Multimodel combination by a bayesian hierarchical model: Assessment of ice accumulation over the Oceanic Arctic Region

Malaak Kallache, Elena Maksimovich, Paul-Antoine Michelangeli, Philippe Naveau

▶ **To cite this version:**

Malaak Kallache, Elena Maksimovich, Paul-Antoine Michelangeli, Philippe Naveau. Multimodel combination by a bayesian hierarchical model: Assessment of ice accumulation over the Oceanic Arctic Region. Journal of Climate, 2010, 23 (20), pp.5421-5436. 10.1175/2010JCLI3107.1 . hal-03200934

HAL Id: hal-03200934
https://hal.science/hal-03200934

Submitted on 17 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multimodel Combination by a Bayesian Hierarchical Model: Assessment of Ice Accumulation over the Oceanic Arctic Region

MALAAK KALLACHE
CLIMPACT and LSCE-IPSL, Paris, France

ELENA MAKSIMOVICH
LOCEAN-IPSL, Paris, France

PAUL-ANTOINE MICHELANGELI
CLIMPACT, Paris, France

PHILIPPE NAVEAU
LSCE-IPSL, Gif-sur-Yvette, France

(Manuscript received 3 March 2009, in final form 20 April 2010)

ABSTRACT

The performance of general circulation models (GCMs) varies across regions and periods. When projecting into the future, it is therefore not obvious whether to reject or to prefer a certain GCM. Combining the outputs of several GCMs may enhance results. This paper presents a method to combine multimodel GCM projections by means of a Bayesian model combination (BMC). Here the influence of each GCM is weighted according to its performance in a training period, with regard to observations, as outcome BMC predictive distributions for yet unobserved observations are obtained. Technically, GCM outputs and observations are assumed to vary randomly around common means, which are interpreted as the actual target values under consideration. Posterior parameter distributions of the authors' Bayesian hierarchical model are obtained by a Markov chain Monte Carlo (MCMC) method. Advantageously, all parameters—such as bias and precision of the GCM models—are estimated together. Potential time dependence is accounted for by integrating a Kalman filter. The significance of trend slopes of the common means is evaluated by analyzing the posterior distribution of the parameters. The method is applied to assess the evolution of ice accumulation over the oceanic Arctic region in cold seasons. The observed ice index is created out of NCEP reanalysis data. Outputs of seven GCMs are combined by using the training period 1962–99 and prediction periods 2046–65 and 2082–99 with Special Report on Emissions Scenarios (SRES) A2 and B1. A continuing decrease of ice accumulation is visible for the A2 scenario, whereas the index stabilizes for the B1 scenario in the second prediction period.

1. Introduction

Today it is well accepted that global climate change has an effect on the polar regions (cf. Parry et al. 2007). The characteristics of the feedback is still under investigation and is related to important questions such as future sea level rise (cf. Charbit et al. 2008) and the influence of the polar regions on midlatitude patterns of atmospheric circulation and precipitation (Stroeve et al. 2007). General circulation model (GCM) projections are instrumental to exploring the effect of climate change in the future (see, e.g., Baettig et al. 2007). These models are run under contemporary conditions and possible future scenarios that reflect assumptions about the evolution of environment and society, especially the potential change in CO_2 emissions and concentrations. In this work two common greenhouse gas and aerosol emission scenarios (Nakicenovic and Swart 2000)—Special Report on Emissions Scenarios (SRES) A2 and B1—are employed for projections. The GCMs are based on complex dynamics and their performance varies across space and time. It is therefore not obvious whether to reject or

Corresponding author address: Malaak Kallache, CLIMPACT, 79 rue du Faubourg Poissonnière, 75009 Paris, France.
E-mail: mk@climpact.com

DOI: 10.1175/2010JCLI3107.1

© 2010 American Meteorological Society

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

prefer a certain GCM when projecting into the future. A simple combination of the different GCM outputs, which we denote the multimodel ensemble, might average out extremes. For diverse application fields it has nevertheless been shown that multimodel ensemble averaging (MEA) or Bayesian model averaging can generate better projections than each single model (cf. Raftery et al. 2005; van Loon et al. 2007).

Bayesian inference allows for the incorporation of expert knowledge and the assessment of parameter uncertainties. The approach has been increasingly used in climate change studies (see, e.g., Tol and De Vos 1998; Rougier 2007). For a general discussion of Bayesian statistics, refer to Congdon (2003) or Gelman et al. (2004). Combining multimodel ensembles by using a Bayesian approach has been pursued by a variety of studies. Among them, several are related to climate change and aim at the enhancement of projections. Coelho et al. (2004) regressed the multimodel ensemble mean on observations. This does not allow for a differentiation of bias and variance of the diverse models. Raftery et al. (2005) regressed observations on a weighted linear combination of model projections. Weights, which indicate the importance of each model, and regression parameters were estimated in a training period for which observations were available. Then they were transferred to a prediction period. Regression on the sum of model projections may cause overfitting when too many models are used. Therefore, the number of models and the length of the training period play an important role in this approach. Min and Hense (2006) proceeded likewise, where the weights for the linear combination of model outputs were obtained from Bayes factors. Tebaldi et al. (2005) treated observations and model outputs as random variables. Both data sources varied around common means. These means were interpreted as unobserved values of the assessed variable, the target values. In this way a connection between model outputs and observations is established. Tebaldi et al. (2005) explored the change of mean temperature between the training and prediction periods. Here stationarity of the variable was assumed within each period. Berliner and Kim (2008) followed a similar approach, but they allowed for time-dependent variables under consideration by applying a data assimilation framework. The target values were modeled as autocorrelated hidden states, which are influenced by covariates. They estimated a potential bias of each model as averaged deviation from these target values.

The study presented here is in line with Tebaldi et al. (2005) and Berliner and Kim (2008). We allow for a temporal gap between the training and prediction periods. That is, contrary to Tebaldi et al. (2005), projections are not regressed on model outputs of the training period. Model outputs and observations are assumed to vary around common means, that is, both have an error. In the training period, model outputs and observations are combined to a weighted average, whereas we assume that observations have no bias. In this way a potential bias and the variability of each GCM are estimated in comparison to the observations. This differs from Berliner and Kim (2008), who did not directly involve observations to estimate these parameters. A Kalman filter is integrated in a Markov chain Monte Carlo (MCMC) routine to estimate the parameters of our Bayesian hierarchical model (Chib and Greenberg 1996). This allows for sequential prediction and for updating steps and the assessment of time-dependent variables, which are especially required when analyzing the effect of climate change. The common means of model outputs and observations are modeled as hidden states comprising a stationary autoregressive and a trend component. As application we assess the evolution of an ice index.

We proceed as follows: In section 2 the data are described and the construction of the ice index for the oceanic Arctic region is presented. The Bayesian model combination (BMC) framework is outlined in section 3 and its application to the ice index is illustrated in section 4. Last, a summary of our results and a conclusion are given in section 5.

2. Ice accumulation over the oceanic Arctic region

Sea ice area reductions are related to increases of surface air temperature in the polar regions (Hassol 2004); such reductions have been well documented by observational and modeling studies (cf. Zhang and Walsh 2006; Cavalieri and Parkinson 2008). Temperature-based ice indices represent primarily sea ice thickness and volume: the intensity of sea ice accumulation during the freezing season is a result of heat fluxes between ocean and air reservoirs through snow-covered sea ice. Sea ice accumulation has, for example, been estimated by Anderson (1961) and Maykut (1986), who utilize daily near-surface temperature fields for this purpose. Sea ice thickness and volume are important for sea ice–related feedbacks (Stroeve et al. 2007). They influence, for example, ice forming and drifting processes.

