

Attribution of Extreme Events in Arctic Sea Ice Extent

MEGAN C. KIRCHMEIER-YOUNG

Pacific Climate Impacts Consortium, and Canadian Centre for Climate Modelling and Analysis, University of Victoria, Victoria, British Columbia, Canada

FRANCIS W. ZWIERS

Pacific Climate Impacts Consortium, University of Victoria, Victoria, British Columbia, Canada

NATHAN P. GILLET

Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, British Columbia, Canada

(Manuscript received 2 June 2016, in final form 28 September 2016)

ABSTRACT

Arctic sea ice extent (SIE) has decreased over recent decades, with record-setting minimum events in 2007 and again in 2012. A question of interest across many disciplines concerns the extent to which such extreme events can be attributed to anthropogenic influences. First, a detection and attribution analysis is performed for trends in SIE anomalies over the observed period. The main objective of this study is an event attribution analysis for extreme minimum events in Arctic SIE. Although focus is placed on the 2012 event, the results are generalized to extreme events of other magnitudes, including both past and potential future extremes. Several ensembles of model responses are used, including two single-model large ensembles. Using several different metrics to define the events in question, it is shown that an extreme SIE minimum of the magnitude seen in 2012 is consistent with a scenario including anthropogenic influence and is extremely unlikely in a scenario excluding anthropogenic influence. Hence, the 2012 Arctic sea ice minimum provides a counterexample to the often-quoted idea that individual extreme events cannot be attributed to human influence.

1. Introduction

Sea ice extent (SIE) in the Arctic has decreased throughout the satellite record (Vaughan et al. 2013a). Loss of Arctic sea ice has implications in many areas (IPCC 2014; Serreze et al. 2007), such as ecosystems, transportation, fisheries/commerce, and Arctic communities. Arctic SIE reached a minimum of $4.28 \times 10^6 \text{ km}^2$ in September 2007 (Stroeve et al. 2008), during a period

of strong Arctic sea ice decline (Stroeve et al. 2007; Comiso et al. 2008; Serreze et al. 2007). Although the rate of sea ice loss decreased subsequently, both short-period trends were within the range of natural variability plus an all-forcing (natural and anthropogenic) signal (Swart et al. 2015). A new minimum was reached in 2012, with the Arctic experiencing a September average SIE of only $3.62 \times 10^6 \text{ km}^2$ (Fetterer et al. 2002, a continuously updated sea ice index). To what extent can these extreme minimum events in Arctic SIE be attributed to anthropogenic influence on the climate?

Climate change detection and attribution methodologies have been widely used to detect changes in climate variables and attribute these to anthropogenic influences (Stott et al. 2010; Hegerl and Zwiers 2011; Bindoff et al. 2013). According to Hegerl et al. (2010), climate change detection aims to determine if observed changes are inconsistent with internal variability, while attribution goes a step farther in showing an observed

Denotes Open Access content.

Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-16-0412.s1>.

Corresponding author e-mail: Megan Kirchmeier-Young, mkirch@uvic.ca

DOI: 10.1175/JCLI-D-16-0412.1

© 2017 American Meteorological Society

change is only consistent with a certain forcing scenario (e.g., natural plus anthropogenic).

Recent focus has been on the attribution of extreme events, especially those of particular societal interest (Peterson et al. 2013; Herring et al. 2015; Christidis et al. 2015). Event attribution aims to determine if a specific, usually extreme, event can be attributed to anthropogenic influences. A common metric of event attribution is the fraction of attributable risk (FAR; Stott et al. 2004; Stone and Allen 2005), which quantifies the additional risk of an event's occurrence due to anthropogenic forcings from that due to natural forcings alone by comparing the probabilities of that event under each forcing scenario. Reviews of event attribution are found in Stott et al. (2016) and Hulme (2014) and in a new report from the National Academies of Science, Engineering, and Medicine (NASEM 2016) that summarizes the methodology and current state of the science of event attribution.

Arctic temperatures have been increasing (Hartmann et al. 2013) for the last several decades and will likely continue to increase at a greater rate than at lower latitudes because of Arctic amplification (Serreze and Barry 2011). Gillett et al. (2008) attributed recent trends in Arctic temperatures to anthropogenic influences. Najafi et al. (2015) distinguished between the separate effects of greenhouse gases (GHGs) and other anthropogenic forcings (OANT; including aerosols) on Arctic temperature. They detected a signal from both GHG and OANT in the observed trend and demonstrated that OANT can act against GHG-induced warming, offsetting this warming by about 60%. Understanding the attribution of Arctic temperature changes allows for a more robust attribution of Arctic sea ice changes, as SIE is strongly influenced by the temperature of the region.

In regards to Arctic sea ice, Vinnikov et al. (1999) demonstrated that observed and modeled trends in Arctic sea ice were unlikely to be the result of internal variability alone. Gregory et al. (2002) found modeled SIE trends (from a single model), when including anthropogenic forcing, were consistent with observations, while simulations with only natural forcing exhibited no significant trend. In an observation-based analysis, Notz and Marotzke (2012) showed that both trends and extreme minima in Arctic SIE are inconsistent with internal variability alone and that the observed trend shows a strong correlation to the trend in atmospheric CO₂, a main component of anthropogenic forcings. Through a detection and attribution analysis, Min et al. (2008) found an anthropogenic influence on the time series of annual Arctic sea ice anomalies beginning in the 1990s; for individual months, both anthropogenic and all-forcing signals were detected from May through December.

It has been shown that internal variability can strongly influence decreasing trends in SIE, especially on shorter time scales (Kay et al. 2011) and mostly in the form of the Atlantic multidecadal oscillation (Day et al. 2012). Specifically, for the 2012 record minimum of Arctic SIE, it has been shown that this event is inconsistent with internal variability alone (Zhang and Knutson 2013) and was likely influenced by warmer surface air temperatures and sea ice memory, which includes ice thinning (Guemas et al. 2013). To our knowledge, an event attribution analysis of the 2012 minimum and other such extreme events in Arctic SIE has not yet been done. This would be useful because it would provide a quantitative assessment of whether anthropogenic forcing was necessary for the event to occur and whether it is sufficient for such events to continue to occur repeatedly in the future. Characterizing the causes and implications of climate change through specific events provides policymakers with another way in which to understand the implications of the continuing greenhouse gas increases. In addition, an event attribution analysis could be of interest to those studying impacts of the 2012 minimum and other such extremes.

This study focuses on attribution of extreme events in Arctic SIE after first performing a detection and attribution analysis for the time series. In particular, large ensembles from the Second Generation Canadian Earth System Model (CanESM2; Arora et al. 2011) and the Community Earth System Model, version 1, (CESM1; Kay et al. 2015) are used. An advantage of a large ensemble is the ability to generate many realizations that allow for a robust sampling of internal climate variability as simulated by that model under transient conditions. As event attribution results can be very sensitive to the framing of the question (NASEM 2016; Stott et al. 2016; Shepherd 2016; Otto et al. 2012), several metrics are used to define the extreme minimum in 2012. This paper is organized as follows. Section 2 includes a discussion of the observations and model ensembles used for the analysis. In section 3, a description of the attribution methodology is presented, including analyses of temporal patterns and specific events. The results are presented in section 4, followed by a discussion and conclusions in section 5. The main goal of this study is to quantify the extent to which extreme events in Arctic SIE (such as that in 2012) can be attributed to anthropogenic influences. Furthermore, the analysis will set up an attribution assessment of future extreme minima.

2. Data

a. Observations

Observations of Arctic SIE were acquired from the National Snow and Ice Data Center (NSIDC). NSIDC

calculates a sea ice index (Fetterer et al. 2002) that represents the SIE north of 30.98°N. Data are available beginning in November 1978. Two satellite datasets are used to create the product based on observations from several passive microwave sensors. A brief overview of the use of satellites in monitoring sea ice can be found in Vaughan et al. (2013b).

In calculating the sea ice index, Fetterer et al. (2002) first convert satellite brightness temperatures to sea ice concentrations (SICs) for each grid box. Then mean monthly SICs are calculated. Grid boxes are designated as “ice” or “no ice” using an SIC threshold of 15%, which is commonly used for the calculation of SIE (Vaughan et al. 2013a). The sum of areas of the ice grid boxes in the Northern Hemisphere is calculated for each month. The resulting SIEs are presented in millions of square kilometers.

b. Models

1) CANESM2 LARGE ENSEMBLE

CanESM2 (Arora et al. 2011) is a coupled Earth system model developed and run by the Canadian Centre for Climate Modelling and Analysis (CCCma) and used in phase 5 of the Coupled Model Intercomparison Project (CMIP5). The CMIP5 submission included five ensemble members run with historical forcings from 1850 to 2005. A large-ensemble project expanded this initial ensemble by branching each of the original five simulations into 10 ensemble members each in 1950, resulting in a total of 50 ensemble members. Different realizations were generated by changing the seed of a random number generator in the cloud parameterization.

Simulations were run with both natural forcings (NAT), which include solar and volcanic influences only, and combined natural and anthropogenic forcings (ALL). The anthropogenic forcings include human influences, such as greenhouse gases, aerosol emissions, and land use. The CanESM2 large ensemble has data through 2100 for ALL simulations, with years beyond 2005 utilizing representative concentration pathway 8.5 (RCP8.5) simulations (van Vuuren et al. 2011), while NAT simulations were run through 2020. An assessment of the standard deviation across the 10 realizations branched from the same simulation (Fig. S1 in the supplementary material) indicates that eliminating the first 10 years of data is more than sufficient for SIE in the ensemble members to diverge. Note that other climate elements may take longer or shorter times to diverge (e.g., ocean heat content vs atmospheric water vapor content) and that SIE does not require the divergence of the full ocean variability. SIE is calculated as the sum of the areas of all grid boxes north of 30°N with an SIC exceeding 15% during that month.

Figure 1a shows time series of SIE for the ensemble means of ALL (blue) and NAT (green) responses, compared to the observations in black. For each forcing scenario, the shading represents the area between the 5th and 95th percentile values across the ensemble members. It is apparent that the CanESM2 ALL data are biased low relative to the observations. This bias is stronger during the boreal summer and autumn months (Fig. 2) and is associated with model sea ice that is too thin and thus melts away too quickly with seasonally warming temperatures (Shu et al. 2015; N. Swart 2015, personal communication).

For the following analysis to be more relevant to recent values of Arctic sea ice extent, the CanESM2 output was adjusted such that the ALL simulations better matched the observations during the common period. Without bias correction, it would be difficult to assess an event of the magnitude of the observed record. Since anomalies relative to the 1981–2010 period compared well between the ALL ensemble mean and observations (Fig. 3b), a simple offset was used for the adjustment. The difference (by month) between the 1981–2010 mean of the observations and the mean of the ALL ensemble mean was added to each ensemble member from both ALL and NAT simulations. The resulting ensemble means are shown in darker colors in Fig. 1a. This method of bias correction, while changing the magnitude of the SIE values, maintains the ensemble spread and the differences between the ALL and NAT results. As an alternative method, the SIC threshold was decreased, but the resulting SIE values are still biased low relative to the observations, and thus the mean-adjustment method was chosen as a simple approach for making bias adjustments. A caveat, however, is that this approach does not account for the relationship between variability and the mean state (Fig. 4). The dependence of variability on the mean state will be discussed in more detail in a subsequent section.

