting processes, two temperature variables are included to represent the thermal influence on the ice from above and below (e.g., [Zhang et al. 2000](#); [Weeks 2010](#)): near-surface air temperature (TAS) and sea surface temperature (SST). Atmospheric surface pressure (PSL) is used as a proxy for the influence on ice drift due to atmospheric winds near the surface (e.g., [Thorndike and Colony 1982](#); [Spren et al. 2011](#)). For each of these variables, three different metrics are calculated over the whole historical time period (1979–2012): the climatological mean (indicated by $_c$), the trend ($_t$), and the interannual (“year-to-year”) variability ($_i$). The calculated diagnostic results from each model and observation/reanalysis are shown in the supplemental information (see Figs. S1–S5 in the online supplemental material).

c. Uncertainty estimation

In addition to the MDER method, we use the method of [Hawkins and Sutton \(2009, 2011\)](#) to quantify different sources of uncertainty in the CMIP5 multimodel projections for the twenty-first-century SIE (2006–2100): internal variability, model uncertainty, and scenario uncertainty. Numerous studies have applied this method to Arctic variables such as CMIP3 summer Arctic sea ice extent ([Lique et al. 2016](#)), CMIP3 Arctic temperature and precipitation ([Hodson et al. 2012](#)), and CMIP5 September Arctic sea ice thickness ([Melia et al. 2015](#)). In the following, we describe the sources of uncertainty and how they can be isolated, following [Hawkins and Sutton \(2009, 2011\)](#). It is important to note that no ensemble

TABLE 1. Overview of the 15 diagnostics used in this study, including the reanalyses or observations used to constrain the models and their respective references. All diagnostics were calculated over the years 1979–2012. “Summer” is the average of June–August.

Acronym	Diagnostic	Reanalysis or observation	Reanalysis or observational value
SIE_c	Climatological mean September Arctic sea ice extent	NSIDC-NT (Cavalieri et al. 1996; Walsh et al. 2015)	$(6.14 \pm 0.16) \times 10^6 \text{ m}^2$
SIE_t	September Arctic sea ice extent trend		$(-0.65 \pm 0.01) \times 10^6 \text{ m}^2 \text{ decade}^{-1}$
SIE_i	Interannual variability of September Arctic sea ice extent		$0.95 \times 10^6 \text{ m}^2$
SIT_c	Climatological mean September Arctic sea ice thickness	PIOMAS (Zhang and Rothrock 2003)	$1.10 \pm 0.05 \text{ m}$
SIT_t	September Arctic sea ice thickness trend		$-0.254 \pm 0.002 \text{ m decade}^{-1}$
SIT_i	Interannual variability of September Arctic sea ice thickness		0.28 m
TAS_c	Climatological mean summer Arctic surface air temperature	ERA-Interim (Dee et al. 2011)	$274.9 \pm 0.07 \text{ K}$
TAS_t	Summer Arctic surface air temperature trend		$0.26 \pm 0.01 \text{ K decade}^{-1}$
TAS_i	Interannual variability of summer Arctic surface air temperature		0.40 K
SST_c	Climatological mean summer Arctic sea surface temperature	HadISST (Rayner et al. 2003)	$273.79 \pm 0.04 \text{ K}$
SST_t	Summer Arctic sea surface temperature trend		$0.183 \pm 0.003 \text{ K}$
SST_i	Interannual variability of summer Arctic sea surface temperature		0.25 K
PSL_c	Climatological mean September Arctic surface pressure	ERA-Interim (Dee et al. 2011)	$1011.20 \pm 0.45 \text{ hPa}$
PSL_t	September Arctic surface pressure trend		$-0.89 \pm 0.04 \text{ hPa decade}^{-1}$
PSL_i	Interannual variability of September Arctic surface pressure		2.66 hPa

averages are taken in the calculation of the uncertainty estimates (as opposed to the diagnostic calculations; see section 2d for details).

1) INTERNAL VARIABILITY

To quantify the contribution of internal variability to the overall uncertainty of the model ensemble, each SIE projection from all models is individually fit with a fourth-order polynomial over the time period 2006–2100, using the least squares method and creating a smooth fit. We can write the predictions X from each model m , for scenario s and year t , as

$$X_{m,s,t} = x_{m,s,t} + i_m + \varepsilon_{m,s,t}, \quad (3)$$

where i is the reference value (year 2000 SIE), x the smooth fit, and ε the residual. The internal variability component V is computed from the multimodel mean of the variances of the residuals:

$$V = \sqrt{\frac{1}{N_m} \sum_m \sigma_{s,t}(\varepsilon_{m,s,t})}, \quad (4)$$

where $\sigma_{s,t}$ is the variance across all scenarios and across time and N_m is the number of models. Note that here,

because the variance is computed over all time values, V is constant in time by design. This is a limitation of this method, especially since interannual variability has been shown to increase as the ice thins. The internal variability obtained with this method is compared in section 3b to a more recent method by Jahn et al. (2016), who used a large ensemble of a single ESM to estimate internal variability.

2) MODEL UNCERTAINTY

Model uncertainty is estimated from the variance in the fits $x_{m,s,t}$ of each scenario. The multisenario mean model uncertainty $M(t)$ is calculated as follows:

$$M(t) = \sqrt{\frac{1}{N_s} \sum_s \sigma_m(x_{m,s,t})}, \quad (5)$$

where N_s is the number of scenarios.

3) SCENARIO UNCERTAINTY

The CMIP5 projections are divided into sets of scenarios of which each assumes a different amount of external forcing due to the emission of climate-relevant substances. More specifically, the scenarios represent four different representative concentration

TABLE 2. The 29 CMIP5 models that are used in this study. All available ensemble members (EM) were used for each model, and an ensemble mean was calculated for each model prior to MDER calculations. The numbering is in accordance with Fig. 2 and starts from 02 for technical reasons.

No.	Model	Institute	EM	Reference
02	BCC-CSM1.1	Beijing Climate Center (BCC)	1	Wu et al. (2014)
03	BCC-CSM1.1(m)		1	
04	CanESM2	Canadian Centre for Climate Modeling and Analysis (CCCma)	5	Arora et al. (2011)
05	CCSM4	National Center for Atmospheric Research (NCAR)	5	Gent et al. (2011)
06	CESM1-CAM5		3	
07	CMCC-CM	Centro Euro-Mediterraneo per I Cambiamenti Climatici	1	Vichi et al. (2011)
08	CMCC-CMS		1	
09	CNRM-CM5	Centre National de Recherches Météorologiques–Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CNRM-CERFACS)	1	Voltaire et al. (2012)
10	CSIRO-Mk3.6.0	Commonwealth Scientific and Industrial Research Organization/ Queensland Climate Change Centre of Excellence (CSIRO-QCCCE)	10	Rotstayn et al. (2012)
11	EC-EARTH	European EC-Earth consortium	2	Hazeleger et al. (2010)
12	FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS	1	Li et al. (2013)
13	GFDL CM3	National Oceanic and Atmospheric Administration – Geophysical Fluid Dynamics Laboratory (NOAA-GFDL)	3	Donner et al. (2011)
14	GFDL-ESM2G		1	Dunne et al. (2013)
15	GFDL-ESM2M		1	
16	GISS-E2-R	NASA Goddard Institute for Space Studies	2	Schmidt et al. (2014)
17	HadGEM2-AO	Met Office Hadley Centre	1	Martin et al. (2011)
18	HadGEM2-CC		1	
19	HadGEM2-ES		3	
20	INM-CM4	Russian Institute for Numerical Mathematics	1	Volodin et al. (2010)
21	IPSL-CM5A-LR	Institute Pierre-Simon Laplace (ISPL)	4	Dufresne et al. (2013)
22	IPSL-CM5A-MR		1	
23	IPSL-CM5B-LR		1	
24	MIROC5	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, and National Institute for Environmental Studies	3	Watanabe et al. (2011)
25	MIROC-ESM		1	
26	MIROC-ESM-CHEM		1	
27	MPI-ESM-LR	Max Planck Institute for Meteorology (MPI-M)	3	Giorgetta et al. (2013)
28	MPI-ESM-MR		3	
29	MRI-CGCM3	Meteorological Research Institute (MRI)	1	Yukimoto et al. (2012)
30	NorESM1-M	Norwegian Climate Centre (NorClim)	1	Iversen et al. (2013)

pathways (RCPs; van Vuuren et al. 2011) that assume a different radiative forcing (W m^{-2}) by the end of 2100: RCP2.6, RCP4.5, RCP6.0, and RCP8.5. The scenario uncertainty is calculated from the variance of the multimodel means over the RCP scenarios:

$$S(t) = \sqrt{\frac{1}{N_m} \sigma_s \left(\sum_m x_{m,s,t} \right)}. \quad (6)$$

For the estimate of scenario uncertainty, we only consider the scenarios RCP4.5 and RCP8.5, because the variable SIC is not available from all 29 models for RCP2.6 and 6.0, which were considered optional in CMIP5. This means that this study considers a medium and a high (“business-as-usual”) RCP scenario for the uncertainty estimates. However, we repeated the analysis with the 19 models for which output for

RCP2.6 was available. The results of this additional analysis are described in section 3b.

d. Models, observations, and reanalyses

For this study, an ensemble of 29 CMIP5 models from 17 different institutes or modeling centers (Table 2) is used. We could not use all (~40) CMIP5 models because not all models provided the required output for all five variables and the two scenarios used in this study. Even though not all CMIP5 models are strictly speaking Earth system models, we will refer to them as such for simplicity.