Here atmospheric surface temperatures (SATs), 2 m above the surface, are used to calculate an ice index as a proxy for the ice accumulation over the oceanic Arctic region. This ice index has been developed within the European Union (EU) project Developing Arctic Modeling and Observing Capabilities for Long-term Environmental Studies (DAMOCLES; available online at http://www.damocles-eu.org). We use daily SAT data to calculate the index, either provided by GCMs or by

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

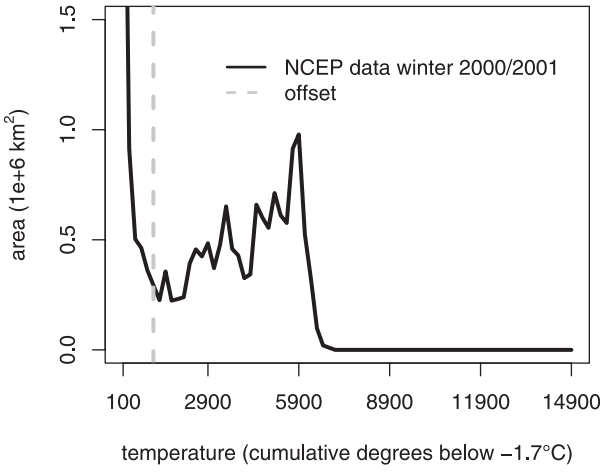

FIG. 1. Accumulated area over oceanic Arctic region, classified by iciness. The level of iciness is measured by the sum of degrees below −1.7°C per freezing season (September–May). The ice index is created as an integral over the black curve, weighted by the classes of iciness (*x* axis). Areas below the offset (gray line) are disregarded.

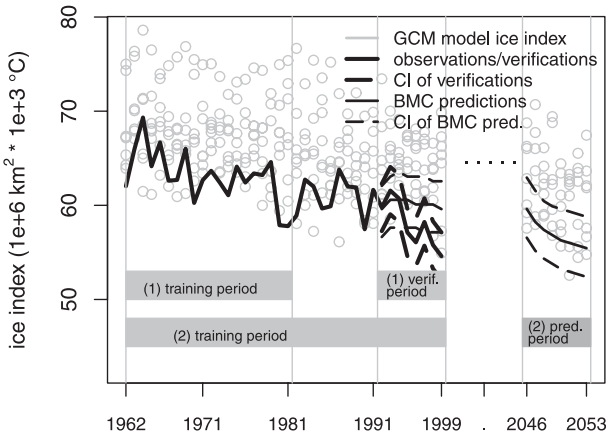

FIG. 2. Outline of the BMC framework: first, the framework is inspected by projecting from a training period to a verification period, where observations are also given (1); then, a training period is chosen to project to a future period (2). The winter ice index generated from NCEP reanalysis data (with 66% confidence intervals in the verification period) is shown as thick black lines, and the BMC projections (with 66% credibility intervals) as thin black lines.

National Centers for Environmental Prediction (NCEP) reanalysis data. The ice index was designed in the following way: The area over the oceanic Arctic region was classified regarding its iciness over the freezing season (September–May). The level of iciness was estimated by taking at each grid point the sum of degrees below −1.7°C over the season. The iciness classes had a width of 200°C. Data at different grid points represent areas of distinct size; in this way, area and classes of iciness are linked. Figure 1 exemplarily shows the levels of iciness of the oceanic Arctic region for the freezing season 2000/01 and NCEP reanalysis data. The winter ice index was obtained by taking the integral of the classified area, that is, the black curve in Fig. 1. Here the summands of the Riemann integral have been weighted with the iciness class levels and the area of the warmest classes, where the sum of degrees below −1.7°C is smaller than 1100°C, were disregarded. In this way the focus is set on the colder levels, and the intensity of freezing conditions during the polar winter is explored. This ice index, therefore, is related to sea ice volume accumulated by the end of winter rather than to sea ice extent. Further details are provided in Maksimovich and Gascard (2010). We analyze the ice index for the historical period 1962–99 and projection periods 2046–65 and 2082–99 and SRES A2 and B1. NCEP reanalysis data are used to calculate the observed ice index. Outputs of 13 Intergovernmental Panel on Climate Change (IPCC) models cover all three periods and are therefore chosen for our analysis (see Fig. 3 for model acronyms). Within the historical period, the models are run with the twentieth-century run (20c3m) scenario, that is, with increasing greenhouse gases and anthropogenic sulfate aerosol forcing as observed through the twentieth century. The model surface temperature data have been obtained from the Program for Climate Model Diagnosis and Intercomparison (available online at www-pcmdi.llnl.gov).

3. Bayesian model combination

Let BMC denote our technique to average over many competing models by setting up a statistical model. The quantities of interest are parameters of this model, which are estimated while evaluating GCM outputs and observations. The effect of each GCM is weighted according to its performance in a training period, where observations are available for comparison. In this way model uncertainty is incorporated in the calculation of the BMC projections. Expert knowledge may be integrated as prior information. The aim of BMC is the improvement of predictive performance. For the analysis we proceed in two steps, as illustrated in Fig. 2. First, the BMC framework is inspected on a verification period. Here the BMC projections are compared to observations not used in the calibration. The BMC results in predictive distributions. We take the mean of these distributions as actual BMC projections. Their uncertainty is deduced by calculating credibility intervals from the predictive distributions. In Fig. 2 BMC projections of the ice index (black line) are depicted together with GCM projections (gray dots) and observations (thick black lines). For the verification period shown, the BMC projections are located much closer to the observations

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

than the projections of each single model. A more detailed evaluation of the performance of the BMC is provided by using scores, see section 4a. In a second step, we choose a training period to project to future periods, where no observations are available. The bias of the GCMs is estimated in the training period and maintained for the projections. Figure 2 shows, for example, that the BMC projections lie below the output of each single model, just like the observations in the training period. Our goal is to project the yet unobserved observations Y_0, which we take as mean of the predictive distribution $[Y_0|\mathbf{D}]$. Here the data vector $\mathbf{D} = (\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_M, \mathbf{Y}_1, \ldots, \mathbf{Y}_M)$ is given by the vectors of observations $\mathbf{X}_0$ and model outputs in the training period $\mathbf{X}_i$ and in the prediction period $\mathbf{Y}_i$. In our example $\mathbf{X}_0$ is a winter ice index created of surface temperature over the oceanic Arctic region and $\mathbf{X}_i$ denotes this ice index computed by model i over the same region and at the same period.