2) CESM1 LARGE ENSEMBLE

A large ensemble of the CESM1 simulations is also available, branching off their CMIP5 ensemble simulations in 1920 (Kay et al. 2015). The output is a 30-member ensemble for ALL forcings with useable data covering 1930–2080. Forcings from the RCP8.5 simulations are used for years beyond 2005. Figure 1b shows the CESM1 ensemble spread compared to observations. Although the magnitude of September SIE from CESM1 agrees well with observations during the earlier period, the declining trend in SIE is weaker. Nevertheless, the difference between the trends in the model's ensemble mean and in the observations is consistent with internal variability as the observed trend falls within the 5th to

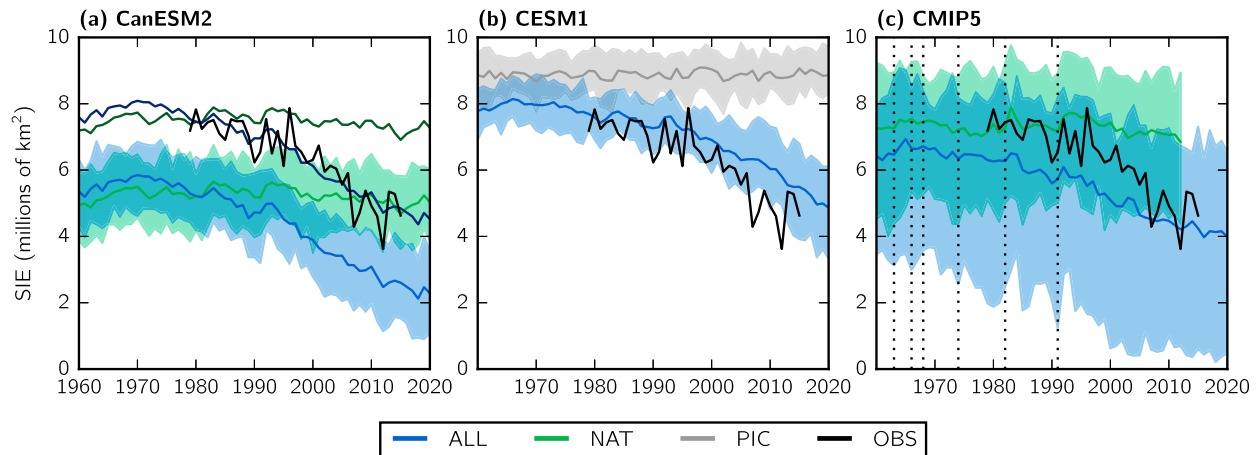


FIG. 1. Time series of September SIE from the large ensembles of (a) CanESM2 and (b) CESM1 and (c) for the CMIP5 ensemble. The solid lines indicate the ensemble mean, and the shaded regions represent the 5th–95th percentile range across the ensemble. Observations are shown by the thick black lines. For CanESM2, the mean-adjusted ensemble means are also plotted. In (c), the vertical dashed lines indicate the timing of major volcanic eruptions from an updated dataset of Sato et al. (1993).

95th percentile range of trends across the individual CESM1 ensemble members. Note that the phases of internal multidecadal variability in observed (Zhang 2015) and model simulated realizations of Arctic sea ice are likely different, such that the effect of natural internal multidecadal variability would be averaged out in the ensemble mean, whereas it is present in the observations. The seasonal cycle from CESM1 (Fig. 2) shows fairly good agreement with the observed seasonal cycle, although the model SIE decreases less through the summer months than observed. This assessment is similar to that presented by Jahn et al. (2012) for a related version of this model (CCSM4).

The CESM1 large ensemble does not include runs with only NAT forcings, so the preindustrial control (PIC) simulation will be used for comparison instead. If the 1000-yr PIC run is divided into segments of a length similar to the observations, the result is about six fewer ensemble members than the ALL simulation. As an extended period is plotted here, there are about half as many ensemble members for the PIC, although the event attribution analysis will use a PIC ensemble of the same size as the ALL ensemble for the slightly shorter period required. However, the spread of the PIC ensemble changes very little with increased ensemble size.

3) IPSL ENSEMBLE

As the CanESM2 shows a low bias relative to observations and the CESM1 agrees fairly well, a third model showing a high bias was chosen from the CMIP5 archives. The IPSL-CM5 model (Dufresne et al. 2013) from the Institute Pierre-Simon Laplace (IPSL) was chosen from those showing a high bias of September SIE

relative to the observations during the overlapping period because of the size of its ensemble and the availability of simulations with NAT forcings. The full model name is IPSL-CM5A-LR, but this ensemble will simply be referred to as IPSL hereinafter. IPSL provides only three ensemble members. However, the accompanying preindustrial control run is long and should sufficiently sample the internal variability. IPSL data beyond 2005 are forced with the RCP8.5 simulations.

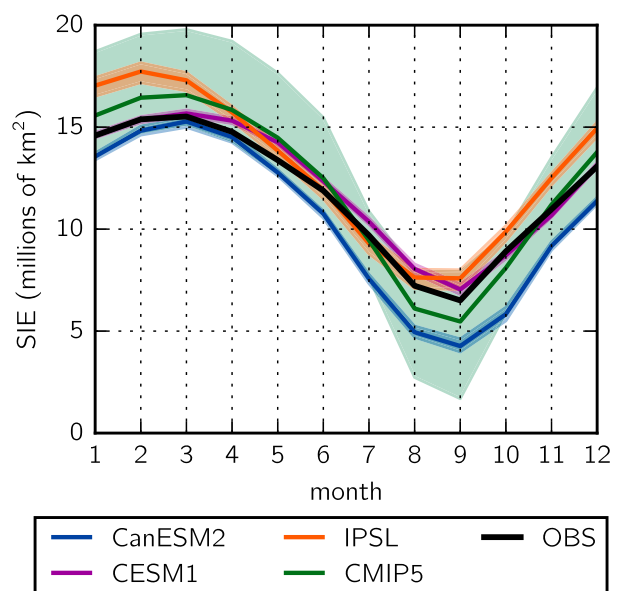


FIG. 2. Seasonal cycle of SIE for each model averaged over 1981–2010. Solid lines represent the ensemble mean from the ALL forcings scenario, and the shaded region indicates the 5th–95th percentile range.

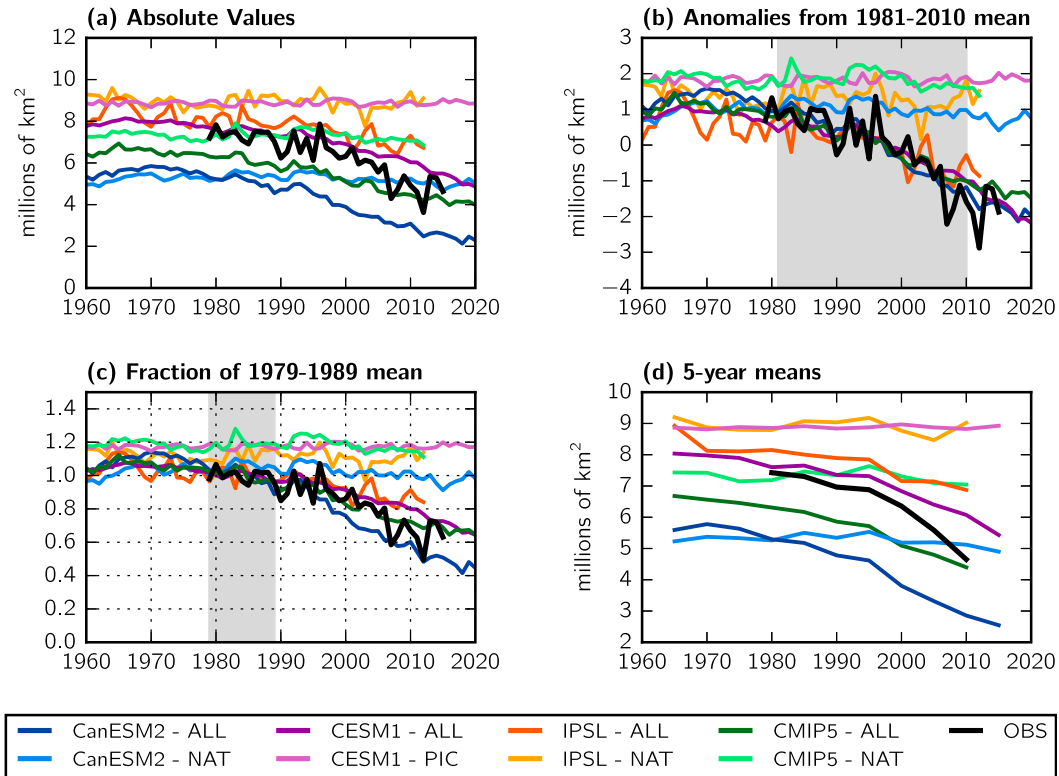


FIG. 3. Time series of four methods to characterize SIE events: (a) absolute values, (b) anomalies relative to the 1981–2010 mean, (c) the fraction of the 1979–89 mean, and (d) nonoverlapping 5-yr averages. The observations are shown in black, and the ensemble mean from each of the four ensembles and each of the two forcing scenarios are shown by colors. The shaded regions in (b) and (c) depict the reference period. NAT or PIC anomalies and fractions are calculated relative to the mean from the corresponding ALL forcing simulations.

Because of its small ensemble, IPSL is excluded from Fig. 1, but a comparison of its ensemble means with observations can be found in Fig. 3a. Similar to that of CESM1, the SIE trend from IPSL is weaker than that in observations, although, again, this could be influenced by internal variability. Using a similar method to that for CanESM2, the IPSL data were bias corrected using a simple shift so that the mean for the 1981–2010 period for the ALL ensemble mean matches that from observations. This bias adjustment results in shifting the IPSL data down by approximately $1 \times 10^6 \text{ km}^2$ in September.

4) CMIP5 ENSEMBLE

Finally, an ensemble from the collection of models included in CMIP5 is also utilized. Historical simulations are available through 2005, and RCP4.5 simulations are used to extend the results through 2012. Choosing RCP4.5 over RCP8.5 allows for a larger ensemble. As the NAT data limit the analysis through 2012, the choice of RCP would make little difference; the radiative forcings for RCP4.5 and RCP8.5 begin diverging from each other around 2020 (van Vuuren

et al. 2011). All models that submitted historical, RCP4.5, and natural-forcing simulations for the desired time period(s) are included. All models with NAT simulations ending in 2005 were excluded. The resulting ensemble has 40 members in total from nine different models (see Table S1 in the supplementary material). However, for most of the subsequent analyses, a 35-member ensemble is used, after removing the CSIRO model. The Arctic sea ice from the CSIRO model is a clear outlier and differs greatly from the observations (Massonnet et al. 2012); the CSIRO model also greatly underestimates SIE variability (Fig. S2 in the supplementary material). The effects of including this model will be discussed in section 4. An overview of the CMIP5 models' performance for sea ice can be found in Shu et al. (2015). Only one of the models selected (IPSL-CM5A-MR) is included in the IPCC's selection of the best five models for sea ice based on a multifaceted comparison to observations (Collins et al. 2013); unfortunately, most of these models did not provide the NAT simulations required for subsequent analyses. It should be noted that the CMIP5 ensemble includes

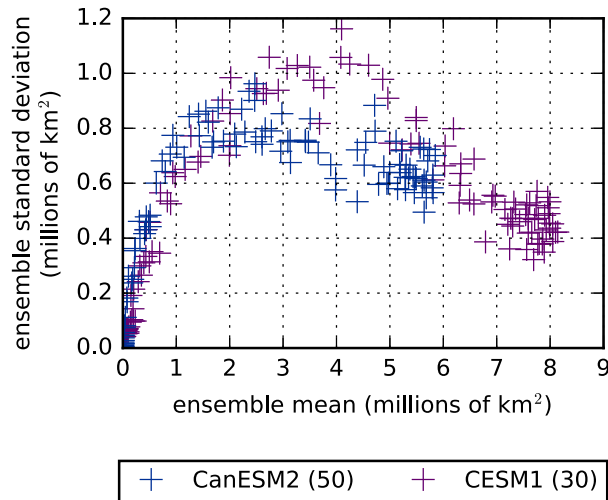


FIG. 4. Mean vs standard deviation calculated across the ensemble for each September in 1960–2080 for CanESM2 and CESM1. Numbers in parentheses in the legend indicate the number of ensemble members.

contributions from two of the three models discussed above: CanESM2 and IPSL.