The historical time period used with the diagnostics (1979–2012) is determined by two constraints. The earliest start is the year 1979 because reliable observations or reanalyses of the required variables are not available for earlier times. The final year is a compromise between extending the CMIP5 historical model experiments ending

in 2005 as long as possible while keeping the impact of the RCP scenario chosen to extend the historical experiments to a minimum. Here, we extended the time period of the historical simulations up to the year 2012 using results from the corresponding RCP4.5 simulations. We did not use RCP8.5 results for extension of the historical runs since these were not available from all models and for all variables used. To minimize the influence of the actual scenario used for extension, the simulations were only extended until 2012 where differences between RCP4.5 and RCP8.5 are still small.

Our target period for future SIE is 2020–44. This 25-yr period is selected to 1) start sufficiently far from the end date of the historical experiments to increase the signal in SIE decline and 2) end before SIE from any model approaches zero, since correlations between historical diagnostics and future SIE may become spurious otherwise. Although we consider this target period as the most suitable period for the purpose of this study, we also tested the sensitivity of the results to the selected period. We found that using a longer period (2016–64) than the 25-yr period 2020–44 gives similar results, suggesting that our conclusions are not very sensitive to the exact target period. The target SIE projections used with the MDER regression algorithm are from RCP8.5 simulations.

For each model experiment, some models run different realizations (ensemble members) with slightly different initial conditions to sample internal variability. We selected only those ensemble members from each model that provide simultaneously all scenarios and variables that are used in this study. To account for the different ensemble sizes in the diagnostic calculations, we calculated for each model an ensemble average of the diagnostic values, which is then used in the further analyses. Thus, the multimodel metrics are not biased toward models with many ensemble members (see also Massonnet et al. 2012). Note that the ensemble average is taken after the diagnostic calculations of the 1979–2012 climatological mean, trends, and interannual variability (section 2b) since calculating the diagnostics on ensemble means would give incorrect estimates of, especially, interannual variability.

The observations and reanalyses for each diagnostic are listed in Table 1 and are provided as monthly means. In the following, we briefly describe the five datasets used in this study.

- We use satellite observations of SIC from the National Snow and Ice Data Center (NSIDC; Walsh et al. 2015). The product is available from 1978 to present and was processed with the NASA-Team retrieval algorithm

(NT; Cavalieri et al. 1996) from data of *Nimbus-7* SMMR and DMSP SSM/I and SSMIS passive microwave sensors. The spatial resolution is $25 \text{ km} \times 25 \text{ km}$.

- Sea ice thickness reanalyses are taken from the Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS; Zhang and Rothrock 2003), a coupled ice–ocean model forced with National Centers for Environmental Prediction (NCEP) reanalyses and assimilating observations of sea ice concentration and sea surface temperature. Satellite observations of SIT are spatially and temporarily limited (e.g., Kwok et al. 2009; Tilling et al. 2015). Since we need complete temporal and Arctic-wide coverage in this study, we use the PIOMAS reanalyses. PIOMAS was compared to SIT satellite observations and found to be a good estimate of the observed SIT in numerous studies (e.g., Lindsay and Zhang 2006; Schweiger et al. 2011; Laxon et al. 2013; Stroeve et al. 2014). PIOMAS reanalyses are commonly used in studies analyzing sea ice thickness (e.g., Melia et al. 2015; Dirkson et al. 2017; Labe et al. 2018).
- The Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST; Rayner et al. 2003) provides SST data. It is a global reanalysis product combining data from the Met Office Marine Data Bank (MDB), the Global Telecommunications System (GTS), and the Comprehensive Ocean–Atmosphere Dataset (COADS) and has a spatial resolution of $1^\circ \times 1^\circ$.
- For TAS and PSL, we use the European Centre for Medium-Range Weather Forecasts interim reanalysis (ERA-Interim; Dee et al. 2011) data. The data assimilation system used is based on the Integrated Forecast System, cycle 31r2 (IFS-Cy31r2), and is available from 1979 with monthly updates and an approximate horizontal resolution of 80 km.

3. Application of MDER to Arctic sea ice projections

MDER calculates a regression model from the historical diagnostics that best predicts future SIE (section 3a). The model weights obtained with this regression model are then applied to twenty-first-century SIE projections to narrow model uncertainty (section 3b) and improve the predictions of YOD (section 3c).

a. Diagnostic selection and regression model

Figure 1 shows the absolute correlation coefficients of all CMIP5 historical diagnostics (1979–2012; see Table 1) with future SIE (2020–44). The climatological mean September Arctic sea ice extent (SIE_c) is the diagnostic with by far the highest correlation coefficient ($r = 0.91$)

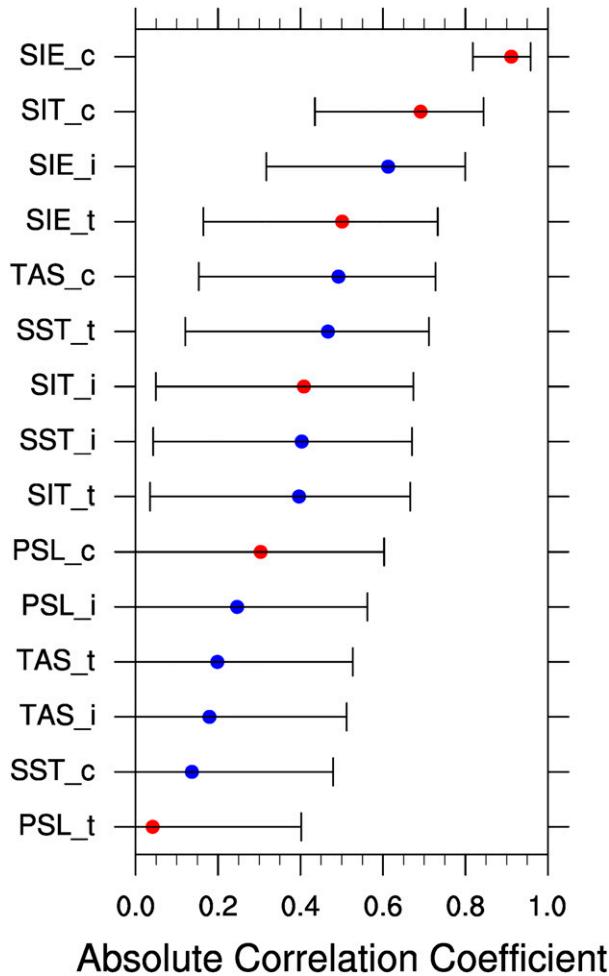


FIG. 1. Absolute correlation coefficients between the diagnostics (see Table 1) calculated from historical simulations (1979–2012) and the future SIE (2020–44) from simulations under the RCP8.5 scenario. The correlation coefficients have been calculated for the 29 CMIP5 models given in Table 2. Positive coefficients are shown in red and negative coefficients are shown in blue; error bars indicate the 95% confidence intervals around the correlation coefficients. Correlations with confidence intervals that include zero are not statistically significant.

and smallest uncertainty range. This means that biases in the historical mean sea ice extent dominate the projections of near-term future sea ice extent: models that simulate a small (large) SIE in the past simulate a small (large) SIE in the near future. Nine out of the 15 diagnostics show a statistically significant correlation with future SIE. After SIE_c, the climatological mean sea ice thickness (SIT_c) shows the largest correlation to future SIE with $r = 0.69$. Thinner ice breaks up and melts out more easily in summer (e.g., Bitz and Roe 2004; Kwok 2018; Petty et al. 2018) and thus results in a smaller future summer ice extent, but this relationship has a significantly larger uncertainty than SIE_c.

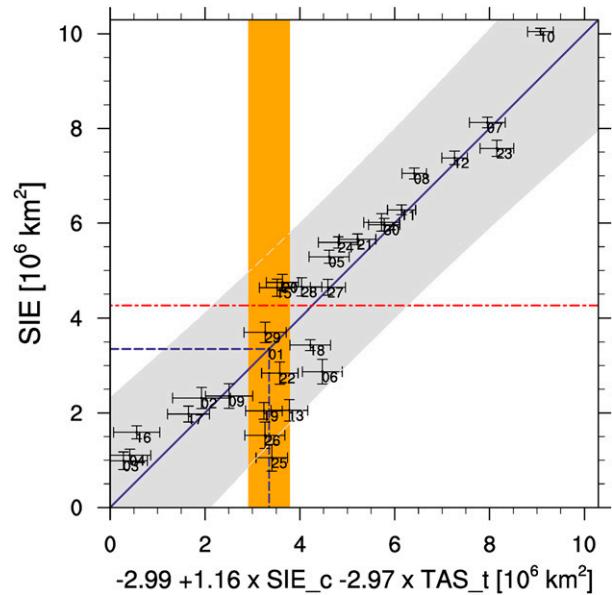


FIG. 2. Scatterplot showing the relationship between the future climatological mean (2020–44) September Arctic sea ice extent (RCP8.5) and the regression model ($-2.99 + 1.16 \times \text{SIE}_c - 2.97 \times \text{TAS}_t$) applied to the historical time period (1979–2012). The models are numbered according to Table 2 and error bars show one standard deviation of the mean values. The solid blue line is the least squares linear fit to the models and gray shading indicates the 95% prediction interval for the linear regression. The orange vertical bar shows one standard deviation around the observed climatological mean value (blue dashed vertical line) estimated from NSIDC and ERA-Interim observations. The dashed horizontal lines indicate the unconstrained multimodel-mean prediction (red) and the constrained MDER prediction (blue).