Bayesian inference presumes that parameters are not point estimates but have a distribution. For each model parameter, say, θ_i, a prior distribution is assumed, which comprises our knowledge about the uncertainty of the parameter values. Then we learn from the data $\mathbf{D}$ to obtain the posterior distribution of the parameters, say, $[\boldsymbol{\Theta}|\mathbf{D}]$. This is formalized with Bayes's theorem (cf. Gelman et al. 2004). Let $[\mathbf{D}|\boldsymbol{\Theta}]$ denote the conditional distribution of $\mathbf{D}$ given $\boldsymbol{\Theta}$, that is, the likelihood of the data. We can say that $[\boldsymbol{\Theta}|\mathbf{D}]$ is proportional to the product of likelihood and prior distribution of the parameters, that is,

$$[\boldsymbol{\Theta}|\mathbf{D}] \propto [\boldsymbol{\Theta}][\mathbf{D}|\boldsymbol{\Theta}]. \tag{1}$$

Integration over all parameters leads to the predictive distribution of Y_0 given $\mathbf{D}$:

$$[Y_0|\mathbf{D}] = \int_{\boldsymbol{\Theta}} [Y_0|\boldsymbol{\Theta}][\boldsymbol{\Theta}|\mathbf{D}]\, d\boldsymbol{\Theta}. \tag{2}$$

Equation (2) implies the conditional independence of Y_0 and $\mathbf{D}$ given $\boldsymbol{\Theta}$. For further details see Congdon (2003).

a. State-space model

Climate projections for the Arctic often exhibit trends, which are superimposed by decadal oscillations (Döscher et al. 2009). To be able to distinguish and track characteristics due to potentially different processes, we model the evolution of the ice index in time. To do so a sequential data assimilation procedure is employed within our Bayesian framework (cf. DelSole 2007).

We permit a gap of G' time steps between the training and prediction periods and introduce $G = T + G'$; thus $t = G + 1, \ldots, G + P$ is the time index of the prediction period of length P (see Fig. 2). At each time point t in $1, \ldots, T, G + 1, \ldots, G + P$, there is a prediction and update step of the parameter distributions. Let $\mathbf{D}_t$ denote the vector of data per time step,

$$\mathbf{D}_t = \begin{cases} (X_{0_t}, X_{1_t}, \ldots, X_{M_t}) & \text{for} \quad t = 1, \ldots, T, \\ (Y_{1_t}, \ldots, Y_{M_t}) & \text{for} \quad t = G + 1, \ldots, G + P. \end{cases} \tag{3}$$

The relations between model outputs and observations are given by the *data equations*

$$\begin{aligned} X_{i_t} &= c_T + \mu_t + d_t + b_i + \epsilon_{i_t} \quad \text{for} \quad t = 1, \ldots, T, \quad i = 0, \ldots, M \quad \text{and} \\ Y_{i_t} &= c_P + \nu_t + \delta_t + b_i + e_{i_t} \quad \text{for} \quad t = G + 1, \ldots, G + P, \quad i = 1, \ldots, M, \end{aligned} \tag{4}$$

with $b_0 = 0$, and Gaussian distributions for the noise, that is, $\epsilon_{i_t} \sim N(0, \lambda_i^{-1}), e_{i_t} \sim N[0, (\gamma\lambda_i)^{-1}]$. Here $N(\mu, \sigma^2)$ denotes normal distribution with mean μ and variance σ^2 as parameters; λ_i is the precision of model i and is equal to the reciprocal value of the variance. The precision of the observations λ_0 reflects the natural variability specific to the season and other physical factors. In our model λ_0 is externally given and estimated from the observations. The parameter γ allows for a different model precision for the training and prediction periods. It is assumed that all models experience the same degree of change, that is, γ is the same for all models. Moreover, it is presumed that each GCM has a constant bias b_i, and this bias is transferred from the training period to the prediction period.

The common means of model outputs and observations are modeled as composition of an intercept, a stationary autoregressive component, and a trend component. This gives the common means for the training period $\mathrm{MT}_t = c_T + \mu_t + d_t$ and for the prediction period $\mathrm{MP}_t = c_P + \nu_t + \delta_t$. Empirical series often exhibit autocorrelated noise. Here the assumption of uncorrelated noise may bias the estimate of the magnitude of the systematic change (cf. Bloomfield 1992). Therefore, we allow for an autocorrelated part, which is a common approach in time series analysis (cf. Cohn and Lins 2005; Harvey et al. 2007). The results for the ice index assessment confirm this choice: we find significant autocorrelation for all settings analyzed (see section 4d). The stationary autoregressive components μ_t and ν_t and the linear trend

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

components d_t and δ_t are modeled separately. By doing so, we differentiate systematic changes and other variations of the common mean (see, e.g., West 1997).

The temporal dependence is modeled by the following *state-space equations:*

$$\begin{aligned}
\mu_t &= \phi\mu_{t-1} + \omega_{\mu_t} \quad \text{for} \quad t = 1, \ldots, T,\\
d_t &= d_{t-1} + k_1 + \omega_{d_t} \quad \text{for} \quad t = 1, \ldots, T,\\
\nu_t &= \phi\nu_{t-1} + \omega_{\nu_t} \quad \text{for} \quad t = G+1, \ldots, P, \quad \text{and}\\
\delta_t &= \delta_{t-1} + k_2 + \omega_{\delta_t} \quad \text{for} \quad t = G+1, \ldots, P,
\end{aligned} \tag{5}$$

with $\omega_{\mu_t} \sim N(0, \lambda_\mu^{-1})$, $\omega_{d_t} \sim N(0, \lambda_d^{-1})$, $\omega_{\nu_t} \sim N(0, \lambda_\nu^{-1})$, and $\omega_{\delta_t} \sim N(0, \lambda_\delta^{-1})$ being independent and identically distributed. The state-space parameters are constructed such that they have the Markov property (Chib and Greenberg 1996). Here μ_t and ν_t are autoregressive (AR) components of order one [AR(1) processes]. We suppose short-term variability to be an intrinsic characteristic of the system assessed; therefore, the autoregressive parameter ϕ is transferred from the training period to the prediction period. On the other hand, there may be a gap of several decades between the training and prediction periods. Consequently, transferring trend slope estimates from the training period to the prediction period is not feasible. Here the unknown potential trend in the prediction period is therefore inferred from the GCM projections. In this way trend slope and intercept may differ in both periods; k_1 and k_2 denote the trend slopes times Δt, which is the time difference between two consecutive observations (see, e.g., Dethlefsen and Lundbye-Christensen 2005).

The parameter vector at time t is given by

$$\boldsymbol{\Theta}_t = \begin{cases} (c_T, \mu_t, d_t, \lambda_1, \ldots, \lambda_M, \phi, k_1, b_1, \ldots, b_M, \lambda_\mu, \lambda_d) & \text{for} \quad t = 1, \ldots, T, \\ (c_P, \nu_t, \delta_t, \lambda_1, \ldots, \lambda_M, \gamma, \phi, k_2, b_1, \ldots, b_M, \lambda_\nu, \lambda_\delta) & \text{for} \quad t = G+1, \ldots, G+P. \end{cases} \tag{6}$$

Conditional independence from the data of previous time steps given the parameter values is supposed, that is,

$$\begin{aligned}
[\boldsymbol{\Theta}_t|\boldsymbol{\Theta}_{t-1}, \mathbf{D}_1, \ldots, \mathbf{D}_{t-1}] &= [\boldsymbol{\Theta}_t|\boldsymbol{\Theta}_{t-1}] \quad \text{for} \quad t = 1, \ldots, T, G+1, \ldots, G+P,\\
[\mathbf{D}_t|\boldsymbol{\Theta}_t, \mathbf{D}_1, \ldots, \mathbf{D}_{t-1}] &= [\mathbf{D}_t|\boldsymbol{\Theta}_t] \quad \text{for} \quad t = 1, \ldots, T, G+1, \ldots, G+P.
\end{aligned} \tag{7}$$

Thus, the model given by Eqs. (4) and (5) has the form

$$\mathbf{D}_t = \mathbf{H}_t\boldsymbol{\Theta}_t + \mathbf{K}_t + \boldsymbol{\Sigma}_t \quad \text{for} \quad t = 1, \ldots, T, G+1, \ldots, G+P \quad \text{and} \tag{8}$$

$$\boldsymbol{\Theta}_t = \mathbf{F}_t\boldsymbol{\Theta}_{t-1} + \mathbf{U}_t + \mathbf{W}_t \quad \text{for} \quad t = 1, \ldots, T, G+1, \ldots, G+P, \tag{9}$$

with $\boldsymbol{\Sigma}_t$ and $\mathbf{W}_t$ being the respective error matrices. The state-space parameters are assumed to be independent; consequently, both matrices have nonzero values only in the diagonal; $\mathbf{H}_t$, $\mathbf{K}_t$, $\mathbf{F}_t$, and $\mathbf{U}_t$ contain either zeros or constants and can be deduced from Eqs. (4) and (5).