An illustration of the CMIP5 ensemble spread and a comparison to observations are shown in Fig. 1c. The CMIP5 ensemble shows considerably more spread than either of the single-model ensembles (Figs. 1a,b). Some of the models included in this ensemble exhibit strong high or low biases relative to the observed values. The spread of biases is also evident in Fig. 2, where the CMIP5 multimodel ensemble shows a much wider range of seasonal patterns across the ensemble than the single-model ensembles. There is also large variability across ensemble members in the magnitude of trends in sea ice extent (Fig. 1c). The ensemble mean for CMIP5 is calculated with weights such that each model lends equal weight to the mean value. For absolute values, the CMIP5 ensemble mean shows a low bias, although the anomalies agree fairly well with observations (Fig. 3). The models included in CMIP5 present improved representations of observed SIE values and trends over the earlier model set of CMIP3 (Flato et al. 2013).

5) COMPARISON OF ENSEMBLES

Figure 3a displays the ensemble mean from the two scenarios for each model and the observations. The low bias from CanESM2 is strong, whereas the high bias of IPSL is weaker. The CESM1 ensemble mean compares fairly well to the observations in terms of SIE magnitude, while the CMIP5 ensemble mean is biased low. The CanESM2, CMIP5, and CESM1 time series in Fig. 3 show greatly reduced variance compared to the observations, as they represent ensemble averages over 50, 35,

and 30 realizations, respectively. Figure S2 compares the standard deviation of the observed time series to the standard deviation from each ensemble member from all models. Most models show slightly greater variability of SIE across the time series compared to the observations, although there is good agreement among the ALL, NAT, and PIC forcing scenarios within each model. CESM1 and a couple of CMIP5 models underestimate the variability compared to observations.

For each model, the NAT and ALL responses show similar SIE values during the earlier period, but the ALL results begin to deviate from the NAT responses around 1990 (Fig. 3a). The separation of the ALL and NAT responses can be seen more distinctly by viewing time series of anomalies. Figure 3b shows the ensemble means from each set as anomalies relative to the ALL ensemble mean from the 1981–2010 period. The observations and ALL responses show decreasing trends, with the Arctic losing roughly $2 \times 10^6 \text{ km}^2$ of sea ice extent over a 30-yr period, although this value differs by ensemble and ensemble member (not shown). In contrast, the NAT (and CESM1 PIC) responses show no obvious trends in SIE but do exhibit some evidence of low-frequency variability and a response to individual volcanic forcing events (see Fig. 1c).

3. Methodology

a. Detection and attribution of the temporal evolution

Although the goal of this study is an event attribution analysis, a preliminary detection and attribution analysis is performed for the SIE time series as the detection and attribution of long-term change in SIE or a closely related variable, such as surface air temperature (Najafi et al. 2015), strengthens event attribution findings (NASEM 2016) by demonstrating that external forcing has discernibly changed the background state against which extreme events occur. Separate detection and attribution analyses were performed for the time series of the annual SIE values and for the time series of September SIE values. Most detection and attribution studies follow the total least squares (TLS) methodology presented in Allen and Stott (2003). The TLS problem is set up as follows:

$$\begin{aligned} y &= y^* + \varepsilon_0, \\ x_i &= x_i^* + \varepsilon_i, \quad \text{and} \\ y^* &= \sum_{i=1}^m \beta_i x_i^*, \end{aligned} \quad (1)$$

where y represents the observations, y^* the true climate response to all acting forcings, each x_i the modeled response to one of m forcings, and x_i^* the noise-free response to that forcing that is anticipated by climate

models. The noise on y , denoted by ε_0 , is assumed to represent internal climate variability, while the noise on x_i , denoted by ε_i , is a result of both internal variability and the finite ensemble used to estimate the model response. It is the inclusion of ε_i that is the difference between TLS and a simpler, ordinary least squares approach.

The TLS coefficients β_i are also referred to as scaling factors and are determined via a method described in [Van Huffel and Vandewalle \(1991\)](#):

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X} - \sigma_{m+1}^2 \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{Y}), \quad (2)$$

where σ_{m+1}^2 is the square of the last singular value of the matrix $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}]$. Here,

$$\begin{aligned} \mathbf{X} &= \mathbf{C}_r^{-1/2} (\sqrt{n_{\text{ens}}} \mathbf{X}_0) \quad \text{and} \\ \mathbf{Y} &= \mathbf{C}_r^{-1/2} \mathbf{Y}_0, \end{aligned} \quad (3)$$

where prewhitening of the ensemble mean responses \mathbf{X}_0 and the observed values \mathbf{Y}_0 is achieved by multiplying by the inverse square root of the regularized covariance matrix \mathbf{C}_r of a climate noise matrix. Prewhitening rotates and scales the observations and mean responses so that noise in the transformed vectors consists of uncorrelated components with unit variance (thus the analogy to white noise). Prewhitening simplifies the fitting of the TLS regression model ([Allen and Stott 2003](#)) and optimizes the detection problem by maximizing the signal-to-noise ratio in the ordinary least squares sense (e.g., [Mitchell et al. 2001](#)). The use of \mathbf{C}_r is based on the methodology of [Ribes et al. \(2013\)](#) and will be discussed in more detail subsequently. The first m columns of \mathbf{Z} are the prewhitened mean model response patterns, and the final column ($m + 1$) holds the prewhitened observations. Furthermore, as an ensemble mean possesses a damped variance compared to a single realization (and the observations), the variance of the ensemble mean response is first rescaled using the square root of the ensemble size n_{ens} ([Allen and Stott 2003](#); [Ribes et al. 2013](#)). As an example, for September, \mathbf{Y}_0 is the time series of observations (black line in [Fig. 1](#)), and \mathbf{X}_0 includes the ensemble mean time series responses from ALL and/or NAT forcings (see [Fig. 1](#)). Both \mathbf{Y}_0 and the columns of \mathbf{X}_0 are expressed as anomalies relative to the period mean.

One-signal analysis involves comparing the model response to a single forcing ($m = 1$) to the observations. A scaling factor different from 0 indicates the signal is “detected,” while a scaling factor consistent with unity implies the simulated and observed patterns are consistent in magnitude, which provides evidence that may contribute to attribution ([Mitchell et al. 2001](#)). Under the assumption of a linear combination of responses to

forcings ([Gillett et al. 2004](#); [Hegerl and Zwiers 2011](#); [Tett et al. 1999](#)), a two-signal analysis ($m = 2$) can be used. Here, scaling factors are fit jointly for x_{ALL} and x_{NAT} that can then be used to derive the scaling factors for the combination of ANT and NAT forcings using

$$\hat{\beta}_{\text{ALL}} = \hat{\beta}_{\text{ANT}} + \hat{\beta}_{\text{NAT}}. \quad (4)$$

A combination of the scaling factors is used to assess the ANT influence instead of combining the responses directly, as the TLS methodology assumes independence between the errors ε_i ([Ribes et al. 2013](#)).

It is common to use the prewhitening operator of [Allen and Stott \(2003\)](#) and [Allen and Tett \(1999\)](#), which often involves a dimension reduction based on the EOFs of internal variability from control simulation(s). In contrast, [Ribes et al. \(2013\)](#) present a method that produces an improved, full-rank estimate of the covariance matrix of internal variability, which makes it possible to derive the prewhitening operator without resorting to EOF truncation. This method, called regularized optimal fingerprinting (ROF), will be used here. The ROF method also includes slightly different methods for performing the residual consistency test (RCT; [Allen and Tett 1999](#)) and calculating the confidence intervals for the scaling factors ([Ribes et al. 2013](#)). The RCT is used to check that the estimate of ε_0 from the TLS analysis is consistent with the simulated internal variability.

An estimate of the covariance matrix representing internal variability is required for this analysis, to be used in calculation of the prewhitening operator and in the RCT, with separate estimates used for the two purposes. These covariance matrices are typically constructed by dividing the model’s control run into segments of the same length as the observations and then dividing the collection of segments into two samples. However, it has been shown that the internal variability of SIE can be dependent on the mean state ([Swart et al. 2015](#)). [Figure 4](#) plots the standard deviation across the ensemble for each year’s September value against the ensemble mean for that month [cf. [Fig. S6](#) in [Swart et al. \(2015\)](#)]. For both CanESM2 and CESM1, September SIE values at the high and low ends of the SIE spectrum show decreased standard deviation compared to the midrange SIE values. This pattern is season dependent. Using a preindustrial control run that sees only larger values of SIE may therefore underestimate the internal variability. An alternate method is to subtract the ensemble mean from each ALL member and calculate the internal variability covariance matrix from this set of realizations. (Note that one member must be removed here to maintain an independent set.) IPSL was not included in [Fig. 4](#) because of its small ensemble

(three members). Instead, the IPSL analysis will utilize its 1000-yr control run to calculate the covariance matrix. The CMIP5 ensemble was also excluded from Fig. 4 because of its large ensemble spread. The use of multiple models may also mask the relationship investigated here. It was found that, when applicable, using the difference from the ensemble mean instead of control simulations resulted in only small differences in the detection/attribution results that did not alter the conclusions.