The other diagnostics with statistically significant correlations with future SIE include historical ice trends and interannual variability, historical SST trends and variability, and climatological air temperature. This means that projections of SIE are in particular influenced by biases in past sea ice conditions and Arctic surface temperatures, but not so much by biases in Arctic wind patterns as estimated by the proxy surface pressure. The values for each diagnostic from models and observations are given in Figs. S1–S5, and the scatterplots in Figs. S6 and S7 show the correlations between all diagnostics and future SIE.

The linear combination of multiple diagnostics (parsimonious regression model) that best predicts future SIE is calculated from the pool of diagnostics listed in Table 1 by the iterative stepwise regression algorithm. MDER selected two diagnostics for the regression model, namely SIE_c and TAS_t, with the resulting regression model equation $-2.99 + 1.16 \times \text{SIE}_c - 2.97 \times \text{TAS}_t$ (see Fig. 2). This means that future SIE is constrained by a linear combination of a bias correction in the historical simulations of SIE and the trend in Arctic surface

temperature. The coefficient of determination of the regression is $R^2 = 0.82$.

Figure S1 in the supplemental material reveals that the unconstrained values of SIE_c from most models show a positive bias compared to the NSIDC-NT observations (negative constant in the regression model equation). The positive sign of the SIE_c term reflects the positive correlation between historical and future SIE, and the negative sign of the TAS_t term shows that models that simulate a strong positive temperature trend in the Arctic in the historical simulations tend to simulate a smaller SIE in the future. Most models show a negative bias in TAS_t, suggesting that the reduction in the constrained SIE is partly due to the TAS_t diagnostic. This is not surprising since Arctic temperatures have risen about twice as much as the global average (Bellucci et al. 2015), which can be partly explained by the loss of sea ice (Screen et al. 2013).

The selected diagnostics also show that the linear combination of diagnostics that best predicts the target variable does not necessarily contain the diagnostics with the highest correlation to the target variable, since the correlation coefficient between TAS_t and future SIE is not statistically significant (see Fig. 1). MDER has selected TAS_t as a predictor despite the weak relationship between TAS_t and future SIE because MDER is a stepwise approach, and a variable is added at a given step if it explains a significant fraction of residual variance from the previous step. In other words, the variance explained by a combination of SIE_c and TAS_t is significantly larger in terms of the F test than that explained by SIE_c only. Therefore, it is not necessary that there is a strong relation between TAS_t and future SIE because such a relation may be masked, for example, by a stronger relation between SIE_c and future SIE. In fact, a sensitivity test without TAS diagnostics resulted in a regression model that only included SIE_c and the strength of the constraint was roughly halved. This suggests that the two diagnostics have a roughly similar contribution to the constraint (not shown).

Applying the regression model equation to the observed SIE_c and TAS_t (blue dashed lines in Fig. 2) we can constrain the CMIP5 projections of SIE. This constraint reduces the projected multimodel climatological mean SIE (red dashed line) by 0.9 million km² (i.e., from 4.25 to 3.35 million km²). This means that applying MDER to SIE results in an over 20% smaller Arctic sea ice extent between 2020 and 2044 compared to the unconstrained CMIP5 multimodel mean.

To estimate the impact of internal variability on the 1979–2012 historical Arctic SIE and TAS trends used in our analysis (Kay et al. 2011; Swart and Fyfe 2013; Swart et al. 2015), Fig. 3 shows trend distributions that were calculated

over the whole 34-yr time period from the results of the 29 CMIP5 models and from a large initial-condition ensemble obtained with the Community Earth System Model (CESM LE; Kay et al. 2015; see also section 3b; Jahn et al. 2016). The assumption is that the spread in large initial-condition ensembles (round-off level perturbation) represents the internal variability of the climate system within the context of a particular climate model. Comparing the standard deviation of the CMIP5 trends to the one obtained from the CESM LE gives an estimate of the contribution of internal variability to the spread in the CMIP5 SIE trends. For SIE, the standard deviation of the CESM LE trends (0.21 million km² decade⁻¹) is slightly smaller than one of the CMIP5 trends (0.27 million km² decade⁻¹), suggesting that internal variability is an important but not the only factor determining the spread in the CMIP5 SIE trends. The standard deviations in the TAS trends show a qualitatively similar behavior (Fig. 3, bottom).

Unfortunately, the method often used in literature to account for contribution of internal variability (Fyfe et al. 2013; Swart and Fyfe 2013; Swart et al. 2015) could not be applied here, since it requires models with multiple ensemble members. Here, we can use only one to very few ensemble members per model since we need the exact same ensemble members to be available for all diagnostics. For the five variables required here only one ensemble member is available for most of the models (see number of ensemble members from each model in Table 2).

b. Uncertainty estimation and the potential to narrow model uncertainty

To test if the regression model is overfitted and to investigate whether the MDER method actually gives a more precise estimate of future SIE than the unweighted multimodel mean, we cross-validate the results using a pseudoreality approach. This approach selects each model in turn as a reference by which to benchmark the other models (see section 2a). Figure 4 shows the RMSE for all pseudorealities considered, both for the unweighted multimodel-mean (uMMM) and the MDER results. The results show that RMSE_{MDER} (0.93 million km²) is about 62% smaller than RMSE_{uMMM} (2.48 million km²). Similarly, the 25th–75th percentiles of the error ensembles are more than halved from 1.8 million km² (uMMM) to 0.8 million km² (MDER). The RMSE of the uMMM prediction basically reflects the intermodel spread in the projections of the 2020–44 mean SIE and is largely influenced by cases where the pseudoreality is an outlier model. In contrast, with MDER we use the information of the historical SIE (1979–2012) from the pseudoreality (“reference”) model and use the other