To keep the number of parameters low, we chose a simple structure of the state-space variables; however, more complicated structures can be easily integrated (cf. Harvey et al. 2007). The distinction between trend and autoregressive components may be difficult, as outlined in Kallache et al. (2005). A state-space model, as presented in Eq. (5), is a possibility to represent time series, which may consist of those both components. Its appropriateness has been explored by Koop and Van Dijk (2000), for example. Moreover, we verify the adequateness of our model given by Eqs. (4) and (5); the results are provided in section 4a.

b. Sequential updating

As common in data assimilation, alternating update and prediction steps of the state-space variables are performed. Here this is also done in the prediction period, and the GCM outputs are treated as biased observations. A projection Y_{0_t} is based on this evolution.

The *update formula* or *filtering formula* is given by

$$\begin{aligned}
[\boldsymbol{\Theta}_{t-1}|\mathbf{D}_1, \ldots, \mathbf{D}_{t-1}] &= \frac{[\mathbf{D}_{t-1}|\boldsymbol{\Theta}_{t-1}][\boldsymbol{\Theta}_{t-1}|\mathbf{D}_1, \ldots, \mathbf{D}_{t-2}]}{[\mathbf{D}_{t-1}|\mathbf{D}_1, \ldots, \mathbf{D}_{t-2}]}\\
&\propto [\mathbf{D}_{t-1}|\boldsymbol{\Theta}_{t-1}][\boldsymbol{\Theta}_{t-1}|\mathbf{D}_1, \ldots, \mathbf{D}_{t-2}].
\end{aligned} \tag{10}$$

With respect to $\boldsymbol{\Theta}_t$, the marginal likelihood of the data $[\mathbf{D}_t|\mathbf{D}_1, \ldots, \mathbf{D}_{t-1}]$ is a constant, so we neglect this term. This update step is equivalent to the calculation of the posterior distribution of the parameters at time step t, with the prior distribution given as $[\boldsymbol{\Theta}_t|\mathbf{D}_{t-1}]$.

The *prediction formula* is denoted by

$$\begin{aligned}
&[\boldsymbol{\Theta}_t|\mathbf{D}_1, \ldots, \mathbf{D}_{t-1}]\\
&\quad = \int_{\boldsymbol{\Theta}_{t-1}} [\boldsymbol{\Theta}_t|\boldsymbol{\Theta}_{t-1}][\boldsymbol{\Theta}_{t-1}|\mathbf{D}_1, \ldots, \mathbf{D}_{t-1}]\, d\boldsymbol{\Theta}_{t-1},
\end{aligned} \tag{11}$$

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

and Eqs. (10) and (11) are the Bayesian solution to calculate the state-space model given in Eqs. (8) and (9).

The state-space model consists of linear and Gaussian equations; therefore, we apply the Kalman filter to obtain the posterior distribution of the parameters. In this way the integral in Eq. (11) can be derived by characteristics of the normal distribution. Let $\mathbf{\Theta}_{t-1}|\mathbf{D}_{t-1} \sim N(\hat{\mathbf{\Theta}}_{t-1}, \mathbf{S}_{t-1})$ with $\hat{\mathbf{\Theta}}_{t-1}$ and $\mathbf{S}_{t-1}$ being the expected value and the variance of $\mathbf{\Theta}_{t-1}|\mathbf{D}_{t-1}$, and then $\mathbf{\Theta}_t|\mathbf{D}_{t-1} \sim N(\mathbf{F}_t\hat{\mathbf{\Theta}}_{t-1}\mathbf{F}_t', \mathbf{F}_t\mathbf{S}_{t-1}\mathbf{F}_t' + \mathbf{W}_t)$ (see, e.g., Meinhold and Singpurwalla 1983).[1]

c. Prior distributions

We choose uninformative priors because of the lack of prior knowledge. The precisions λ_i, $i = 1, \ldots, M$ and λ_μ, λ_d, λ_ν, and λ_δ are assumed to have uniform prior densities $U(0, c)$. A uniform prior for the precision λ is proportional to a uniform prior for the standard deviation σ (see, e.g., Gelman 2006). Similarly, we choose $\gamma \sim U(0, c)$ as a prior for γ. The upper bound $c = 1\,000\,000$ is chosen to be large to include any plausible prior value and to avoid an impact of this choice on the results. The uniform distribution is conditionally conjugate to the normal distribution. Therefore, this prior results in a posterior gamma distribution, similar to when choosing an informative gamma distribution as prior. However, by using a uniform prior, we avoid distortions of the posterior (cf. Harvey et al. 2007) and do not have to specify hyperparameters for a gamma prior, which may result in an improper posterior density in case these hyperparameters tend to zero.

For μ_t and d_t a nearly flat prior, that is, a normal distribution with variance near to infinity, is selected, and the same applies for ν_t and δ_t, respectively. We model a trend and an autoregressive component. Therefore, the autoregressive component is assumed to be stationary, and for the parameter ϕ a uniform prior on the interval $(-1, 1)$ is chosen. The trend slope parameters k_1 and k_2, the intercepts c_T and c_P, and the model biases $b_1, \ldots, b_M$ have uniform priors on the real line.

d. Calculation of the posterior distribution

The joint posterior distribution of the parameters $[\mathbf{\Theta}|\mathbf{D}]$ is the target density, which we will obtain by means of Eq. (1) and sequential calculation of the time dependent parameters.