These scaling factors can be used to rescale the model response(s) to better resemble the observed values (Sun et al. 2014; Christidis et al. 2012). One set of subsequent event attribution analyses will be performed using model responses that have been adjusted using the September scaling factors:

$$\begin{aligned} x_{\text{ALL},\beta,j} &= x_{\text{ALL},j} - \bar{x}_{\text{ALL}} + \hat{\beta}_{1,\text{ALL}}(\bar{x}_{\text{ALL}} - \hat{\epsilon}_1) \quad \text{and} \\ x_{\text{NAT},\beta,j} &= x_{\text{NAT},j} - \bar{x}_{\text{NAT}} + \hat{\beta}_{2,\text{NAT}}(\bar{x}_{\text{NAT}} - \hat{\epsilon}_2), \end{aligned} \quad (5)$$

where \bar{x} represents the ensemble mean and j distinguishes the ensemble members. Individual ensemble members and the ensemble means are expressed as anomalies relative to the mean of the ALL ensemble mean over the common period between models and observations. To be consistent with the calculation of the scaling factors, the NAT ensemble mean has had its time mean removed, and both ensemble means are multiplied by the square root of the ensemble size to rescale the variance. Furthermore, $x_{\text{ALL},\beta,j}$, $x_{\text{NAT},\beta,j}$, \bar{x}_{ALL} , and \bar{x}_{NAT} are prewhitened using the regularized covariance matrix, again for consistency with the ROF procedure used to fit the scaling factors. Then the appropriate ALL or NAT ensemble mean is subtracted from each member. Next, the ALL ensemble mean is scaled by $\hat{\beta}_{\text{ALL}}$ from the one-signal analysis, while the NAT ensemble mean is scaled by $\hat{\beta}_{\text{NAT}}$ from the two-signal analysis (ANT/NAT), taking into account the error in the response from the TLS analysis, $\hat{\epsilon}$. The rescaled ensemble means are added back to each ALL and NAT ensemble member, and the inverse of the prewhitening operator is used to return to the regular space. Because the scaling factors were fit using anomalies, the β -adjusted data will only be used with anomalies in subsequent analyses. Although the reference period for the calculation of the anomalies is slightly different (1979–2014 instead of 1981–2010) in order to match the procedure for calculating $\hat{\beta}$, we expect this to have minimal influence over the comparison of results.

b. Event attribution of extremes

Event attribution is used to assess the influence of anthropogenic forcings on the magnitude or frequency

of occurrence of a particular event. It is typical to choose extreme events with notable impacts (NASEM 2016). While the TLS method above is used for detection and attribution of the temporal pattern of the response to a specific forcing using a full time series record, a different approach is utilized for event attribution. The ensemble of responses can be used to determine the probability of a specific event occurring under each forcing scenario. Two metrics are typically used to assess the differences in probability with different forcings. The first is called the fraction of attributable risk (Stott et al. 2004):

$$\text{FAR} = 1 - \frac{p_0}{p_1}. \quad (6)$$

The FAR represents the fraction of the risk for the occurrence of a particular event that is due to the inclusion of additional forcing from one scenario to the next. Typically, p_1 represents the probability of the event under ALL forcing and p_0 the probability under NAT forcing. An alternate metric is the risk ratio (RR), a simple ratio that describes how many times more likely the occurrence of a specific event is under ALL forcing than under NAT forcing alone:

$$\text{RR} = \frac{p_1}{p_0}. \quad (7)$$

Hannart et al. (2016) present event attribution in terms of necessary and sufficient causation. According to their definitions, requiring the presence of a particular forcing scenario for an event to occur would be necessary causation. In contrast, if the particular forcing scenario always produces the event in question, this would be sufficient causation. Under certain assumptions (monotonicity and exogeneity), the probability of necessary causation [PN, using the notation of Hannart et al. (2016)] is equivalent to the FAR described above. The probability of sufficient causation (PS) can be calculated by

$$\text{PS} = 1 - \frac{1 - p_1}{1 - p_0}. \quad (8)$$

Both PN and PS will be calculated in the following analyses. It should be noted that the equations presented here only apply as PN or PS if the resulting values are greater than 0; if negative, the PN or PS is assigned a probability of 0.

These metrics will be analyzed for decadal periods. Values from each year in the period and each ensemble member will be combined into a pool of data. Through the use of a kernel density estimator

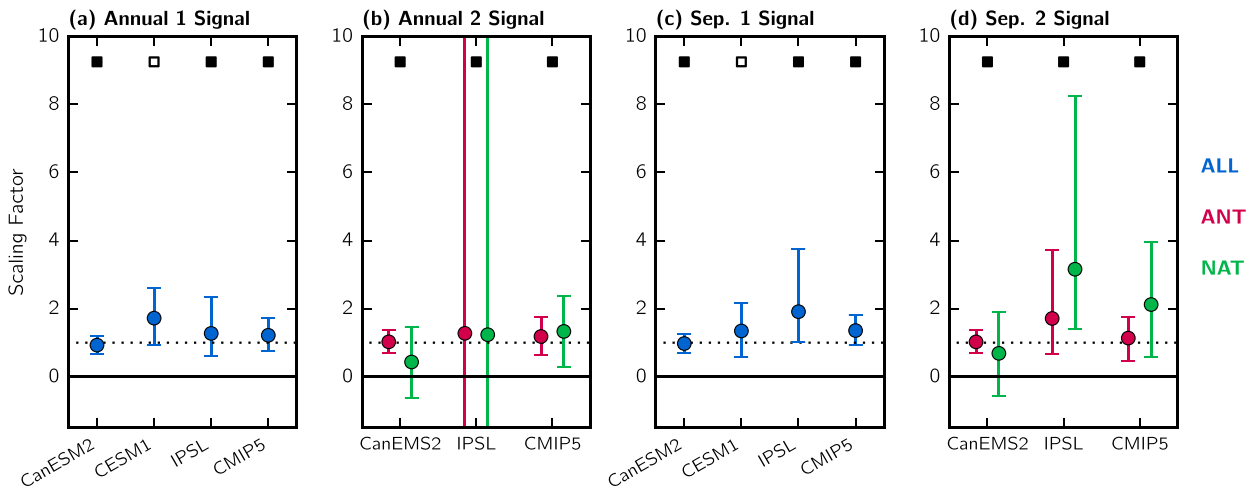


FIG. 5. Scaling factors and 90% confidence intervals for (a),(c) ALL one-signal analysis and (b),(d) the two-signal analysis of ANT (red) and NAT (green) for each model. Results are shown for annual values in (a),(b) and September values in (c),(d). Filled squares across the top of the panels indicate a passing of the residual consistency test, while open squares imply model variability is too low.

(a nonparametric method for estimating a probability density function), the p_1 and p_0 values for specific situations can be assessed. A Gaussian kernel will be used, as it was determined that the results show little sensitivity to this choice. For the event in question, p_0 and p_1 of experiencing an event equal to or more extreme than this threshold can be calculated by integrating the density function. To calculate the uncertainty on the FAR, PS, and RR, a resampling method similar to that of Stott et al. (2004) will be used. The pool of data from each month in the decade and from each ensemble member is resampled with replacement to reproduce a set of data of the same size to use for the fitting of the kernel densities and calculation of the probabilities. With 1000 estimates of the density curves, a nonparametric 90% confidence interval for each of the metrics can be determined.

Event attribution can be sensitive to the framing of the question (NASEM 2016; Otto et al. 2012), which includes the definition of the event. As such, we use several different metrics to define the extreme SIE events. These include the absolute values from each model simulation and anomalies calculated relative to the 1981–2010 period, as discussed previously. Additionally, the event will be represented as a fraction of the 1979–89 climatology; this period was chosen as the earliest period before the influence of anthropogenic forcings becomes strongly apparent and was limited by the observations. Finally, 5-yr means will also be analyzed; with the constraint from the IPSL and CMIP5 NAT forcing scenarios, the period of interest will be the mean of 2008–12. Time

series of these four metrics for each ensemble mean are shown in Fig. 3.

4. Results

a. Detection and attribution of temporal patterns

The scaling factors and corresponding 90% confidence intervals (CIs) for each model are displayed in Fig. 5 for both the annual time series and September alone. The annual time series is calculated as the sum of individual monthly means. For the one-signal analysis, scaling factors are shown for the ALL forcing, and filled squares imply a passing of the RCT. Consequently, open squares imply the RCT failed, which in this case indicates the model's estimate of internal variability is too low. (None of the cases shown failed the RCT with model internal variability too high.) The estimates of β_{ALL} for the annual one-signal analysis for all four models are inconsistent with zero, implying the signal is detected, and consistent with unity, which indicates good agreement between the observed and simulated SIE changes. All but CESM1 pass the RCT for the annual one-signal analysis, although both CESM1 and IPSL show wide confidence intervals. As the scaling factors are fitted based on SIE anomalies, any biases discussed previously should not affect the detection/attribution results.

In general, similar results are seen for the September one-signal analysis (Fig. 5c) as for the annual time series. One exception is that the wide confidence interval for IPSL's scaling factor does not include unity, implying a trend in the forced response that is smaller in magnitude than that observed. For CMIP5, the CI

just barely includes 1.0; a value of $\hat{\beta}_{\text{ALL}}$ greater than 1.0 would imply a smaller trend in the ensemble mean for September, and this is consistent with the results in Figs. 1c and 3.

For the two-signal analysis (Figs. 5b,d), the linear combination of ALL and NAT forcings is used to derive the scaling factors for ANT (red) and NAT (green). Two-signal results are not shown for CESM1, as this model does not have NAT forcing simulations. The CIs for β_{NAT} are generally wider than those for β_{ANT} or the one-signal β_{ALL} , expressing larger uncertainty on this estimate of β_{NAT} . All three models pass the RCT for the two-signal analyses. The ANT signal is detected for both CanESM2 and CMIP5 with a scaling factor consistent with 1. CanESM2 shows a very tight CI around $\hat{\beta}_{\text{ANT}}$, although the CI for β_{NAT} includes both 0 and 1. While we cannot state that the signal is detected, we also cannot state that it is different in magnitude from the NAT signal that may be present in the observations. Thus, the best interpretation of this scaling factor range would be that the signal is not detected but there is large uncertainty in the estimate (as in Christidis et al. 2012). Thus, it would be deduced that the observed signal is almost entirely explained by anthropogenic forcings using this model. For CMIP5, the NAT signal is also detected for both September and the annual time series, and $\hat{\beta}_{\text{NAT}}$ is larger than $\hat{\beta}_{\text{ANT}}$ with a CI that includes 1.0 but does not include 0, implying detection of this signal of a magnitude possibly similar to the observations. The detection of a NAT signal is likely influenced by the effect of volcanic forcing (see Fig. 1c). For IPSL, the CIs for both ANT and NAT in the annual time series are unbounded, which inhibits any detection or attribution inferences. For IPSL in September, the NAT signal is detected, even though greater than 1, while $\hat{\beta}_{\text{ANT}}$ is consistent with unity and exhibits a narrower CI. All three models pass the RCT for the two-signal analyses. Note that the β_{ANT} and β_{ALL} estimates are identical in the case of the two-signal analysis [see Eq. (4)].

The results of this detection and attribution analysis indicate that the CanESM2 large ensemble is the most suitable for an analysis of the attribution of extreme events in the SIE time series because of its strong detection of the anthropogenic signal and passing of the RCT. The CMIP5 ensemble is also fairly suitable, although the September ALL response is slightly smaller in magnitude than observed. CESM1 shows the necessary detection results for the one-signal analysis but fails to pass the RCT; variability in the residuals that is significantly larger than the model's internal variability decreases the credibility of the attribution result. IPSL passes the RCT for all cases, and

its wide CIs are likely due to the small size of its ensemble.

b. Event attribution

1) RESULTS FOR GENERAL THRESHOLDS

The attribution of a specific event involves the comparison of the probability of occurrences of that event under different forcing scenarios. This process is demonstrated through Fig. 6 for the adjusted CanESM2 data. Figure 6a uses the first of the four metrics for defining SIE events (Fig. 3), the absolute values, which are expressed in millions of square kilometers. The three left panels of Fig. 6 are density curves for ALL (blue) and NAT (green) from the model response for each decadal period. These curves are estimated from $10 \text{ yr} \times 50$ ensemble members. For this model, the NAT curve changes very little with time, while the ALL curve shifts farther to the left (to smaller SIE values) with each subsequent decade. Similar plots from previous decades (not shown) exhibited only small differences between the ALL and NAT curves; separation of the forced signals beginning in the 1990s agrees with the result of Min et al. (2008).