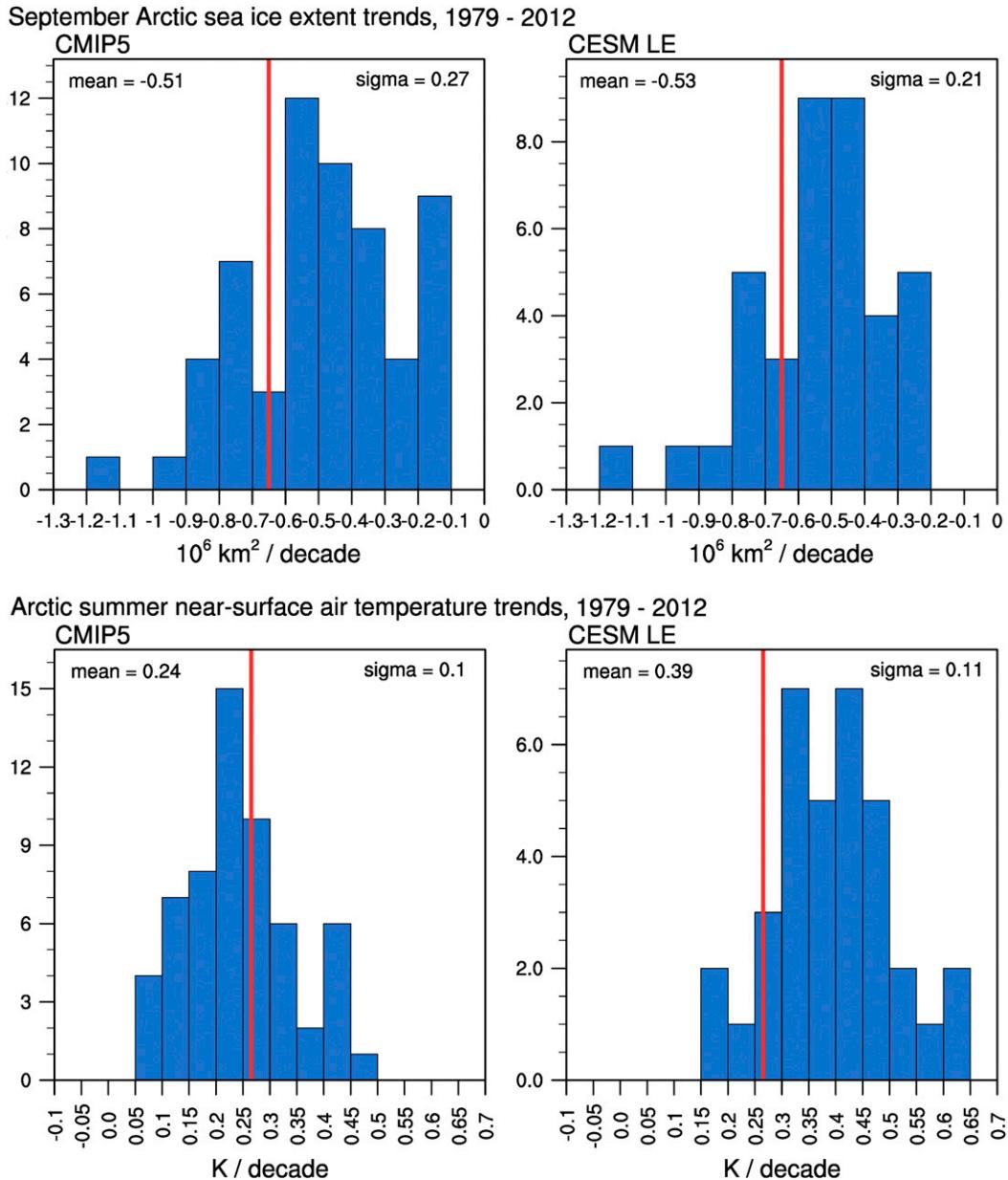


FIG. 3. Frequency distributions of (top) SIE_t and (bottom) TAS_t calculated (left) from all 29 CMIP5 models and (right) from the 38-member CESM LE over the time period 1979–2012. The red vertical lines represent the trends from (top) NSIDC-NT observations or (bottom) ERA-Interim data. The values for the mean and the variance (σ) of each trend distribution are given at the top of each panel. The top-left panel is similar to Fig. 2c in Swart et al. (2015).

models to estimate the change in mean SIE between 1979–2012 and 2020–44. The reduction of RMSE in MDER in comparison to that in uMMM indicates a predictive skill of MDER and suggests that overfitting is not an issue in our calculations.

To assess the full potential of MDER for reducing the prediction uncertainty, we apply the model weights calculated by MDER to obtain a weighted multimodel

mean. We hereby combine the MDER approach with the method introduced by Hawkins and Sutton (2009) that separates total prediction uncertainty into the three components internal variability, model uncertainty, and scenario uncertainty (see section 2c). Figure 5a shows time series of the three sources of uncertainty in unweighted SIE projections. The dominant source of uncertainty throughout the whole time period is model

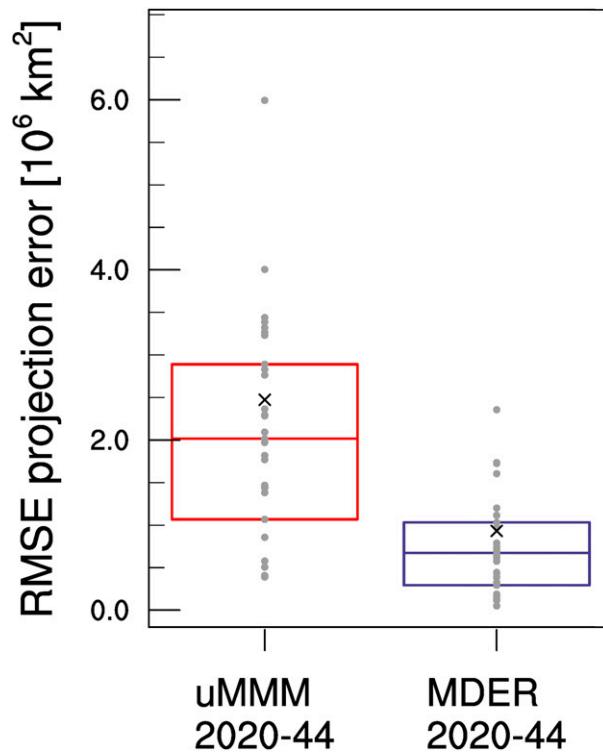


FIG. 4. RMSE differences between the multimodel-mean future SIE and the pseudoreality of future SIE estimated for different pseudorealities (gray dots). The RMSE is calculated for two cases: the uMMM and the MDER methods. The crosses show the RMSE for each case and the boxes (red: uMMM; blue: MDER) give the 25th–75th percentiles across the error ensembles. The horizontal middle line inside each box indicates the median of the model ensemble.

uncertainty, averaging around ± 2.5 million km^2 , which is due to the large intermodel spread in SIE projections between different CMIP5 models (Kay et al. 2011; Stroeve et al. 2012b; Liu et al. 2013; Swart et al. 2015). Differences in SIE due to different RCP scenarios start to emerge after 2020 and become increasingly larger in the second half of the twenty-first century. Thus, the scenario uncertainty increases with time and becomes more important than internal variability after 2050. Scenario uncertainty, however, remains smaller than model uncertainty until at least 2100. It is important to note that only simulations of the scenarios RCP4.5 and RCP8.5 were used here, since not enough models have run the RCP2.6 and RCP6.0 scenarios (see section 2c). As a sensitivity test, the analyses have been repeated with 19 CMIP5 models for which the three scenarios RCP2.6, RCP4.5, and RCP8.5 were available (not shown), and model uncertainty still remained the dominant source over the whole time period in this method.

The internal variability component is assumed constant in time by Hawkins and Sutton (2009), which is a

limiting factor (see also section 2c). With this method, it has a value of ± 0.6 million km^2 and is the least important source of uncertainty after 2050. Other studies find a larger contribution of internal variability in other variables (Melia et al. 2015; Jahn et al. 2016; Lique et al. 2016; see also section 3b). This raises the question whether the method of Hawkins and Sutton (2009) really captures the internal variability to its full extent. Jahn et al. (2016) estimated internal variability in Arctic SIE projections from a large ensemble of simulations with the Community Earth System Model, as did Swart and Fyfe (2013) for Antarctic sea ice area trends.

To compare our results to the Jahn et al. (2016) method, we repeated their approach here. Figure 6 shows the results for the 38-member large ensemble (CESM LE; gray), which was forced with RCP8.5, and the 15-member medium ensemble (CESM ME; blue), which was forced with RCP4.5. The ensemble spreads were calculated as the standard deviation at each time step across all ensemble members. The two ensemble spreads are similar to each other (around 1.5 million km^2) until more and more ensemble members of CESM LE reach an SIE of 0, which leads the CESM LE ensemble spread to approach 0 as well. If we assume that the ensemble spread represents internal variability, the Jahn et al. (2016) estimate is more than twice the internal variability from the Hawkins and Sutton (2009) method. The internal variability estimated by Jahn et al. (2016) quantifies variability as produced by a particular model for a particular scenario and could therefore be different for other models. In contrast, the estimate from Hawkins and Sutton (2009) is obtained by performing a statistical fit to the simulated time series and quantifying internal variability as the residual from the smooth fit across a number of models and scenarios. These are obviously different metrics and the large difference in estimated values indicates that a true estimate of SIE internal variability that accounts for multiple models is not yet possible. Regardless, it appears that the Hawkins and Sutton method may underestimate internal variability in the case of September SIE.

Model weighting has the potential to narrow uncertainties in climate model projections (Hawkins and Sutton 2009, 2011; Melia et al. 2015; Knutti et al. 2017; Eyring et al. 2019). MDER can be used to produce model weights by calculating the regression of historical diagnostics and future SIE (see sections 3a and 2a). Note that these MDER weights are different from the classical performance-based model weights, since they are not directly proportional to model biases and can be negative (Bracegirdle and Stephenson 2012). Table S1 lists all models and their weights. We recalculate the three types of uncertainty weighting the models with the MDER weights. By applying the weights calculated