The likelihood of the data is given by

$$[\mathbf{D}|\mathbf{\Theta}] = [\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_M, \mathbf{Y}_1, \ldots, \mathbf{Y}_M|\mathbf{\Theta}] \propto \prod_{i=1}^{M}\left(\sqrt{\gamma^P}\sqrt{\lambda_i^{(T+P)}}\exp\left\langle -\frac{\lambda_i}{2}\left\{\sum_{t=1}^{T}[X_{i_t} - (c_T + \mu_t + d_t + b_i)]^2 + \gamma\sum_{t=G+1}^{G+P}[Y_{i_t} - (c_P + \nu_t + \delta_t + b_i)]^2\right\}\right\rangle\right) \times \exp\left\{-\frac{\lambda_0}{2}\sum_{t=1}^{T}[X_{0_t} - (c_T + \mu_t + d_t)]^2\right\}. \quad (12)$$

To consider the dependencies within the parameters, we separately assess the constant components of the parameter vector $\mathbf{\Theta}_{\text{stat}} = (c_T, c_P, \gamma, \phi, k_1, k_2, \lambda_1, \ldots, \lambda_M, b_1, \ldots, b_M, \lambda_\mu, \lambda_\nu, \lambda_d, \lambda_\delta)$ and the state-space components μ_t, d_t, ν_t, and δ_t. The posterior distribution of the constant parameters is given by

$$[\mathbf{\Theta}_{\text{stat}}|\mu, d, \nu, \delta, \mathbf{D}] \propto [\mathbf{D}|\mathbf{\Theta}] \times [\mu, d, \nu, \delta|\mathbf{\Theta}_{\text{stat}}] \times [c_T] \times [c_P] \times [\gamma] \times [\lambda_1] \times \ldots \times [\lambda_M] \times [\lambda_\mu] \times [\lambda_d] \times [\lambda_\nu] \times [\lambda_\delta] \times [\phi] \times [k_1] \times [k_2] \times [b_1] \times \ldots \times [b_M]. \quad (13)$$

Inference of the analytic form of the posterior distribution of the parameters cannot be drawn, but the priors of the stationary parameters are conjugate to the likelihood. Thus, the marginal conditional densities of the parameters are given, and a MCMC simulation through a Gibb's sampler is employed to approximate the marginal posterior distributions of the parameters (cf. Tebaldi et al. 2005). Further details are provided in the appendix. The posterior distribution of the intercepts c_T and c_P is normal with variance $[T(\lambda_0 + \sum_{i=1}^{M}\lambda_i)]^{-1}$ and $(\gamma P\sum_{i=1}^{M}\lambda_i)^{-1}$. The mean of the intercepts is a weighted average of observations and model outputs, whereas the precisions λ_i serve as weights:

[1] The vector $\mathbf{F}_t'$ is the transpose of $\mathbf{F}_t$.

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

$$\mu_{c_T} = \frac{\sum_{t=1}^{T}\left\langle \sum_{i=1}^{M}\{\lambda_i[X_{i_t} - (\mu_t + d_t + b_i)]\} + \lambda_0[X_{0_t} - (\mu_t + d_t)]\right\rangle}{T\left(\lambda_0 + \sum_{i=1}^{M}\lambda_i\right)} \quad \text{and} \quad \mu_{c_P} = \frac{\sum_{i=1}^{M}\sum_{t=G+1}^{G+P}\lambda_i[Y_{i_t} - (\nu_t + \delta_t + b_i)]}{P\sum_{i=1}^{M}\lambda_i}. \tag{14}$$

The posterior distribution of the state-space parameters $[\mu_t, \nu_t, d_t, \delta_t|\boldsymbol{\Theta}, \mathbf{D}]$ at each time step $t = 1, \ldots, T$, $G + 1, \ldots, G + P$ is obtained by applying the update and prediction steps of Eqs. (10) and (11) by means of a Kalman filter–based simulation smoother (Meinhold and Singpurwalla 1983; De Jong and Shephard 1995; Durbin and Koopman 2002; Harvey et al. 2007). The Kalman filter and smoother are run at each Gibb's sampling iteration (i) with the current stationary parameters $\boldsymbol{\Theta}_{\text{stat}}^{(i)}$.

The predictive distribution Y_{0_t} at time t is derived by using the corresponding parameter vector $\boldsymbol{\Theta}_t$ for Eq. (2). This parameter vector is actually calculated by utilizing all data information $\mathbf{D}$, because a Kalman smoother is employed for the state-space variables (cf. Dethlefsen and Lundbye-Christensen 2005).

e. Approximation of the predictive distribution

Let the conditional distribution of Y_0 given the parameters be $Y_0|\boldsymbol{\Theta} \sim N(c_P + \nu_t + \delta_t, \lambda_0^{-1})$. The assumption of a similar variability of the observations in the training and prediction periods is commonly used (cf. Min et al. 2004; Raftery et al. 2005). Other modeling approaches would require detailed physical knowledge of the prediction period. We use Monte Carlo integration to calculate the predictive distribution of Y_{0_t}; that is, equation (2) is interpreted as the calculation of the expectation of $[Y_0|\boldsymbol{\Theta}]$, which is the predictive distribution (Davison 2003). As an approximation of the posterior predictive distribution, we get

$$[Y_{0_t} = y_{0_s}|\mathbf{D}] = R^{-1}\sum_{r=1}^{R}[Y_0 = y_{0_s}|\boldsymbol{\Theta}_r], \tag{15}$$

where $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_R$ represent draws from the posterior probability of the parameters $[\boldsymbol{\Theta}|\mathbf{D}]$, which are obtained by Gibb's sampling. Equation (15) is evaluated for each y_{0_s}, $s = 1, \ldots, S$, where the set of y_{0_s} covers the whole domain of the posterior predictive distribution; R denotes the number of Gibb's sampling iterations kept for evaluation (see the appendix for further details).

4. Ice index assessment

The BMC framework is applied to an ice index representing the ice accumulation over the oceanic Arctic region for a freezing season (see section 2). The index is derived from surface temperature, a climatic variable, and is not directly related to sea ice physics.

We combine single runs of several GCMs; thus, the natural variability of the GCMs themselves is not considered. However, Gregory et al. (2002) and Zhang and Walsh (2006) do not find a major influence of internal model variability on the trends of sea ice extent. Given the strong decrease of the ice accumulation index for the prediction periods (see section 4d), we deduce that internal model variability may also be neglected for the quantitative interpretation of the ice index assessment results.

Our results are verified by comparisons with an ice index created from reanalysis data. Bromwich et al. (2007) find NCEP reanalysis data to be a suitable tool to study the Arctic region, despite some deficiencies. Sufficiently accurate observations were not available until now. Sea ice volume or thickness can only be measured with large uncertainty, since the data are derived from submarine sonar measurements, which do not have sufficient coverage (Gregory et al. 2002), or from satellite measurements (Laxon et al. 2003; Kwok and Rothrock 2009).

a. Verification

For verification, training periods of different lengths are used, namely, freezing seasons spanning 5 (1977–81), 10 (1972–81), 15 (1967–81), and 20 (1962–81) yr. The verification period covers from 1992 to 1999 and the derived predictive distributions are compared to observations for this period.

We utilize common measures for verification, such as the mean absolute error (MAE) or the root-mean-square error (RMSE). Furthermore, we employ the continuous ranked probability score (CRPS) and the ignorance score (IS) to evaluate distributions. The CRPS gets better the closer a verification y is to the center of the predictive cumulative distribution function $F(\cdot)$. The CRPS is the integral of the Brier score at all possible threshold values t for the continuous predictand and is defined as

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty}[F(t) - H(t - y)]^2\,dt, \tag{16}$$

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

where $H(t - y)$ denotes the Heaviside function with $H(t - y) = 0$ for $t < y$ and $H(t - y) = 1$ otherwise (see Gneiting et al. 2005). To approximate the CRPS, we use the discretized predictive distribution evaluated at 200 quantiles (cf. Hersbach 2000). The IS is the negative logarithm of the predictive density $f(\cdot)$ at the verification y, that is,

$$\mathrm{IS}(f, y) = -\log f(y). \tag{17}$$

Both scores indicate a better performance when having a lower value. All comparisons are done with the average of these scores over the whole verification period.