The second panel from the right in Fig. 6a displays the FAR curves for each decade (1990s in red, 2000s in yellow, and 2010s in purple), providing the fraction of the risk of experiencing a September SIE value at or below the given threshold that is attributable to the anthropogenic forcing included in the ALL scenario beyond the forcings in the NAT scenario. Note that more extreme (smaller) SIE values are found to the right. The FAR curves all saturate to 1.0 by the time SIE declines to $5 \times 10^6 \text{ km}^2$, indicating that any SIE less than this value is entirely attributable to the inclusion of ANT forcing (i.e., such SIE values could not occur with only NAT forcing). The observed 2012 SIE value (vertical bar) is smaller than the value at which the FAR reaches 1.0. Comparing to the density plots (Fig. 6a, left panels), the SIE thresholds where the FAR saturates to 1.0 are also where the NAT curves reduce to near-zero probabilities. The larger FAR values during earlier decades for larger SIEs are a result of the sensitivity to the shape of the density curves; because of the saturation of the FAR, the shift in the ALL curve with decade has little influence on this metric.

The far-right panel in Fig. 6a presents the RR curves for each of the three decades. The RR is only calculated within the realized range of values in the two responses for that decade, hence the greater extent of the latter decades' curves. By $5 \times 10^6 \text{ km}^2$, this event is approximately 10^4 times more likely to occur under ALL forcing than under NAT forcing alone. The RR values can get

CanESM2-adj.

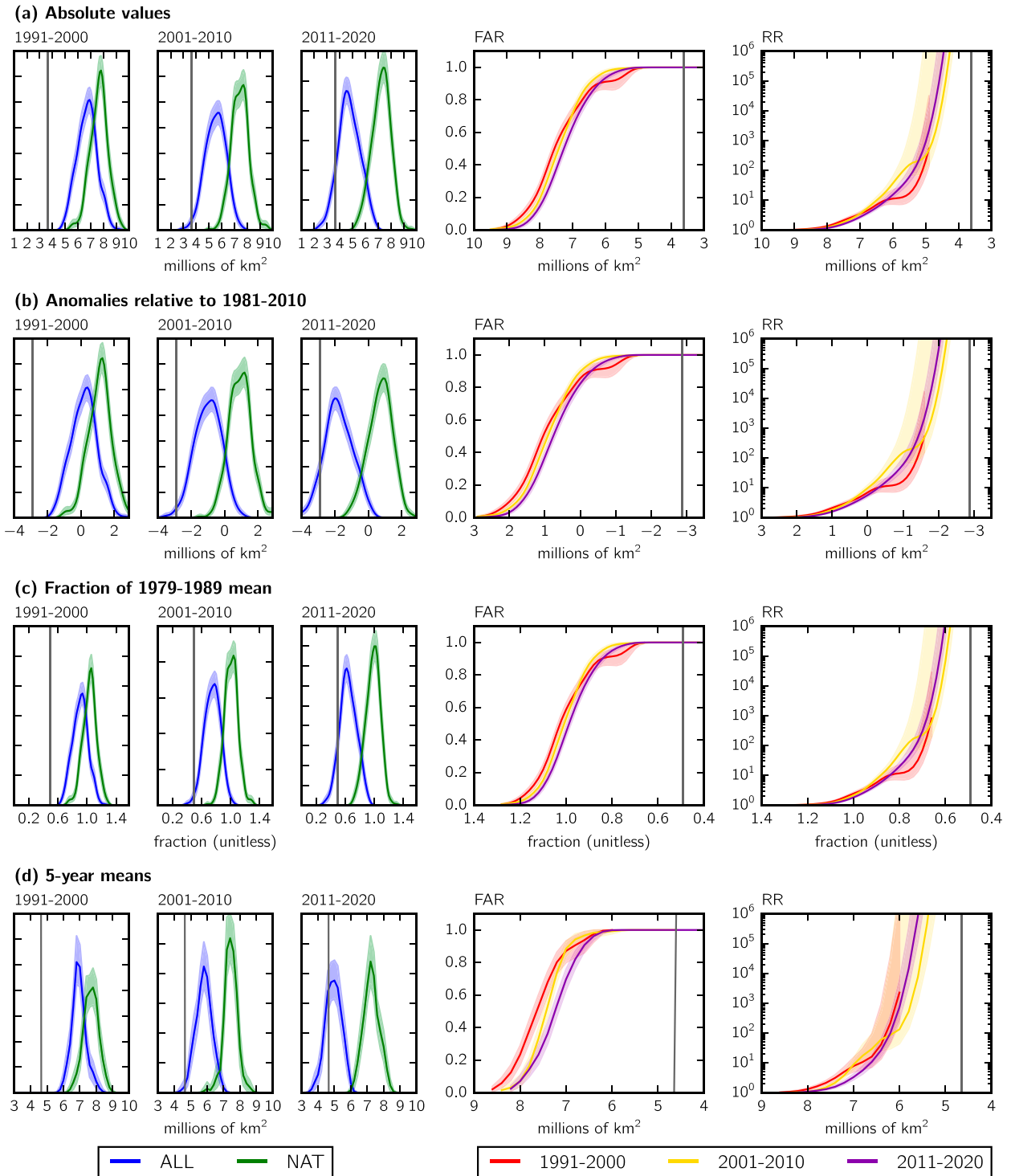


FIG. 6. Event attribution analysis using the adjusted CanESM2 ensemble for four different event definitions: (a) absolute values, (b) anomalies relative to 1981–2010, (c) fraction of the 1979–89 mean, (d) and 5-yr means. Kernel densities by decade for ALL (blue) and NAT (green) for September SIE are shown in the three left panels. Number of values into each curve is 500 from the number of ensemble members multiplied by 10 yr. FAR of an SIE extent more extreme than the given value is shown in the second panels from the right. RR of a more extreme SIE is shown in the right panels; an upper limit of 10^6 is applied. Vertical lines represent the observed 2012 event.

very large for small SIEs, and as p_1 approaches zero and p_0 is near zero, the RR can be sensitive to the estimation of the density curve and its integration. We will only consider RR values up to 10^6 , the inverse of which was determined to be a reasonable error when integrating a Gaussian distribution with single precision (e.g., Cody 1993).

Figure 6 also shows the same analyses for the other three metrics to define SIE events: anomalies relative to 1981–2010 (Fig. 6b), fraction of the 1979–89 mean (Fig. 6c), and 5-yr means (Fig. 6d). No matter how SIE events are characterized, the more extreme events have increased probability in the world with ALL forcings in later decades, while no significant changes by decade are seen in the world with NAT forcings. As a result, the FAR equals 1.0 for events of extreme magnitude, and the RR implies such events are much more likely with ALL forcings. Additionally, the 2012 event is a rare event with each metric (based on the density curves), even under ALL forcings, although an event of this magnitude or more extreme is considerably more likely to occur under ALL forcings than under NAT forcings alone.

Similar analyses for the other models and the unadjusted CanESM2 responses are included in the supplementary material (Figs. S3–S8). It should be noted that, because of the constraints from the availability of NAT forcings, IPSL (Figs. S5 and S6) and CMIP5 (Figs. S7 and S8) results are presented on slightly shifted decades (1983–92, 1993–2002, and 2003–12) from the CanESM2 (Fig. 6 and Fig. S3) and CESM1 (Fig. S4) results for 1990–2000, 2001–10, and 2011–20. Instead of using the same decades, CanESM2 and CESM1 results are presented using a more current, and more relevant, time period. All models and all metrics show increasing separation of the density curves of ALL and NAT with each subsequent decade. Furthermore, the FAR saturates to 1.0 within the realized range of all models. However, this saturation occurs at SIE values much greater than the 2012 event for some models (e.g., CESM1 and IPSL) and at SIE values more extreme than the 2012 event (e.g., the nonrelative metrics for CanESM2 and CMIP5). FAR values less than 1.0 for the 2012 event are likely due to the low bias of CanESM2 climatological SIE and the low bias and large ensemble spread of CMIP5.

Because of its much greater ensemble spread, the density curves for the CMIP5 results (Fig. S7) cover a wider range of values than those for the single-model ensembles. This increased spread results in smaller RR values than are seen with the other models and FAR curves that do not reach 1.0 until more extreme SIE values. Regardless, the conclusions are consistent with

those using other models. The uncertainty for the IPSL curves (Figs. S5 and S6) is much larger than that seen for other models as a result of the small ensemble size. Additionally, the 2012 event is outside the range of absolute values in both the adjusted and unadjusted IPSL responses. Even though the 2012 event is very rare (p_1 is very small), the FAR and RR results indicate the probability of such an event occurring is greater under ALL forcing than under NAT forcing.

To facilitate the comparison between models, Fig. 7 presents plots of the FAR (or PN), PS, and RR using the anomaly metric for all four ensembles. Because of the nature of the adjustment procedure, adjusted and unadjusted results will be the same for the anomaly metric. The decade from 2003 to 2012 is used for all models here, so the curves for CanESM2 and CESM1 are slightly different than those in Fig. 6 and Fig. S4. CanESM2, CESM1, and IPSL show FAR curves that reach 1.0 just before an anomaly of $-1 \times 10^6 \text{ km}^2$. The CMIP5 FAR curve increases at a slower rate than the other models because of the wider range of SIE values found in this ensemble. CMIP5 also presents smaller values of the RR, remaining below 10^6 until after the observed 2012 value of almost $-3 \times 10^6 \text{ km}^2$. Both CanESM2 and CESM1 present much greater values, well over this threshold.

The PS curves in Fig. 7b present large values for positive-anomaly thresholds and generally decrease rapidly with decreasing anomalies. The positive anomaly values are either common in the ALL forcing responses or more likely to occur with NAT forcing alone, resulting in PS values near 1.0, indicating that the inclusion of ALL forcing is sufficient for such events to occur, effectively guaranteeing the occurrence of an event of the given magnitude, or more extreme. Very small PS values, such as those for larger negative anomalies, imply such an event is rare, even under ALL forcing. Comparing the PS curves with those for PN (Fig. 7a), shows that for an extreme event in the current climate, such as that in 2012 (vertical bars), the inclusion of ANT forcing is necessary (PN = 1.0) but not sufficient (PS is small). That is, the inclusion of ALL forcing is required for such an event to occur but does not guarantee the event's occurrence.

The PS curve for the CMIP5 ensemble has slightly lower values for positive anomalies and shows a more gradual decrease, not reaching a PS value of 0 until values more extreme than $-4 \times 10^6 \text{ km}^2$. The difference between the PS results from the multimodel CMIP5 ensemble and from the single-model ensembles is predominantly due to the larger spread across the CMIP5 ensemble. With the inclusion of models with high and low biases, events in the tails of the single-model

ensembles carry more probability in the multimodel ensemble distribution. This disparity is much more pronounced when the CSIRO model is included in the CMIP5 ensemble (Fig. S9 in the supplementary material). In addition to smaller RR values and a more slowly increasing FAR curve, the PS curve shows a very different shape. With the inclusion of the CSIRO model, the ensemble contains more large, positive anomalies, and the greater ensemble mean leads the negative anomalies to be more extreme (Fig. S9). The wider, and multimodal, distribution results in smaller values of PS on either end.

The CMIP5 ensemble represents an “ensemble of opportunity” (Tebaldi and Knutti 2007). Several studies address the interpretation of these ensembles (e.g., Goldstein and Rougier 2009; Rougier et al. 2013), and, in practice, multiple methods are utilized for choosing an ensemble of a subset of the available realizations (e.g., Lutz et al. 2016; Cannon 2015; Collins et al. 2013). The differences between the CMIP5 results in Fig 7 and Fig. S9 demonstrate a potential disadvantage to using the ensemble of opportunity. Including a model deemed to have a very poor representation of the observations can produce a distinct difference in the results and conclusions.