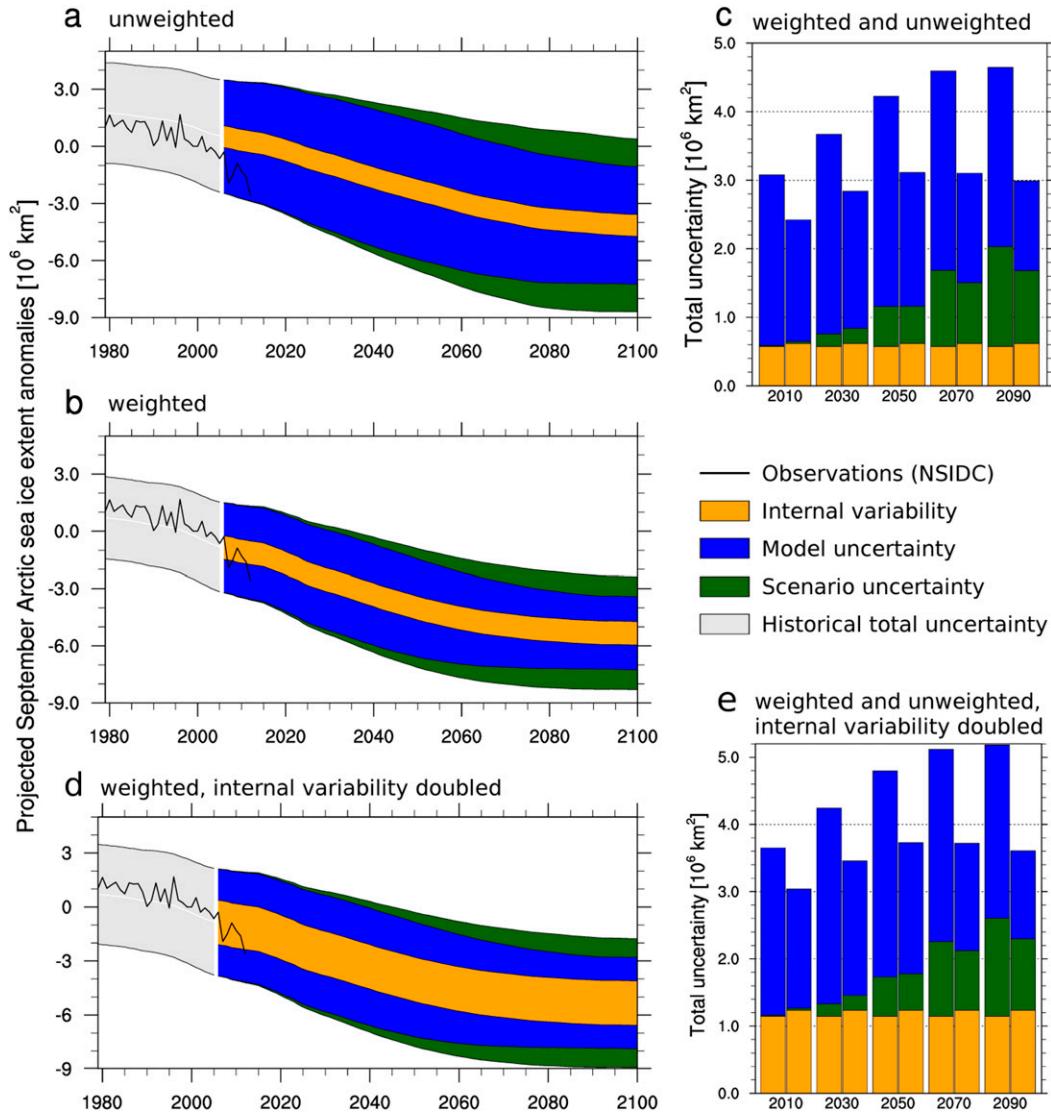


FIG. 5. Total uncertainty in CMIP5 SIE projections separated into the three fundamentally different sources of uncertainty: internal variability (orange), model uncertainty (blue), and scenario uncertainty (green) estimated from the 29 CMIP5 models listed in Table 2 (RCP4.5 and RCP8.5) following Hawkins and Sutton (2009). (a),(b) Time series (1979–2100) of the sources of uncertainty from unweighted SIE projections and from SIE projections weighted with the model weights generated by MDER, respectively. Additionally, the total uncertainty of the historical simulations is shown (gray shading), together with SIE observations (NSIDC; black line). All time series are 10-yr running averages and calculated as anomalies with respect to the observed September 2000 Arctic sea ice extent (6.2 million km^2 ; NSIDC). (c) Sources of uncertainty in five different time steps both for the unweighted (left bars) and weighted case (right bars). (d),(e) As in (b) and (c), respectively, but for the internal variability component, which is doubled in each time step based on a rough estimate with a single ESM large ensemble.

from the target period 2020–44 to the whole projection period we need to assume that the same processes selected by MDER are similarly important for SIE projections during the other periods, because we cannot apply MDER to a longer time period given the spurious correlations occurring when more and more models reach a sea ice extent close or equal to zero (see section 2a). The results of the weighting on the types of uncertainty are

shown in Fig. 5b. Compared to the unweighted case (Fig. 5a) the weighted model uncertainty is considerably smaller, while internal variability is (as expected) not affected by weighting.

The bar charts in Fig. 5c show a more quantitative estimation of this uncertainty reduction. The relative reduction of model uncertainty becomes larger with time: from 30% in 2010 to 50% in 2090. This large uncertainty

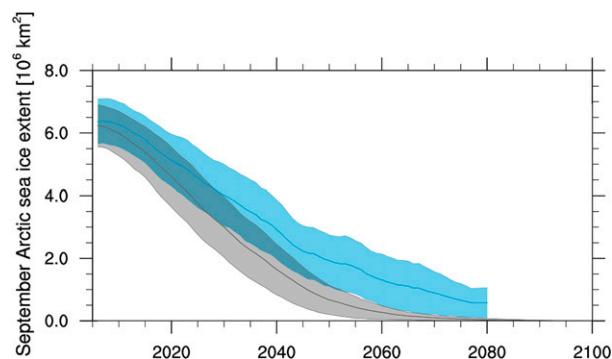


FIG. 6. Ensemble spread (10-yr running-mean standard deviation) in September Arctic sea ice extent simulations performed with CESM, for the 38-member large ensemble (CESM LE; 2006–2100) forced with RCP8.5 (gray) and the 15-member medium ensemble (CESM ME; 2006–80) forced with RCP4.5 (blue). Thick lines represent the respective ensemble means. For details on CESM LE and CESM ME, see Jahn et al. (2016).

reduction is mainly due to a bias correction: the MDER weights are calculated using the two diagnostics SIE_c and TAS_t (section 3a). For SIE_c, the models with the largest biases have the smallest weights, reducing the model spread. Comparing the diagnostic values for SIE_c (Fig. S1) and TAS_t (Fig. S3) reveals that the model with the largest positive SIE_c bias (CSIRO-Mk3.6.0) also has an unrealistically small temperature trend (TAS_t), resulting in a small SIE_t, and thus gets the smallest weight (Table S1). This example illustrates how model weighting can reduce model uncertainty. In the weighted results, the contribution of model uncertainty to the total uncertainty becomes smaller with increasing projection time (from over 80% in 2010 to less than 40% in 2100), but since scenario uncertainty is also slightly reduced by weighting after 2065, model uncertainty remains the most important source of uncertainty in SIE projections throughout the twenty-first century.

To account for a possible underestimation by the Hawkins and Sutton method (as discussed above), Figs. 5d and 5e show a rough estimate of this uncertainty by doubling the internal variability component, as the Jahn et al. method suggests. Here, internal variability is more important than scenario uncertainty until 2100, accounting for 20%–30% of the total uncertainty, and becomes almost equally as important as the weighted model uncertainty toward the end of the twenty-first century. However, even considering a doubled internal variability component, model uncertainty—despite weighting—remains the dominant source of uncertainty throughout this century.

c. Weighting SIE projections and estimating YOD

We now apply the results of MDER to reduce the uncertainty in the multimodel projection of future SIE

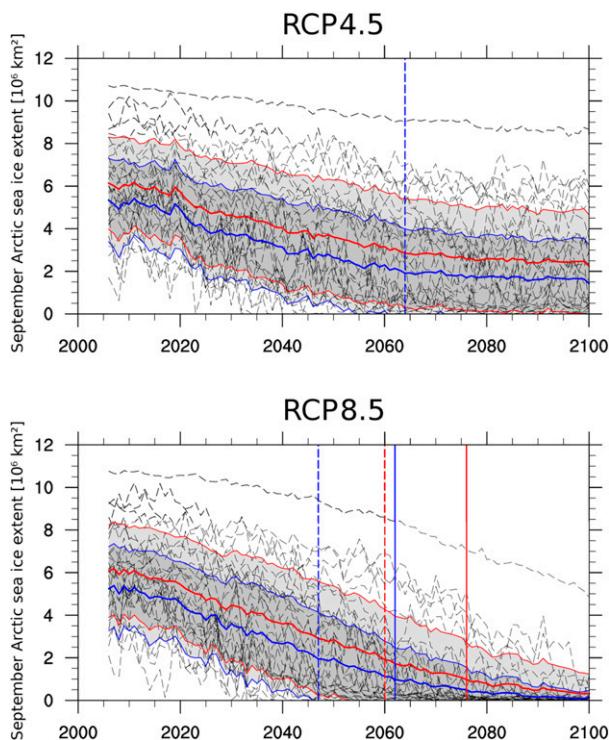


FIG. 7. Time series (2006–2100) of future SIE in the (top) RCP4.5 and (bottom) RCP8.5 scenarios: ensemble means of the CMIP5 models listed in Table 2 (black dashed curves), the uMMM (red curves), and the multimodel mean weighted by MDER (wMMM; blue curves). The gray shading shows the standard deviation of the CMIP5 ensemble, both unweighted (enclosed in thin red lines) and weighted (enclosed in thin blue lines). The vertical lines indicate the first time the multimodel means drop below 2 million km² (YO2; dashed) or the YOD (solid).

by weighting the models with the MDER weights (Fig. 7). For both RCP scenarios, the weighted multimodel-mean SIE (wMMM) is about 1 million km² smaller than uMMM until YOD is reached in the wMMM projection of SIE. The smaller SIE from the wMMM suggests a more pessimistic future for Arctic sea ice than from uMMM, implying an earlier disappearance of the ice. We calculate two thresholds to quantify the differences between uMMM and wMMM: the first year in which the multimodel-mean SIE drops below 2 million km² (YO2) and the multimodel-mean YOD. For RCP4.5 the YOD comes after the year 2100 in both cases but YO2 happens much earlier in case of wMMM compared with uMMM: whereas YO2 happens in uMMM after 2100, YO2 is reached in wMMM in the year 2064 and thus at least 37 years earlier. In RCP8.5 both YOD and YO2 are also earlier in the wMMM: 2047 instead of 2060 in the case of YO2 and 2062 instead of 2076 for YOD, meaning that when applying MDER to reduce the model uncertainty nearly ice-free conditions in the summertime Arctic in the RCP8.5

scenario are reached more than a decade earlier compared with the unweighted CMIP5 multimodel mean.