To further evaluate the Bayesian framework, we carried out simulation studies with artificial data (results not shown). The data were generated from a Gaussian white noise process or from an AR(1) process and a linear trend or no trend was added. Then we calibrated the BMC model in a training period and projected to a verification period. The scores described in this section were used as an evaluation criterion. In all cases the simulation studies showed good performance of the BMC framework and superiority to just taking the multimodel ensemble mean.

b. Size of the multimodel ensemble

Thirteen GCMs cover the historical period and the two projection periods (see section 2). In Fig. 3 the precision λ_i of all models is shown. The precision signifies how closely the GCMs vary around the common means. It is an indicator for the quality of the models, and the contribution of each model to the constant part of the common means, c_T and c_P, is weighted by the precision, which is apparent in Eq. (14).

Six of the GCMs include natural forcing in their 20C3M runs, such as variations in solar input and volcanic forcing. Most of those models have a high precision (see Fig. 3). However, anthropogenic forcing is assumed to be the main factor for the increase of northern SAT (Kaufman et al. 2009) and the decrease of Arctic sea ice extent (cf. Gregory et al. 2002; Johannessen et al. 2004) observed for the last third of the twentieth century. It is expected that Arctic sea ice extent will continue to decline through the twenty-first century also because of atmospheric greenhouse gas loading (Stroeve et al. 2007). To evaluate the potential advantage of using solely GCMs, which include natural forcing, we compare BMC ice index projection skills of the sets of GCMs with and without natural forcing (results not shown). The projection period of 1989–95 is taken, because here Stroeve et al. (2007) found a stronger downward trend for sea ice extent, which is not well reproduced by the IPCC Fourth Assessment Report (AR4) models. Stroeve et al. (2007) attributed this effect to an intense positive

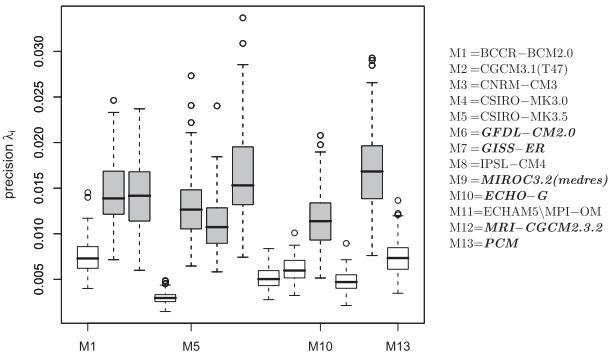

FIG. 3. Box plots of GCM precisions λ_i, which are estimated by using the historical period 1962–99. The models with high precision are accented in gray. Models that include natural forcing in their 20C3M run are listed in bold italic.

state of the winter northern annular mode, which in turn is linked to solar activity (cf. Ruzmaikin and Feynman 2002). Thus, strong differences of the ice index projections for this period are expected. However, the set of GCMs that include natural forcing did not ameliorate the reproduction of the observed decline of the ice index for this period. This indicates that other model characteristics—for example, the implementation of sea ice, ocean and atmospheric physics, and the coupling between those modules—might have a stronger influence on the ice index evolution than the inclusion of natural forcing. Thus, we consider all 13 GCMs as valid candidates for our analysis.

Figure 3 reveals that some models have a very low precision, which points at reducing the number of GCMs. To decide whether to delimit the number of models for the projection, we compare the prediction scores of the set of all models and a selection of seven models with good precision (the precision box plot of the selection is accented gray in Fig. 3). The prediction scores of the smaller set of models are on a par with the set of all models for all verification settings except the one with 15 yr of training length (see Fig. 4). Therefore, we chose to utilize the selection of seven models with the best precision in the further analysis.

c. Length of the training period

The influence of the length of the training period is evaluated by comparing prediction scores for different training period lengths and a verification period from 1992 to 1999 (see section 4a). Results are depicted in Fig. 4. In Figs. 4a,b the RMSE and MAE for the BMC projections and the MEA are shown. Apparently, the BMC projections outperform projections made by just taking the multimodel ensemble average. The CRPS and IS scores for different training period lengths are depicted

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

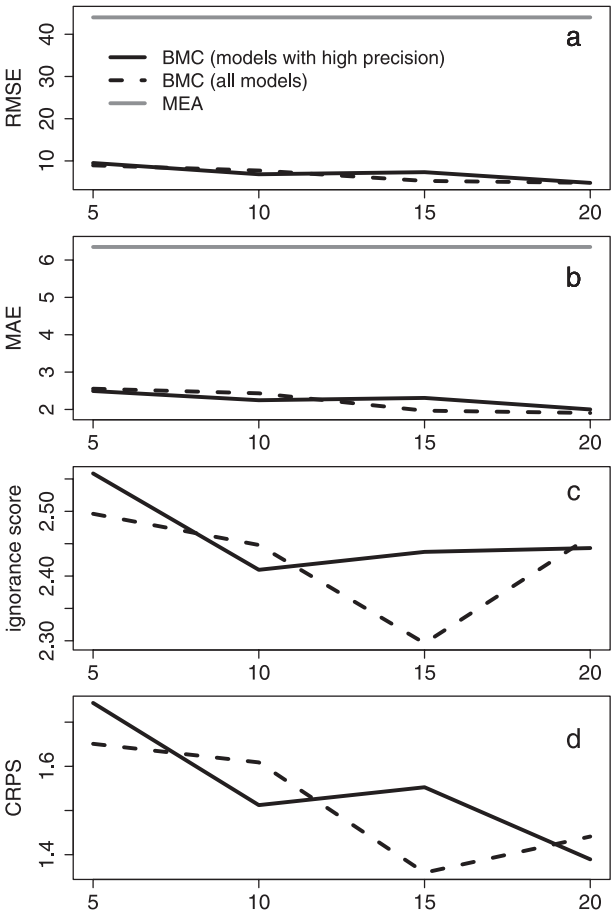

FIG. 4. (a) RMSE, (b) MAE, (c) IS, and (d) CRPS for the BMC projections using the seven models with the highest precision (black solid line), the BMC projections using all the models (black dashed line), and the MEA projections (gray line) for training periods using freezing seasons spanning 5 (1977–81), 10 (1972–81), 15 (1967–81), and 20 (1962–81) yr. The verification period is from 1992 to 1999.

in Figs. 4c,d, respectively. They reveal that a longer training period does not necessarily result in better scores. However, the longest training period leads to comparatively good results. We also could not find any physical reasons to dismiss parts of the historical data, and no signs of overfitting because of too much information were apparent. Therefore, we chose the whole historical period (1962–99) as the training period for the projections. Here also the scores of the set of seven models with high precision, which we choose for the analysis, are on par with the BMC projections of all the models.