2) FOCUS ON THE 2012 RECORD EVENT

To analyze the 2012 event specifically, Table 1 presents values of PN (FAR), PS, and RR from each model for an event as extreme as or more extreme than that observed in 2012. For comparison, the 2012 values of each model and metric are also presented, although the statistics are calculated using model response values from the 2003–12 decade. The 2012 values are taken from the ALL ensemble mean and are thus less extreme than observed, even for a dataset that compares well with the magnitude of observations. Calculating the mean across a large ensemble greatly reduces variability, so while none of these ensemble means see a 2012 value near as extreme as that observed (Fig. 3b), many of their individual members may.

The PN results are consistent across all scenarios, with values at or very close to 1.00 for all models and metrics. Thus, for all four metrics used to define SIE events, essentially all of the risk of the 2012 extreme event is attributable to the combination of anthropogenic and natural forcings (ALL), as opposed to natural forcing alone. Alternatively, a PN value of 1.00 indicates that ALL forcing is required for the 2012 event to occur (compared to only NAT forcings). Most model–metric combinations exhibit RR values greater than 10^6 , indicating that the probability of experiencing an event more extreme than that observed increases substantially

under ALL forcing compared to under NAT forcing. Even those models, such as the CMIP5 ensemble, that present RR values less than 10^6 still show a considerable increase in event probability with ALL forcing.

The PS values are very small for many of the models and metrics, which is because, in these cases, the 2012 event was rare, even in the ALL forcings scenario, and the inclusion of ALL forcings does not guarantee the occurrence of an event of this magnitude. If ALL forcings guaranteed the occurrence of such an event, it would not be “extreme” in the observations or simulations of the current climate. Larger values of PS indicate the 2012 event is more commonly realized in the ALL responses. For example, the 5-yr means for CanESM2 are significantly smaller than the observations (Fig. 3d), and all values from the ALL ensembles during the chosen decade are smaller than the observed 2012 value. This results in a PS value of 1.00, indicating that the inclusion of ALL forcings is sufficient to guarantee an event more extreme than the observed 2012 value in these simulations. In contrast, the 2012 event was rare with most other models and metrics, and the PS values are much smaller. While ALL forcings are necessary for an extreme event of 2012 magnitude to occur (see PN), they are generally not sufficient.

For CanESM2 and IPSL, both the original and adjusted results are shown in Table 1. Unlike IPSL, CanESM2 shows slightly different results between the original and bias-corrected responses. The RR values increase considerably with the bias correction. The bias correction was applied to both the ALL and NAT responses, and, prior to this adjustment, the NAT response overlapped with the observations during the latter decades (Fig. 1). Thus, the bias correction results in a shift of the density curves such that the NAT curve no longer contains the observed 2012 value, greatly decreasing the value of p_0 . The decrease in the PS values from the unadjusted to the adjusted response reflects the shifting of the ALL density curve such that the 2012 event became even more rare.

Table 1 also includes values of PN, PS, and RR for the β -adjusted response anomalies for each model. By rescaling the model responses using the scaling factors from the detection and attribution analysis of trends in the time series, response patterns that more closely match the observations are produced. For a model, such as CanESM2, with scaling factors very close to 1, the effect of this adjustment is minimal. Even for a model, such as IPSL, whose anomaly for the 2012 event is considerably more extreme than the realized anomaly from the unadjusted response, the conclusions drawn from the PN, PS, and RR values remain largely the

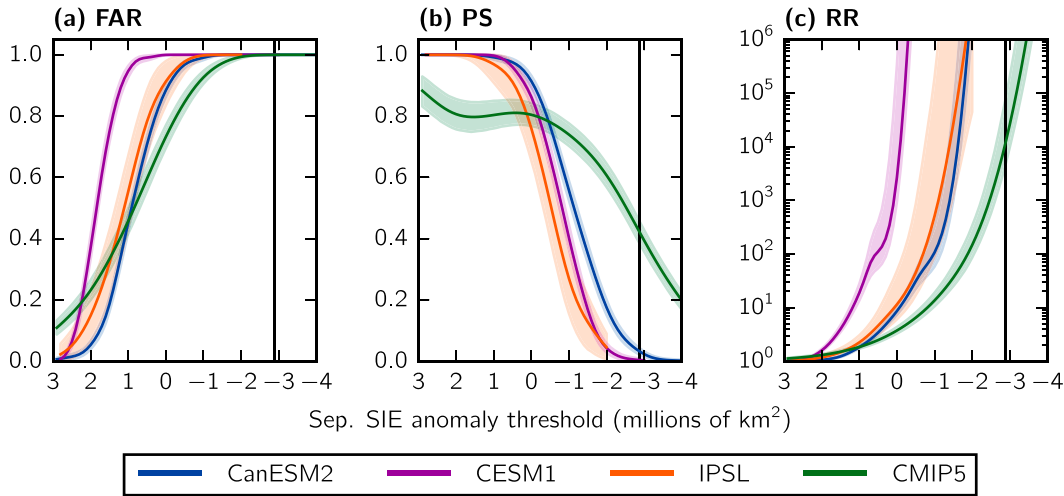


FIG. 7. For each model, (a) FAR (PN), (b) PS, and (c) RR of experiencing a September SIE value more extreme than the given threshold. The probabilities are calculated from densities fit to data from 2003–12 for anomalies relative to 1981–2010. These statistics are calculated only for values within the realized range of the ensemble for this period. Vertical lines represent the observed 2012 event. The RR plot in (c) is cut off at 10^6 .

same, although slightly weaker. Consistency between the results for the original and β -adjusted responses lends further robustness to the conclusion that the 2012 event is consistent with a scenario including anthropogenic influences and is very unlikely to occur with natural forcing alone.

3) FUTURE EXTREMES

To assess the implications of the event attribution analysis for future SIE events, the CanESM2 and CESM1 large ensembles provide model responses through 2100 and 2080, respectively. Figure 8a shows

the ensemble mean anomalies (relative to 1981–2010) for this extended time period. Given the base period means of $4.3 \times 10^6 \text{ km}^2$ for CanESM2 and $7.0 \times 10^6 \text{ km}^2$ for CESM1, the anomaly curves level off around ice-free conditions in the model. To calculate the event attribution statistics for a future decade, the period 2031–40 is chosen as the last decade during which both models still show annually varying September SIE values.

Density curves for each model during this decade are shown in Fig. 8b. For comparison, thirty 10-yr segments are extracted from CESM1’s adjusted PIC and the

TABLE 1. Values for 2012 from observations and four different ensembles (10^6 km^2), as well as the PN (FAR), PS, and RR for experiencing a value more extreme than observed. Four different metrics are used to calculate the SIE: the values for the anomalies, anomalies relative to 1980–2010, the fraction of the 1979–89 mean, and the 5-yr mean event ending in 2012. Models with biases (CanESM2 and IPSL) also include mean-adjusted data (adj), and all models include a version for the anomalies that has been adjusted by scaling factors (β ; Fig. 5). Large RR values exceeding 10^6 are indicated.

	Values				Anomalies				Fraction				5-yr mean			
	2012	PN	PS	RR	2012	PN	PS	RR	2012	PN	PS	RR	2012	PN	PS	RR
Observed	3.62	—	—	—	-2.89	—	—	—	0.49	—	—	—	4.65	—	—	—
CanESM2	2.47	0.98	0.708	45	-1.79	1.00	0.031	$>10^6$	0.48	1.00	0.261	10^4	2.86	0.89	1.000	8.9
CanESM2 (adj)	4.71	1.00	0.031	$>10^6$	-1.79	1.00	0.031	$>10^6$	0.64	1.00	0.032	$>10^6$	5.09	1.00	0.100	$>10^6$
CanESM2 (β)	—	—	—	—	-1.49	1.00	0.013	$>10^6$	—	—	—	—	—	—	—	—
CESM1	5.81	1.00	0.000	$>10^6$	-1.24	1.00	0.003	$>10^6$	0.77	1.00	0.000	$>10^6$	6.07	1.00	0.000	$>10^6$
CESM1 (β)	—	—	—	—	-3.63	1.00	0.223	$>10^6$	—	—	—	—	—	—	—	—
IPSL	6.73	1.00	0.000	$>10^6$	-0.86	1.00	0.000	$>10^6$	0.84	1.00	0.000	$>10^6$	6.87	1.00	0.000	$>10^6$
IPSL (adj)	5.64	1.00	0.000	$>10^6$	-0.86	1.00	0.000	$>10^6$	0.82	1.00	0.000	$>10^6$	5.79	1.00	0.000	$>10^6$
IPSL (β)	—	—	—	—	-3.31	0.75	0.133	3.9	—	—	—	—	—	—	—	—
CMIP5	4.21	0.99	0.61	87	-2.53	1.00	0.420	10^4	0.48	1.00	0.496	10^2	3.16	0.93	0.733	13
CMIP5 (β)	—	—	—	—	-3.98	0.97	0.698	36	—	—	—	—	—	—	—	—