Weighting the multimodel mean has similar effects to weighting model uncertainty (section 3b): outlier models are down-weighted and since more models on average show a positive SIE_c bias than negative bias, and since those models also project a larger future SIE than the others, MDER's bias correction predicts a smaller future multimodel-mean SIE. Similarly, the weighted standard deviation across the CMIP5 projections is about 0.5 million km² smaller than the unweighted standard deviation until wMMM reaches Y02. Note that here the multimodel standard deviation is a different measure than model uncertainty [Eq. (5)] and that the standard deviation is estimated individually for each scenario, which is why the reduction in model uncertainty by weighting (section 3b) is larger than the reduction in the multimodel standard deviation. Here, the weighted standard deviation is still large, amounting up to 5 million km².

Since YOD is clearly scenario-dependent, it provides a measure of the anthropogenically forced response of the climate system. As seen in section 3b, even without model uncertainty the contribution of internal variability is too large to precisely predict YOD. However, the results obtained here strongly suggest an earlier near-disappearance of Arctic sea ice than estimated from an unweighted multimodel mean confirming similar findings by Massonnet et al. (2012).

4. Summary and discussion

We applied the multiple diagnostic ensemble regression method (e.g., Karpechko et al. 2013) to projections of September Arctic sea ice extent (SIE) from 29 CMIP5 models (Table 2). From a set of 15 process-oriented diagnostics (Table 1) known to influence Arctic sea ice, the stepwise regression algorithm within MDER selected a linear combination (parsimonious regression model) of two of these diagnostics that best predict the projected future SIE: the climatological mean SIE (SIE_c) and the trend in Arctic near-surface temperature (TAS_t). Applying this regression model to observations of these two quantities, MDER is used to constrain the SIE projections. Compared to unweighted multimodel-mean projections, MDER results in smaller ice extents by about 1 million km² and an earlier year of near-disappearance of Arctic sea ice (YOD) by more than a decade in the RCP8.5 scenario. By cross-validating the MDER results with a pseudoreality approach we have confirmed that our regression model has not been overfitted. We caution that the predictions of future SIE and YOD provided in this study are affected by unavoidably subjective

preselection of diagnostics, and they can be refined once additional emergent constraints become available. Finding emergent constraints (e.g., Bracegirdle and Stephenson 2012; Borodina et al. 2017) for Arctic sea ice is an ongoing scientific topic. The study showed that MDER has the potential to increase the understanding of which variables mainly control the model uncertainty in the projected Arctic sea ice evolution, and we would like to encourage further process-oriented studies.

Using the method introduced by Hawkins and Sutton (2009, 2011) to separate the total uncertainty in projections of SIE in its three components (internal variability, model uncertainty, and scenario uncertainty), we have identified model uncertainty as the largest source of uncertainty in SIE projections throughout the twenty-first century. We have shown that weighting the models based on MDER's regression model greatly reduces model uncertainty in projections of SIE by 30%–50%. This is possible because the weighting in MDER uses information from two diagnostics allowing the projections to be constrained with observations. Hodson et al. (2012) found that uncertainties in the twentieth-century mean state of the Arctic climate are a major source for uncertainties in Arctic climate projections and that observational constraints can greatly improve the precision of Arctic climate projections, which was also demonstrated here.

However, the other two sources of uncertainty—internal variability and scenario uncertainty—cannot be significantly reduced by applying the MDER method. By analyzing large ensembles of the CESM model, Jahn et al. (2016) found that these two types of uncertainty make up for a combined prediction uncertainty in YOD of about 25 years. We also used this approach here and obtained values for internal variability that were about twice as high as the one obtained with the method by Hawkins and Sutton (2009). While the CESM analysis only quantifies internal variability within the context of a single model, the discrepancy with the curve fitting method of Hawkins and Sutton (2009) suggests that the latter method might underestimate the actual internal variability in SIE. Assuming an internal variability component that is twice as large as the one suggested by Hawkins and Sutton (2009), internal variability becomes as large as the weighted model uncertainty (1.3 million km²). In any case, and despite the lack of a convincing estimate of internal variability, the internal variability is too large for a precise prediction of YOD. However, MDER strongly indicates an earlier disappearance of Arctic sea ice compared to the unconstrained CMIP5 multimodel mean.

For other sea ice variables, two studies find slightly different contributions of different sources of uncertainty to the total uncertainty. Melia et al. (2015) find that model

uncertainty in September sea ice thickness projections only becomes the dominant source of uncertainty after 2022 (before that, the dominant source is internal variability) and that it accounts for maximally 70% of the total uncertainty in sea ice thickness projections, whereas for sea ice extent we find that model uncertainty is the dominant source throughout the twenty-first century, regardless of the method estimating internal variability. For the change in CMIP3 summer (July–September) Arctic sea ice extent, [Lique et al. \(2016\)](#) found internal variability to be more important than model uncertainty until 2020, with a model uncertainty of between 1 and 3 million km². This is similar to the model uncertainty that we find (around 2.5 million km²). However, our estimate of the internal variability component (0.6 million km²) in SIE projections is smaller than in [Lique et al. \(2016\)](#), which is around 1 million km². This could be due to the Hawkins and Sutton method not capturing internal variability to its full extent as discussed above, or due to the use of different model results (CMIP3 vs CMIP5). To further analyze and quantify the contribution of internal variability to the diagnostics used in MDER—especially the trend diagnostics—multiple ensemble members from each model are required. With larger ensembles available in CMIP6, this study can be repeated and internal variability contributions estimated.

This study demonstrates the potential of MDER to reduce model uncertainty in multimodel projections of Arctic SIE and constrain the prediction of YOD. Its strength is partly based on the strong relationship between past and future climatological mean SIE, with a correlation value of $r = 0.91$ ([Fig. 1](#) and [Fig. S6](#)), which is a well-known emergent constraint (e.g., [Boé et al. 2009](#); [Bracegirdle et al. 2015](#); [Borodina et al. 2017](#)). Other studies ([Massonnet et al. 2012](#); [Liu et al. 2013](#); [Melia et al. 2015](#); [Knutti et al. 2017](#); [Screen and Williamson 2017](#)) also use this relationship to constrain projections of Arctic sea ice using different methods such as, model selection ([Massonnet et al. 2012](#)) or bias correction ([Melia et al. 2015](#)). All studies estimate a YOD in the 2040s or 2050s, similar to our results. [Liu et al. \(2013\)](#) combine model selection based on model performance and constraining the model biases with observations. They estimate a YOD between 2054 and 2058. In view of our results with still large uncertainties despite weighting, it is surprising that [Liu et al. \(2013\)](#) find such a narrow estimate for YOD. It may stem from confining their analyses to 5-yr sliding windows.

[Knutti et al. \(2017\)](#) introduce a weighting scheme that accounts for both model performance and model interdependence using sea ice and surface temperature diagnostics, resulting in a faster expected decline of SIE than in the unweighted case. They raise the concern of

selecting the right diagnostics, but argue that picking unsuitable diagnostics will assign random weights to the models and therefore will not influence the results, as long as the model ensemble is large enough. This is one of the key strengths of our approach: the stepwise regression algorithm in MDER filters spurious relationships and only retains the most suitable diagnostics. The strength of the MDER method to constrain the projections can be further improved by finding additional emergent constraints that can be used as diagnostics.

In addition to historical Arctic SIE, the MDER results are also based on historical Arctic temperature trends. Using CMIP3 results, [Mahlstein and Knutti \(2012\)](#) found a nearly linear relationship between Arctic sea ice area and both global and Arctic mean temperature. Using this relationship, they predict the future evolution of Arctic sea ice extent under different stabilized warming scenarios, based on observations and considering internal variability from the models. Their findings suggest that a permanently ice-free Arctic could likely be prevented under 2°C or less stabilized warming.

The results of the different approaches mentioned above are in general agreement with our findings. However, despite all the great community efforts to reduce uncertainties in the projections of Arctic climate, model uncertainty remains too large to give a precise estimate of the timing of the first near-disappearance of Arctic sea ice. In fact, considering the large contribution of irreducible internal variability, it is likely that these exact predictions are impossible. Yet, all of these studies, including this study, hint at a more pessimistic outlook for Arctic sea ice. Recent studies found that reducing Arctic warming to 1.5°C instead of 2°C by the end of the twenty-first century can greatly reduce the number of occurrences of an ice-free Arctic ([Jahn 2018](#); [Screen 2018](#); [Sigmond et al. 2018](#)). It is therefore imperative to further pursue and enhance global mitigation strategies to limit climate change.