d. Results

Projections are carried out for the A2 and B1 scenarios. Scenario A2 is a rather pessimistic scenario; a regionalized, heterogeneous world with high population growth and energy use and slow technological evolution is assumed, which results in high CO_2 concentrations hitting about 840 ppm at the end of the twenty-first century. By contrast B1 expects low population growth and energy use and a medium technical evolution, and is therefore a low-emission scenario. In Fig. 5 exemplarily marginal posterior parameter distributions for the projection to years 2082–99 and the A2 scenario are shown. The posterior parameter distributions are a means to test for significance. The slope of the trend in the training period (Fig. 5a) and prediction period (Fig. 5b) differ in size, but both are significant; that is, zero is not included in the slope distributions. With a length of approximately 20 years, the prediction periods are relatively short. Therefore, the term significance is related to the existence of a trend in presence of noise rather than pointing to an irreversible downward trend. For the B1 scenario, no significant trend could be found for the years 2082–99. The parameter ϕ, shown in Fig. 5c, reflects the autocorrelation present in the data, and in Fig. 5d a model bias is depicted. The BMC projections are the basis for the examination of the future ice accumulation over the oceanic Arctic region. This might give insight into the minimal ice accumulation of the oceanic Arctic ice at the end of the summer, in case the relation to SAT temperature is the same for the end of the twentieth and twenty-first centuries, which is assumed when projecting the ice index. This index allows for a qualitative assessment of the ice accumulation, since quantitatively the ice amount estimated from NCEP reanalysis data differs from the actual amount of ice. Figure 6 shows projections for years 2046–65 and 2082–99 and the A2 (Fig. 6a) and B1 (Fig. 6b) scenarios. Apparently, the variability of the mean of the predictive distributions, which we take as point projection, is very low. This might be a side effect of Kalman filtering. However, the projection is expected to occur within the bandwidth of the whole predictive distribution. Thus, the variability of the observations is more or less preserved for the projections. The first projection period shows a comparable decline of the ice index for both scenarios. Under the A2 scenario, this tendency of reduction of the ice accumulation over the oceanic Arctic region in the freezing season continues in the years 2082–99, whereas for the B1 scenario, interestingly, a stabilization of the ice accumulation is indicated. Correspondingly, Zhang and Walsh (2006) and Gregory et al. (2002) find a decreasing trend for sea ice extent for those scenarios and the twenty-first century. However, they do not analyze the evolution of this trend.

To evaluate the choice of the state-space model, we compare our results in Fig. 7. Here the evolution of the

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

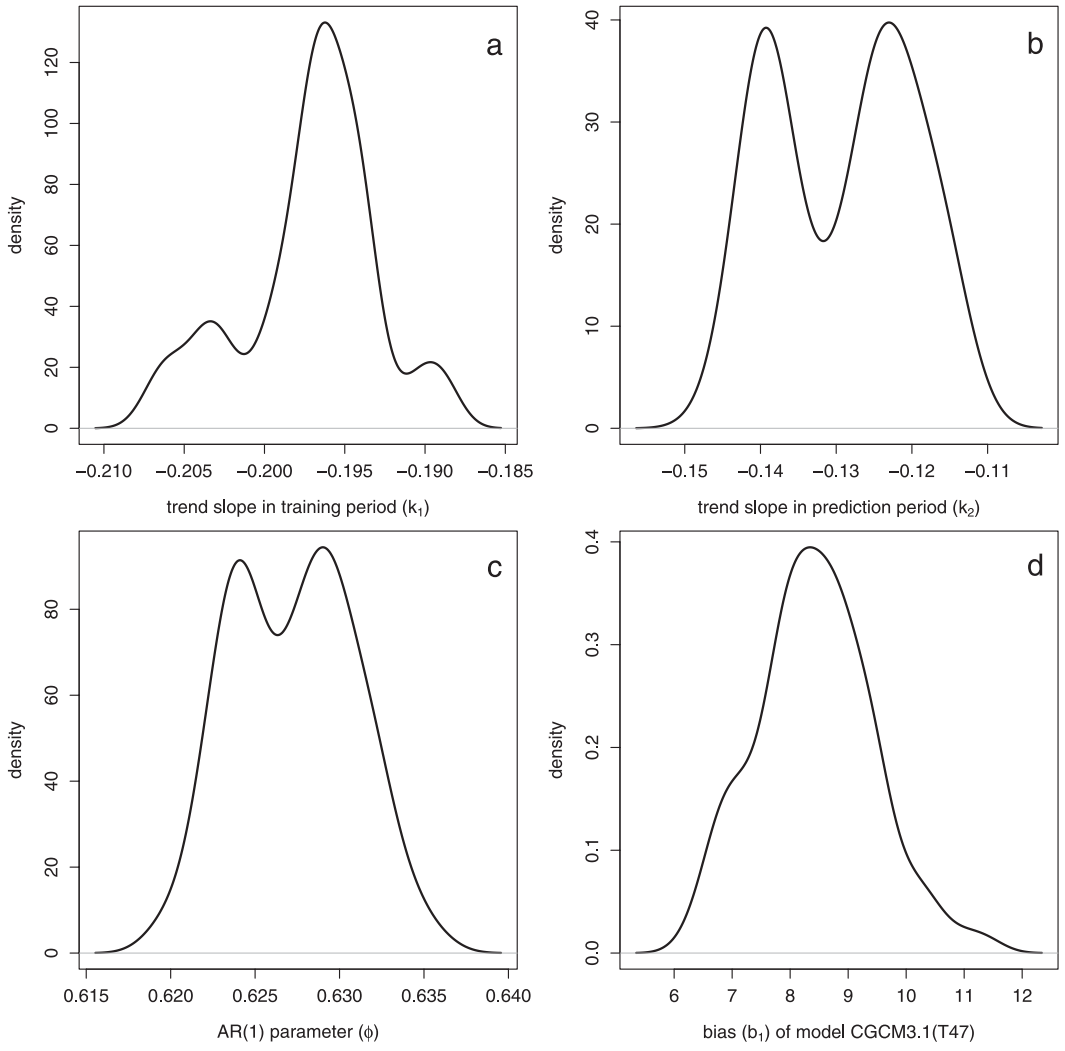

FIG. 5. Posterior distribution of the trend slope in the (a) training period k_1 and (b) prediction period k_2, (c) AR(1) parameter ϕ, and (d) estimated bias b_1 of Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled General Circulation Model, version 3.1 [CGCM3.1(T47)] for the projection to years 2082–99 under the A2 scenario.

distribution of the common means MT_t is shown for the training period (Fig 7a) and MP_t for the prediction period 2082–99 for the B1 scenario (black lines; Fig. 7b). Furthermore, the following state-space models are assessed: The common means being composed of intercept and trend (gray lines), of intercept and AR component (black dashed lines), and of a constant intercept only (gray dashed lines). The width of the densities reflects the uncertainty of the estimate of the common means, not the variability of the projected observations. We find that the projected common means are estimated with approximately the same accuracy as the common means in the training period; all distributions span about 4×10^9 km^2 °C. Furthermore, the less complex state-space models are capable of estimating the projected common means with comparable good accuracy. Similar results are obtained for the other three projection periods (not shown). In Fig. 8 results for the four state-space models are compared for the A2 scenario projection period 2046–65. In Fig. 8a the densities of the difference of the common means in the training and projection periods are shown (the densities of the nonstationary state-space models have been averaged over the training period and the prediction period, respectively). This important figure indicates the expected average change of the ice index. In line with Fig. 7, it is nearly the same for all four state-space models. The mean of the densities is at about -16×10^9 km^2 °C; that is, a decrease of the ice index of 26% is expected. We find for the B1 scenario and this period an expected change of 22.9%, and for the second