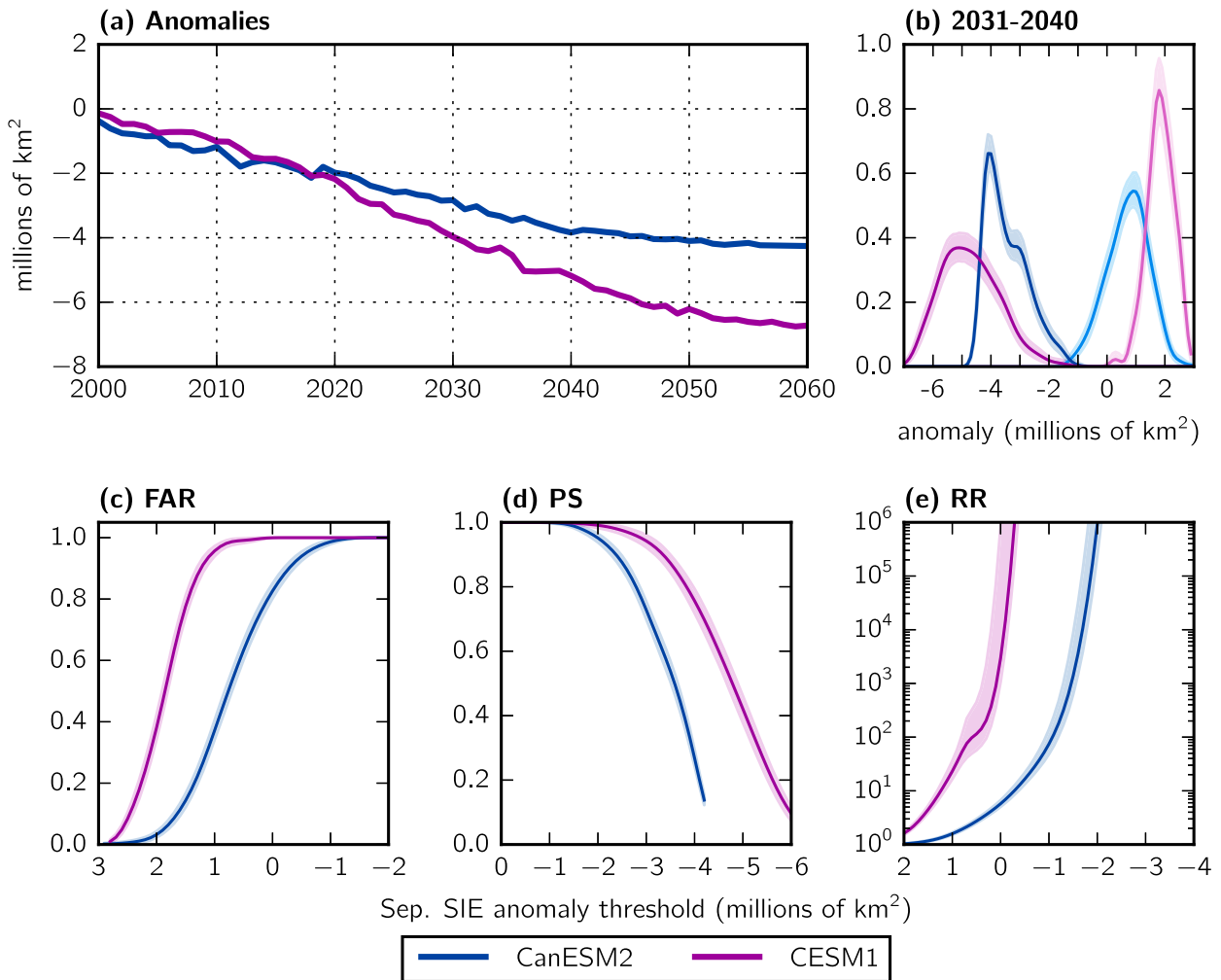


FIG. 8. (a) Anomalies relative to 1981–2010 for CanESM2 and CESM1 extending through 2060. (b) Kernel density estimates for the 2031–40 period for ALL data (darker colors), compared to a selection of the PIC for CESM1 and NAT from 2011–20 for CanESM2 (lighter colors). The resulting (c) FAR, (d) PS, and (e) RR comparing 2031–40 to 2011–20 are plotted only over the realized range of values, and, additionally, an upper limit of 10^6 is applied to the RR plot in (e).

lighter purple density curve is fit. As CanESM2 only has NAT responses through 2020, the NAT curve from 2011–20 will be used as reference (the light blue curve in Fig. 8b). Considering the inconsequential changes in the NAT curve by decade shown in Fig. 6, the choice of NAT period should make little difference here. Figure 6 and Fig. S4 showed diverging density curves with each subsequent decade through the 2010s, and Fig. 8b shows that, by the 2030s, the ALL and NAT curves are almost completely separated for both models.

The FAR, PS, and RR were calculated for a range of anomaly thresholds for both CanESM2 and CESM1 for the 2031–40 period, and the results are also shown in Figs. 8c–e, respectively. The FAR curves look very

similar to those seen in Fig. 7, with a saturated FAR by an anomaly of -1×10^6 km² for CanESM2 and around 1×10^6 km² for CESM1. The PS curves exhibit values close to 1.0 for smaller anomalies, implying that negative anomalies less extreme than -3×10^6 km² (the 2012 extreme event) are almost guaranteed with the inclusion of the additional ALL forcings in this decade. The RR values increase rapidly and reach 10^6 at an anomaly just less than 0 for CESM1 and around -2×10^6 km² for CanESM2. Extending beyond the previous conclusions, any negative anomalies relative to the 1981–2010 mean in future decades are extremely unlikely to have occurred in a NAT-forcing-only scenario and can be almost entirely attributed to the addition of anthropogenic forcings.

4) THE 2015 RECORD MINIMUM IN MARCH MAXIMUM SIE

Finally, although we have focused on September SIE minimum events, extremes in the March maximum are also of interest. Analyses similar to those presented above but looking at the March extremes are included in the supplementary material (Figs. S10–S13). As the current record minimum in the March maximum SIE occurred in 2015, only CanESM2 and CESM1 are used. Both models show decreasing trends in March SIE under ALL forcings (Fig. S10) that correspond well with the observed trend, and the simulated ensemble mean ALL forcings responses from both models can be detected in the observations (Fig. S12). For the 2011–20 period, both CanESM2 and CESM1 show FAR curves that saturate to 1.0 before the magnitude of the 2015 event (Fig. S13). The RR values, while smaller than for September, imply an event more extreme than the current record is still considerably more likely to occur under ALL forcing than under NAT forcing alone. The models disagree on the PS value for the 2015 event, and this is likely a result of the difference in the magnitude of the trends (Figs. S10 and S12) such that the 2015 event is more common in the CanESM2 responses while remaining rare in those from CESM1. In general, the broad conclusions found for September are also valid for the March results.

5. Discussion and conclusions

There are several potential sources of uncertainty in an event attribution analysis. [Christiansen \(2015\)](#) discusses the situation where both p_0 and p_1 are very small, which inhibits the interpretation of the FAR and may be due to the rarity of the particular event under both forcing scenarios, a poor calculation of the probabilities using a distribution with tails that are too light, and/or an inappropriate definition for the event in question. Regarding the second explanation, we have limited assumptions regarding the distribution by using kernel densities and have shown the probabilities to be robust to the choice of kernel. [Christiansen \(2015\)](#) mentions that a common problem with event definitions comes from the spatial region chosen. The choice of area for our study should not be an issue in this regard, as we use a single time series of SIE over an unambiguously defined region. We avoid the other concern with defining the event in question by using multiple metrics to quantify the extreme event. Although the 2012 event is chosen from observations, the model response probabilities are calculated using the containing decade and do not require each model to have observed its record

minimum in the same year. It was shown that the 2012 record minimum Arctic SIE was a rare event, even under the ALL forcing scenarios for all models. However, even though the event in question was rare, it was also shown to be very unlikely to occur with NAT forcing only. The rarity of the 2012 event in the ALL forcing simulations demonstrates the role of internal variability in the occurrence of such an event; while an extreme event like this one was shown to be much more likely to occur under ALL forcing, its occurrence still depends on a particular setup from natural variability.

Both [Christidis et al. \(2013\)](#) and [NASEM \(2016\)](#) emphasize the importance of first assessing a model's ability to reproduce the events in question before performing an event attribution analysis. Although the models used do not perfectly characterize Arctic SIE and related processes, large ensembles allow for a more thorough representation of internal variability. Furthermore, using models with negative (CanESM2) and positive (IPSL) SIE biases, models with underestimated (CMIP5 multimodel ensemble) and well-captured (CanESM2) trends, and ensembles with large spread (CMIP5) or with few ensemble members (IPSL) strengthens the conclusions when results are consistent. No single model captures all observed aspects of Arctic sea ice evolution perfectly; for example, one model with a low bias in the climatology reproduces the trend well (CanESM2). The fact that our results are not strongly sensitive to biases and consistent results are obtained when using a range of models strengthens our conclusions. Although the scaling factors and their accompanying uncertainty ranges may differ between models, the general conclusion is an attribution of trends in Arctic SIE anomalies to the inclusion of anthropogenic influences beyond natural forcings alone. Additionally, although the exact values of the statistics may differ, there is general agreement among models (including the original, mean-adjusted, and β -adjusted responses).

Through a detection and attribution analysis, it was demonstrated that the observed trends in annual and September Arctic SIE anomalies are consistent with the inclusion of anthropogenic forcings and are very unlikely to have occurred under the combined influence of natural forcings and internal variability alone. The event attribution showed agreement among models of a FAR value near 1.0, a small PS value, and a very large RR value for the 2012 event. This means that an event of the magnitude of the 2012 record minimum in Arctic SIE or more extreme is generally entirely attributable to the combination of anthropogenic and natural forcings and that such forcings are likely necessary for the occurrence

of the event (FAR near or equal to 1.0); that the 2012 event is extreme in the current climate, so the inclusion of anthropogenic forcings is not sufficient for event occurrence (small PS); and the event was much more likely to occur under ALL forcings than under NAT forcings alone (very large RR).

It was also shown that, for an event of the magnitude of the 2012 extreme minimum, the inclusion of anthropogenic forcings is a necessary but not a sufficient cause. That is, an event of this magnitude requires anthropogenic influences to occur, but the inclusion of these forcings alone is not enough to guarantee the occurrence of the event. Given that the occurrence of the 2012 event was extremely unlikely in its decade under natural forcings alone, combined with decreasing trends in SIE, it is virtually certain that future September minimum events more extreme than that in 2012 will be attributable to anthropogenic influences and that the inclusion of anthropogenic forcings will become a sufficient cause for events of the magnitude of the 2012 record event.

Acknowledgments. This work was supported by the NSERC Canadian Sea Ice and Snow Evolution (CanSISE) Network (NSERC Grant RGPC-433874-12). We thank Neil Swart, Xuebin Zhang, three anonymous reviewers, and the editor, John Walsh, for providing constructive comments that helped to improve the manuscript. We would also like to thank Neil Swart for his assistance with calculating SIE for the CMIP5 models and Reza Najafi for his help with the application of the detection and attribution methodology. The detection and attribution scripts were adapted from publicly available code from both Yang Feng and Aurélien Ribes.

REFERENCES

- Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Climate Dyn.*, **15**, 419–434, doi:10.1007/s003820050291.
- , and P. A. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting, part I: Theory. *Climate Dyn.*, **21**, 477–491, doi:10.1007/s00382-003-0313-9.
- Arora, V. K., and Coauthors, 2011: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.*, **38**, L05805, doi:10.1029/2010GL046270.
- Bindoff, N. L., and Coauthors, 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 867–952. [Available online at https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_Chapter10_FINAL.pdf.]
- Cannon, A. J., 2015: Selecting GCM scenarios that span the range of changes in a multimodel ensemble: Application to CMIP5 climate extremes indices. *J. Climate*, **28**, 1260–1267, doi:10.1175/JCLI-D-14-00636.1.
- Christiansen, B., 2015: The role of the selection problem and non-Gaussianity in attribution of single events to climate change. *J. Climate*, **28**, 9873–9891, doi:10.1175/JCLI-D-15-0318.1.
- Christidis, N., P. A. Stott, G. S. Jones, H. Shiogama, T. Nozawa, and J. Luterbacher, 2012: Human activity and anomalously warm seasons in Europe. *Int. J. Climatol.*, **32**, 225–239, doi:10.1002/joc.2262.
- , —, A. A. Scaife, A. Arribas, G. S. Jones, D. Copsey, J. R. Knight, and W. J. Tennant, 2013: A new HadGEM3-A-based system for attribution of weather- and climate-related extreme events. *J. Climate*, **26**, 2756–2783, doi:10.1175/JCLI-D-12-00169.1.
- , G. S. Jones, and P. A. Stott, 2015: Dramatically increasing chance of extremely hot summers since the 2003 European heatwave. *Nat. Climate Change*, **5**, 46–50, doi:10.1038/nclimate2468.
- Cody, W. J., 1993: Algorithm 715: SPECFUN—A portable FORTRAN package of special function routines and test drivers. *ACM Trans. Math. Software*, **19**, 22–32, doi:10.1145/151271.151273.
- Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1029–1136. [Available online at http://www.climatechange2013.org/images/report/WG1AR5_Chapter12_FINAL.pdf.]
- Comiso, J. C., C. L. Parkinson, R. Gersten, and L. Stock, 2008: Accelerated decline in the Arctic sea ice cover. *Geophys. Res. Lett.*, **35**, L01703, doi:10.1029/2007GL031972.
- Day, J. J., J. C. Hargreaves, J. D. Annan, and A. Abe-Ouchi, 2012: Sources of multi-decadal variability in Arctic sea ice extent. *Environ. Res. Lett.*, **7**, 034011, doi:10.1088/1748-9326/7/3/034011.
- 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, doi:10.1007/s00382-012-1636-1.
- Fetterer, F., K. Knowles, W. Meier, and M. Savoie, 2002: Sea Ice Index. National Snow and Ice Data Center, accessed 16 September 2015, doi:10.7265/NSQJ7F7W.
- Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866, doi:10.1017/CBO9781107415324.020.
- Gillett, N. P., M. F. Wehner, S. F. B. Tett, and A. J. Weaver, 2004: Testing the linearity of the response to combined greenhouse gas and sulfate aerosol forcing. *Geophys. Res. Lett.*, **31**, L14201, doi:10.1029/2004GL020111.
- , D. A. Stone, P. A. Stott, T. Nozawa, A. Y. Karpechko, G. C. Hegerl, M. F. Wehner, and P. D. Jones, 2008: Attribution of polar warming to human influence. *Nat. Geosci.*, **1**, 750–754, doi:10.1038/ngeo338.
- Goldstein, M., and J. Rougier, 2009: Reified Bayesian modelling and inference for physical systems. *J. Stat. Plann. Inference*, **139**, 1221–1239, doi:10.1016/j.jspi.2008.07.019.
- Gregory, J. M., P. A. Stott, D. J. Cresswell, N. A. Rayner, C. Gordon, and D. M. H. Sexton, 2002: Recent and future changes in Arctic sea ice simulated by the HadCM3 AOGCM. *Geophys. Res. Lett.*, **29**, 2175, doi:10.1029/2001GL014575.
- Guemas, V., F. Doblas-Reyes, A. Germe, M. Chevallier, and D. Salas y Méliá, 2013: Discriminating between sea ice memory, the August 2012 extreme storm, and prevailing