Acknowledgments. This work was funded by the European Union's Horizon 2020 Framework Programme for Research and Innovation "Coordinated Research in Earth Systems and Climate: Experiments, kNowledge, Dissemination and Outreach (CRESCENDO)" project under Grant Agreement 641816. Additional funding was received by the Advanced Earth System Model Evaluation for CMIP (EVal4CMIP) project funded by the Helmholtz Association of German Research Centers. A.K. is supported by the Academy of Finland (Grants 286298 and 319397). The authors acknowledge the World Climate Research Program's (WCRP's) Working Group on Coupled Modeling (WGCM), which is responsible for CMIP, and thank the modeling groups ([Table 2](#)) for

providing their model output. We thank Veronika Eyring (DLR) and Marika Holland (National Center for Atmospheric Research) for their contributions to the study, François Massonnet (Université Catholique de Louvain) for fruitful discussions on an earlier version of the manuscript, and Mattia Righi, Manuel Schlund, and Sabrina Zechlau (DLR) for technical support with the ESMValTool. The authors also thank the three anonymous reviewers and editor James Screen for their helpful comments.

REFERENCES

- Arora, V. K., and Coauthors, 2011: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.*, **38**, L05805, <https://doi.org/10.1029/2010GL046270>.
- Bellucci, A., and Coauthors, 2015: Advancements in decadal climate predictability: The role of nonoceanic drivers. *Rev. Geophys.*, **53**, 165–202, <https://doi.org/10.1002/2014RG000473>.
- Bitz, C. M., and G. H. Roe, 2004: A mechanism for the high rate of sea ice thinning in the Arctic Ocean. *J. Climate*, **17**, 3623–3632, [https://doi.org/10.1175/1520-0442\(2004\)017<3623:AMFTHR>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<3623:AMFTHR>2.0.CO;2).
- Boé, J., A. Hall, and X. Qu, 2009: September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nat. Geosci.*, **2**, 341–343, <https://doi.org/10.1038/ngeo467>.
- Borodina, A., E. M. Fischer, and R. Knutti, 2017: Emergent constraints in climate projections: A case study of changes in high-latitude temperature variability. *J. Climate*, **30**, 3655–3670, <https://doi.org/10.1175/JCLI-D-16-0662.1>.
- 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, <https://doi.org/10.1007/s00382-012-1330-3>.
- , —, J. Turner, and T. Phillips, 2015: The importance of sea ice area biases in 21st century multimodel projections of Antarctic temperature and precipitation. *Geophys. Res. Lett.*, **42**, 10 832–10 839, <https://doi.org/10.1002/2015GL067055>.
- Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. Zwally, 1996: Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Arctic, full record. NASA National Snow and Ice Data Center Distributed Active Archive Center, accessed 11 December 2018, <https://doi.org/10.5067/8GQ8LZQVLOVL>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Dirksen, A., W. J. Merryfield, and A. Monahan, 2017: Impacts of sea ice thickness initialization on seasonal Arctic sea ice predictions. *J. Climate*, **30**, 1001–1017, <https://doi.org/10.1175/JCLI-D-16-0437.1>.
- Donner, L. J., and Coauthors, 2011: The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *J. Climate*, **24**, 3484–3519, <https://doi.org/10.1175/2011JCLI3955.1>.
- Dufresne, J. L., and Coauthors, 2013: Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dyn.*, **40**, 2123–2165, <https://doi.org/10.1007/s00382-012-1636-1>.
- Dunne, J. P., and Coauthors, 2013: GFDL's ESM2 global coupled climate-carbon Earth system models. Part II: Carbon system formulation and baseline simulation characteristics. *J. Climate*, **26**, 2247–2267, <https://doi.org/10.1175/JCLI-D-12-00150.1>.
- Eyring, V., and Coauthors, 2016: ESMValTool (v1.0)—A community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geosci. Model Dev. Discuss.*, **9**, 1747–1802, <https://doi.org/10.5194/gmd-9-1747-2016>.
- , and Coauthors, 2019: Taking climate model evaluation to the next level. *Nat. Climate Change*, **9**, 102–110, <https://doi.org/10.1038/s41558-018-0355-y>.
- Fowler, C., W. J. Emery, and J. Maslanik, 2004: Satellite-derived evolution of Arctic sea ice age: October 1978 to March 2003. *IEEE Geosci. Remote Sens. Lett.*, **1**, 71–74, <https://doi.org/10.1109/LGRS.2004.824741>.
- Fyfe, J. C., and Coauthors, 2013: One hundred years of Arctic surface temperature variation due to anthropogenic influence. *Sci. Rep.*, **3**, 2645, <https://doi.org/10.1038/srep02645>.
- Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973–4991, <https://doi.org/10.1175/2011JCLI4083.1>.
- Giorgetta, M. A., and Coauthors, 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.*, **5**, 572–597, <https://doi.org/10.1002/jame.20038>.
- Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1107, <https://doi.org/10.1175/2009BAMS2607.1>.
- , and —, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dyn.*, **37**, 407–418, <https://doi.org/10.1007/s00382-010-0810-6>.
- Hazeleger, W., and Coauthors, 2010: EC-Earth. *Bull. Amer. Meteor. Soc.*, **91**, 1357–1364, <https://doi.org/10.1175/2010BAMS2877.1>.
- Hodson, D. L. R., and Coauthors, 2012: Identifying uncertainties in Arctic climate change projections. *Climate Dyn.*, **40**, 2849–2865, <https://doi.org/10.1007/s00382-012-1512-z>.
- Holland, M. M., and C. M. Bitz, 2003: Polar amplification of climate change in coupled models. *Climate Dyn.*, **21**, 221–232, <https://doi.org/10.1007/s00382-003-0332-6>.
- Iversen, T., and Coauthors, 2013: The Norwegian Earth System Model, NorESM1-M—Part 2: Climate response and scenario projections. *Geosci. Model Dev.*, **6**, 389–415, <https://doi.org/10.5194/gmd-6-389-2013>.
- Jahn, A., 2018: Reduced probability of ice-free summers for 1.5°C compared to 2°C warming. *Nat. Climate Change*, **8**, 409–413, <https://doi.org/10.1038/s41558-018-0127-8>.
- , J. E. Kay, M. M. Holland, and D. M. Hall, 2016: How predictable is the timing of a summer ice-free Arctic? *Geophys. Res. Lett.*, **43**, 9113–9120, <https://doi.org/10.1002/2016GL070067>.
- Karpechko, A. Y., D. Maraun, and V. Eyring, 2013: Improving Antarctic total ozone projections by a process-oriented multiple diagnostic ensemble regression. *J. Atmos. Sci.*, **70**, 3959–3976, <https://doi.org/10.1175/JAS-D-13-071.1>.
- Kay, J. E., M. M. Holland, and A. Jahn, 2011: Inter-annual to multi-decadal Arctic sea ice extent trends in a warming world. *Geophys. Res. Lett.*, **38**, L15708, <https://doi.org/10.1029/2011GL048008>.
- , and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, <https://doi.org/10.1175/BAMS-D-13-00255.1>.

- Knutti, R., J. Sedláček, B. M. Sanderson, R. Lorenz, E. M. Fischer, and V. Eyring, 2017: A climate model projection weighting scheme accounting for performance and interdependence. *Geophys. Res. Lett.*, **44**, 1909–1918, <https://doi.org/10.1002/2016GL072012>.
- Kwok, R., 2018: Arctic sea ice thickness, volume, and multiyear ice coverage: Losses and coupled variability (1958–2018). *Environ. Res. Lett.*, **13**, 105005, <https://doi.org/10.1088/1748-9326/aae3ec>.
- , G. F. Cunningham, M. Wensnahan, I. Rigor, H. J. Zwally, and D. Yi, 2009: Thinning and volume loss of the Arctic Ocean sea ice cover: 2003–2008. *J. Geophys. Res.*, **114**, C07005, <https://doi.org/10.1029/2009JC005312>.
- Labe, Z., G. Magnusdottir, and H. Stern, 2018: Variability of Arctic sea ice thickness using PIOMAS and the CESM Large Ensemble. *J. Climate*, **31**, 3233–3247, <https://doi.org/10.1175/JCLI-D-17-0436.1>.
- Laxon, S., N. Peacock, and D. Smith, 2003: High interannual variability of sea ice thickness in the Arctic region. *Nature*, **425**, 947–950, <https://doi.org/10.1038/nature02050>.
- , and Coauthors, 2013: CryoSat-2 estimates of Arctic sea ice thickness and volume. *Geophys. Res. Lett.*, **40**, 732–737, <https://doi.org/10.1002/grl.50193>.
- Li, L., and Coauthors, 2013: The Flexible Global Ocean–Atmosphere–Land System Model, grid-point version 2: FGOALS-g2. *Adv. Atmos. Sci.*, **30**, 543–560, <https://doi.org/10.1007/s00376-012-2140-6>.
- Lindsay, R. W., and J. Zhang, 2006: Assimilation of ice concentration in an ice–ocean model. *J. Atmos. Oceanic Technol.*, **23**, 742–749, <https://doi.org/10.1175/JTECH1871.1>.
- Lique, C., M. M. Holland, Y. B. Dibike, D. M. Lawrence, and J. A. Screen, 2016: Modeling the Arctic freshwater system and its integration in the global system: Lessons learned and future challenges. *J. Geophys. Res. Biogeosci.*, **121**, 540–566, <https://doi.org/10.1002/2015JG003120>.
- Liu, J., M. Song, R. M. Horton, and Y. Hu, 2013: Reducing spread in climate model projections of a September ice-free Arctic. *Proc. Natl. Acad. Sci. USA*, **110**, 12 571–12 576, <https://doi.org/10.1073/pnas.1219716110>.
- Mahlstein, I., and R. Knutti, 2012: September Arctic sea ice predicted to disappear near 2°C global warming above present. *J. Geophys. Res. Atmos.*, **117**, D06104, <https://doi.org/10.1029/2011JD016709>.
- Martin, G. M., and Coauthors, 2011: The HadGEM2 family of Met Office Unified Model climate configurations. *Geosci. Model Dev.*, **4**, 723–757, <https://doi.org/10.5194/gmd-4-723-2011>.
- Maslanik, J., J. Stroeve, C. Fowler, and W. Emery, 2011: Distribution and trends in Arctic sea ice age through spring 2011. *Geophys. Res. Lett.*, **38**, L13502, <https://doi.org/10.1029/2011GL047735>.
- Massonnet, F., and Coauthors, 2012: Constraining projections of summer Arctic sea ice. *Cryosphere*, **6**, 1383–1394, <https://doi.org/10.5194/tc-6-1383-2012>.
- , and Coauthors, 2018: Arctic sea-ice change tied to its mean state through thermodynamic processes. *Nat. Climate Change*, **8**, 599–603, <https://doi.org/10.1038/s41558-018-0204-z>.
- Melia, N., K. Haines, and E. Hawkins, 2015: Improved Arctic sea ice thickness projections using bias-corrected CMIP5 simulations. *Cryosphere*, **9**, 2237–2251, <https://doi.org/10.5194/tc-9-2237-2015>.
- Petty, A. A., and Coauthors, 2018: The Arctic sea ice cover of 2016: A year of record-low highs and higher-than-expected lows. *Cryosphere*, **12**, 433–452, <https://doi.org/10.5194/tc-12-433-2018>.
- Rayner, N. A., and Coauthors, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, <https://doi.org/10.1029/2002JD002670>.
- Riahi, K., and Coauthors, 2011: RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, **109**, 33–57, <https://doi.org/10.1007/s10584-011-0149-y>.
- Rotstayn, L. D., and Coauthors, 2012: Aerosol- and greenhouse gas-induced changes in summer rainfall and circulation in the Australasian region: A study using single-forcing climate simulations. *Atmos. Chem. Phys.*, **12**, 6377–6404, <https://doi.org/10.5194/acp-12-6377-2012>.
- Sanderson, B. M., M. Wehner, and R. Knutti, 2017: Skill and independence weighting for multi-model assessments. *Geosci. Model Dev.*, **10**, 2379–2395, <https://doi.org/10.5194/gmd-10-2379-2017>.
- Schmidt, G. A., and Coauthors, 2014: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J. Adv. Model. Earth Syst.*, **6**, 141–184, <https://doi.org/10.1002/2013MS000265>.
- Schweiger, A., and Coauthors, 2011: Uncertainty in modeled Arctic sea ice volume. *J. Geophys. Res.*, **116**, C00D06, <https://doi.org/10.1029/2011JC007084>.
- Screen, J. A., 2018: Arctic sea ice at 1.5 and 2°C. *Nat. Climate Change*, **8**, 362–363, <https://doi.org/10.1038/s41558-018-0137-6>.
- , and D. Williamson, 2017: Ice-free Arctic at 1.5°C? *Nat. Climate Change*, **7**, 230–231, <https://doi.org/10.1038/nclimate3248>.
- , I. Simmonds, C. Deser, and R. Tomas, 2013: The atmospheric response to three decades of observed Arctic sea ice loss. *J. Climate*, **26**, 1230–1248, <https://doi.org/10.1175/JCLI-D-12-00063.1>.
- Sigmond, M., J. C. Fyfe, and N. C. Swart, 2018: Ice-free Arctic projections under the Paris Agreement. *Nat. Climate Change*, **8**, 404–408, <https://doi.org/10.1038/s41558-018-0124-y>.
- Snape, T. J., and P. M. Forster, 2014: Decline of Arctic sea ice: Evaluation and weighting of CMIP5 projections. *J. Geophys. Res.*, **119**, 546–554, <https://doi.org/10.1002/2013JD020593>.
- Sprenn, G., R. Kwok, and D. Menemenlis, 2011: Trends in Arctic sea ice drift and role of wind forcing: 1992–2009. *Geophys. Res. Lett.*, **38**, L19501, <https://doi.org/10.1029/2011GL048970>.
- Stroeve, J. C., M. C. Serreze, M. M. Holland, J. E. Kay, J. Malanik, and A. P. Barrett, 2012a: The Arctic’s rapidly shrinking sea ice cover: A research synthesis. *Climatic Change*, **110**, 1005–1027, <https://doi.org/10.1007/s10584-011-0101-1>.
- , V. Kattsov, A. Barrett, M. Serreze, T. Pavlova, M. Holland, and W. N. Meier, 2012b: Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations. *Geophys. Res. Lett.*, **39**, L16502, <https://doi.org/10.1029/2012GL052676>.
- , A. Barrett, M. Serreze, and A. Schweiger, 2014: Using records from submarine, aircraft and satellites to evaluate climate model simulations of Arctic sea ice thickness. *Cryosphere*, **8**, 1839–1854, <https://doi.org/10.5194/tc-8-1839-2014>.
- Swart, N. C., and J. C. Fyfe, 2013: The influence of recent Antarctic ice sheet retreat on simulated sea ice area trends. *Geophys. Res. Lett.*, **40**, 4328–4332, <https://doi.org/10.1002/grl.50820>.
- , —, E. Hawkins, J. E. Kay, and A. Jahn, 2015: Influence of internal variability on Arctic sea-ice trends. *Nat. Climate Change*, **5**, 86–89, <https://doi.org/10.1038/nclimate2483>.
- 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, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Thorndike, A. S., and R. Colony, 1982: Sea ice motion in response to geostrophic winds. *J. Geophys. Res.*, **87**, 5845–5852, <https://doi.org/10.1029/JC087iC08p05845>.

- Tilling, R. L., A. Ridout, A. Shepherd, and D. J. Wingham, 2015: Increased Arctic sea ice volume after anomalously low melting in 2013. *Nat. Geosci.*, **8**, 643–646, <https://doi.org/10.1038/ngeo2489>.
- van Vuuren, D. P., and Coauthors, 2011: The representative concentration pathways: An overview. *Climatic Change*, **109**, 5–31, <https://doi.org/10.1007/s10584-011-0148-z>.
- Vichi, M., and Coauthors, 2011: Global and regional ocean carbon uptake and climate change: Sensitivity to a substantial mitigation scenario. *Climate Dyn.*, **37**, 1929–1947, <https://doi.org/10.1007/s00382-011-1079-0>.
- Voldoire, A., and Coauthors, 2012: The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dyn.*, **40**, 2091–2121, <https://doi.org/10.1007/s00382-011-1259-y>.
- Volodin, E. M., N. A. Dianskii, and A. V. Gusev, 2010: Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izv. Atmos. Ocean. Phys.*, **46**, 414–431, <https://doi.org/10.1134/S000143381004002X>.
- von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 494 pp.
- Walsh, J. E., W. L. Chapman, and F. Fetterer, 2015: Gridded Monthly Sea Ice Extent and Concentration, 1850 Onward, version 1. National Snow and Ice Data Center, accessed 11 December 2018, <https://doi.org/10.7265/N5833PZ5>.
- Wang, M., and J. E. Overland, 2009: A sea ice free summer Arctic within 30 years? *Geophys. Res. Lett.*, **36**, L07502, <https://doi.org/10.1029/2009GL037820>.
- Watanabe, S., and Coauthors, 2011: MIROC-ESM: Model description and basic results of CMIP5-20c3m experiments. *Geosci. Model Dev.*, **4**, 845–872, <https://doi.org/10.5194/gmd-4-845-2011>.
- Weeks, W. F., 2010: *On Sea Ice*. University of Alaska Press, 664 pp.
- Wenzel, S., V. Eyring, E. P. Gerber, and A. Y. Karpechko, 2016: Constraining future summer austral jet stream positions in the CMIP5 ensemble by process-oriented multiple diagnostic regression. *J. Climate*, **29**, 673–687, <https://doi.org/10.1175/JCLI-D-15-0412.1>.
- Wu, T., and Coauthors, 2014: An overview of BCC climate system model development and application for climate change studies. *Acta Meteor. Sin.*, **28**, 34–56, <https://doi.org/10.1007/s13351-014-3041-7>.
- Yukimoto, S., and Coauthors, 2012: A new global climate model of the Meteorological Research Institute: MRI-CGCM3—Model description and basic performance. *J. Meteor. Soc. Japan*, **90A**, 23–64, <https://doi.org/10.2151/jmsj.2012-A02>.
- Zhang, J. L., and D. A. Rothrock, 2003: Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Mon. Wea. Rev.*, **131**, 845–861, [https://doi.org/10.1175/1520-0493\(2003\)131<0845:MGSIWA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0845:MGSIWA>2.0.CO;2).
- , —, and M. Steele, 2000: Recent changes in Arctic sea ice: The interplay between ice dynamics and thermodynamics. *J. Climate*, **13**, 3099–3114, [https://doi.org/10.1175/1520-0442\(2000\)013<3099:RCIASI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<3099:RCIASI>2.0.CO;2).