- warm conditions [in “Explaining Extreme Events of 2012 from a Climate Perspective”]. *Bull. Amer. Meteor. Soc.*, **94**, S20–S22.
- Hannart, A., J. Pearl, F. E. L. Otto, P. Naveau, and M. Ghil, 2016: Causal counterfactual theory for the attribution of weather and climate-related events. *Bull. Amer. Meteor. Soc.*, **97**, 99–110, doi:10.1175/BAMS-D-14-00034.1.
- Hartmann, D. L., and Coauthors, 2013: Observations: Atmosphere and surface. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 159–254. [Available online at http://www.climatechange2013.org/images/report/WG1AR5_Chapter02_FINAL.pdf.]
- Hegerl, G. C., and F. Zwiers, 2011: Use of models in detection and attribution of climate change. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 570–591, doi:10.1002/wcc.121.
- , and Coauthors, 2010: Good practice guidance paper on detection and attribution related to anthropogenic climate change. *Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Detection and Attribution of Anthropogenic Climate Change*, T. F. Stocker et al., Eds., 9 pp. [Available online at http://www.ipcc-wg2.gov/meetings/EMs/IPCC_D%26A_GoodPracticeGuidancePaper.pdf.]
- Herring, S., M. Hoerling, J. Kossin, T. Peterson, and P. Stott, Eds., 2015: Explaining extreme events of 2014 from a climate perspective. *Bull. Amer. Meteor. Soc.*, **96** (Suppl.), S1–S172, doi:10.1175/BAMS-ExplainingExtremeEvents2014.1.
- Hulme, M., 2014: Attributing weather extremes to ‘climate change’: A review. *Prog. Phys. Geogr.*, **38**, 499–511, doi:10.1177/0309133314538644.
- IPCC, 2014: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects*. Cambridge University Press, 688 pp.
- Jahn, A., and Coauthors, 2012: Late-twentieth-century simulation of Arctic sea ice and ocean properties in the CCSM4. *J. Climate*, **25**, 1431–1452, doi:10.1175/JCLI-D-11-00201.1.
- 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, doi: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, doi:10.1175/BAMS-D-13-00255.1.
- Lutz, A., H. ter Maat, H. Biemans, A. Shrestha, P. Wester, and W. Immerzeel, 2016: Selecting representative climate models for climate change impact studies: An advanced envelope-based selection approach. *Int. J. Climatol.*, **36**, 3988–4005, doi:10.1002/joc.4608.
- Massonnet, F., T. Fichefet, H. Goosse, C. M. Bitz, G. Philippon-Berthier, M. M. Holland, and P. Y. Barriat, 2012: Constraining projections of summer Arctic sea ice. *Cryosphere*, **6**, 1383–1394, doi:10.5194/tc-6-1383-2012.
- Min, S. K., X. Zhang, F. W. Zwiers, and T. Agnew, 2008: Human influence on Arctic sea ice detectable from early 1990s onwards. *Geophys. Res. Lett.*, **35**, L21701, doi:10.1029/2008GL035725.
- Mitchell, J. F. B., D. J. Karoly, G. C. Hegerl, F. W. Zwiers, M. R. Allen, and J. Marengo, 2001: Detection of climate change and attribution of causes. *Climate Change 2001: The Scientific Basis*, Cambridge University Press, 697–738.
- Najafi, M. R., F. W. Zwiers, and N. P. Gillett, 2015: Attribution of Arctic temperature change to greenhouse-gas and aerosol influences. *Nat. Climate Change*, **5**, 246–249, doi:10.1038/nclimate2524.
- NASEM, 2016: *Attribution of Extreme Weather Events in the Context of Climate Change*. The National Academies Press, 186 pp., doi:10.17226/21852.
- Notz, D., and J. Marotzke, 2012: Observations reveal external driver for Arctic sea-ice retreat. *Geophys. Res. Lett.*, **39**, L08502, doi:10.1029/2012GL051094.
- Otto, F. E. L., N. Massey, G. J. Van Oldenborgh, R. G. Jones, and M. R. Allen, 2012: Reconciling two approaches to attribution of the 2010 Russian heat wave. *Geophys. Res. Lett.*, **39**, L04702, doi:10.1029/2011GL050422.
- Peterson, T. C., M. P. Hoerling, P. A. Stott, and S. C. Herring, Eds., 2013: Explaining extreme events of 2012 from a climate perspective. *Bull. Amer. Meteor. Soc.*, **94** (Suppl.), S1–S74, doi:10.1175/BAMS-D-13-00085.1.
- Ribes, A., S. Planton, and L. Terray, 2013: Application of regularised optimal fingerprinting to attribution. Part I: Method, properties and idealised analysis. *Climate Dyn.*, **41**, 2817–2836, doi:10.1007/s00382-013-1735-7.
- Rougier, J., M. Goldstein, and L. House, 2013: Second-order exchangeability analysis for multimodel ensembles. *J. Amer. Stat. Assoc.*, **108**, 852–863, doi:10.1080/01621459.2013.802963.
- Sato, M., J. E. Hansen, M. P. McCormick, and J. B. Pollack, 1993: Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.*, **98**, 22 987–22 994, doi:10.1029/93JD02553.
- Serreze, M. C., and R. G. Barry, 2011: Processes and impacts of Arctic amplification: A research synthesis. *Global Planet. Change*, **77**, 85–96, doi:10.1016/j.gloplacha.2011.03.004.
- , M. M. Holland, and J. Stroeve, 2007: Perspectives on the Arctic’s shrinking sea-ice cover. *Science*, **315**, 1533–1536, doi:10.1126/science.1139426.
- Shepherd, T. G., 2016: A common framework for approaches to extreme event attribution. *Curr. Climate Change Rep.*, **2**, 28–38, doi:10.1007/s40641-016-0033-y.
- Shu, Q., Z. Song, and F. Qiao, 2015: Assessment of sea ice simulations in the CMIP5 models. *Cryosphere*, **9**, 399–409, doi:10.5194/tc-9-399-2015.
- Stone, D. A., and M. R. Allen, 2005: The end-to-end attribution problem: From emissions to impacts. *Climatic Change*, **71**, 303–318, doi:10.1007/s10584-005-6778-2.
- Stott, P. A., D. A. Stone, and M. R. Allen, 2004: Human contribution to the European heatwave of 2003. *Nature*, **432**, 610–614, doi:10.1038/nature03089.
- , N. P. Gillett, G. C. Hegerl, D. J. Karoly, D. A. Stone, X. Zhang, and F. Zwiers, 2010: Detection and attribution of climate change: A regional perspective. *Wiley Interdiscip. Rev.: Climate Change*, **1**, 192–211, doi:10.1002/wcc.34.
- , and Coauthors, 2016: Attribution of extreme weather and climate-related events. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 23–41, doi:10.1002/wcc.380.
- Stroeve, J., M. M. Holland, W. Meier, T. Scambos, and M. Serreze, 2007: Arctic sea ice decline: Faster than forecast. *Geophys. Res. Lett.*, **34**, L09501, doi:10.1029/2007GL029703.
- , M. Serreze, S. Drobot, S. Gearheard, M. Holland, J. Maslanik, W. Meier, and T. Scambos, 2008: Arctic sea ice extent plummets in 2007. *Eos, Trans. Amer. Geophys. Union*, **89**, 13–14, doi:10.1029/2008EO020001.
- Sun, Y., X. Zhang, F. W. Zwiers, L. Song, H. Wan, T. Hu, and H. Yin, 2014: Rapid increase in the risk of extreme summer heat in eastern China. *Nat. Climate Change*, **4**, 1082–1085, doi:10.1038/nclimate2410.
- Swart, N. C., J. C. Fyfe, E. Hawkins, J. E. Kay, and A. Jahn, 2015: Influence of internal variability on Arctic sea-ice trends. *Nat. Climate Change*, **5**, 86–89, doi:10.1038/nclimate2483.

- 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, doi:10.1098/rsta.2007.2076.
- Tett, S., P. Stott, M. Allen, W. Ingram, and J. Mitchell, 1999: Causes of twentieth-century temperature change near the Earth's surface. *Nature*, **399**, 569–572, doi:10.1038/21164.
- Van Huffel, S., and J. Vandewalle, 1991: *The Total Least Squares Problem: Computational Aspects and Analysis*. Society for Industrial and Applied Mathematics, 300 pp.
- van Vuuren, D. P., and Coauthors, 2011: The representative concentration pathways: An overview. *Climatic Change*, **109**, 5–31, doi:10.1007/s10584-011-0148-z.
- Vaughan, D., and Coauthors, 2013a: Observations: Cryosphere. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 317–382, doi:10.1017/CBO9781107415324.012.
- , and Coauthors, 2013b: Observations: Cryosphere supplementary material. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., 4SM-1–4SM-10. [Available online at https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/supplementary/WG1AR5_Ch04SM_FINAL.pdf.]
- Vinnikov, K. Y., and Coauthors, 1999: Global warming and Northern Hemisphere sea ice extent. *Science*, **286**, 1934–1937, doi:10.1126/science.286.5446.1934.
- Zhang, R., 2015: Mechanisms for low-frequency variability of summer Arctic sea ice extent. *Proc. Natl. Acad. Sci. USA*, **112**, 4570–4575, doi:10.1073/pnas.1422296112.
- , and T. R. Knutson, 2013: The role of global climate change in the extreme low summer Arctic sea ice extent in 2012 [in “Explaining Extreme Events of 2012 from a Climate Perspective”]. *Bull. Amer. Meteor. Soc.*, **94**, S23–S26.