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

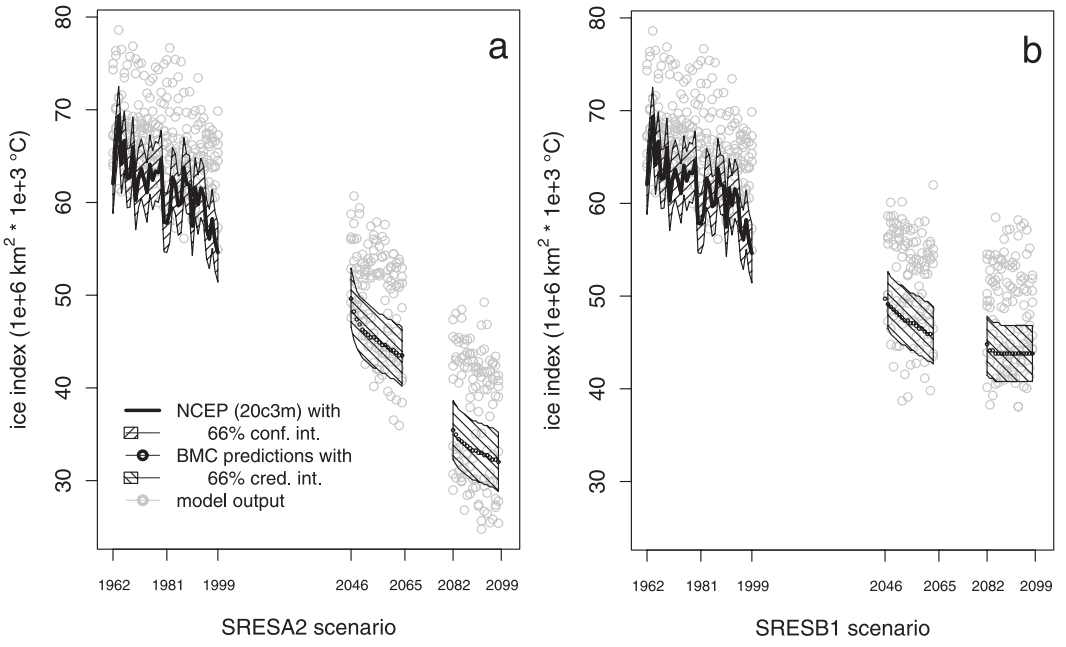

FIG. 6. Ice index generated from NCEP reanalysis data (black line, with 66% confidence interval) and BMC projections of the ice index (black dots, with 66% credibility interval) into future periods 2046–65 and 2082–99 for the (a) A2 and (b) B1 scenarios.

prediction period 2082–99 expected changes of 28.4% for the B1 scenario and 45.7% for the A2 scenario, which indicates nearly a halving of the ice volume. In Fig. 8b the evolution of the projected observations Y_0 are shown. Here the advantage of including a trend component becomes obvious (in case the GCM projections reflect the right trend behavior). The inclusion of an AR component (black dashed line) does not lead to deviations of the projections of a constant common mean (gray dashed line). However, the nonzero estimate

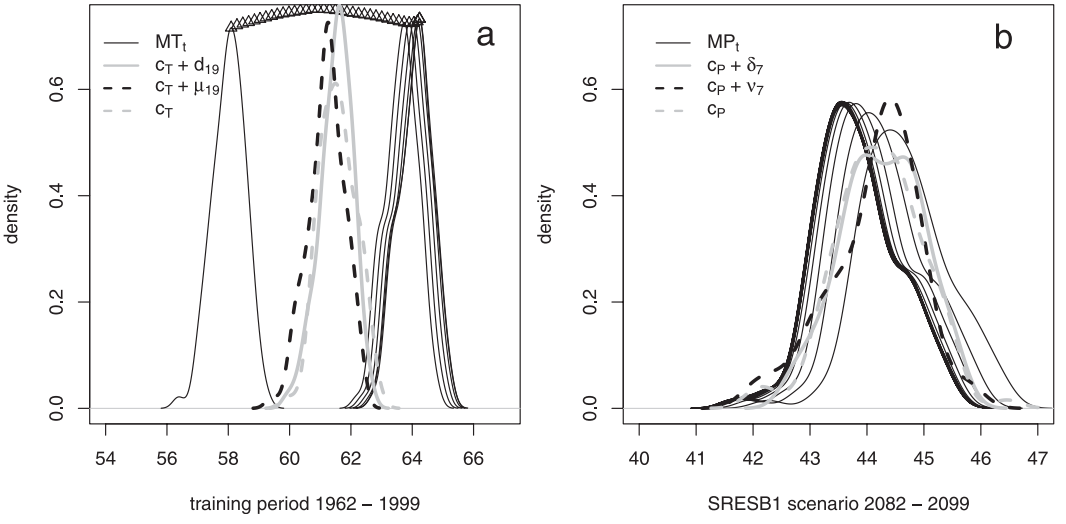

FIG. 7. Evolution of the distribution of the common means: (a) MT_t in the training period and (b) MP_t in the prediction period 2082–99 of the B1 (black lines) scenario. In addition, results for other state-space models are shown, including the common means composed of intercept and trend (gray lines), of intercept and AR component (black dashed lines), and of an intercept (gray dashed lines). For clarity some of the common mean distributions in the training period are indicated by black triangles and exemplarily only one distribution is depicted for the other nonstationary state-space models.

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

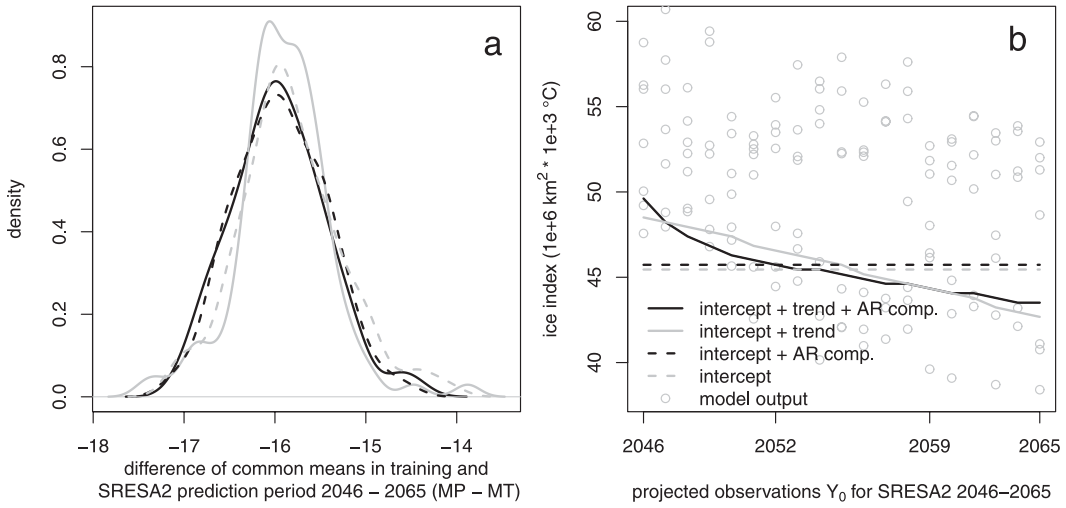

FIG. 8. (a) Expected average change of the ice index and (b) evolution of the projected observations Y_0 for years 2046–65 under the A2 scenario. Results for the different state-space models are depicted, i.e., common means comprising all three components (black lines), intercept and trend (gray lines), intercept and AR component (black dashed lines), and intercept (gray dashed lines).

of the AR parameter ϕ hints at the necessity of including an AR component. This component causes differences in the trend behavior (a comparison of black and gray lines), as discussed earlier.

In case constant common means are assumed, the approach of Tebaldi et al. (2005) is recovered (results are depicted as dashed gray lines in Figs. 7, 8). Thus, the main difference to this approach consists of nonstationary projections. Furthermore, we do not explicitly introduce any dependence between GCM outputs in the training and prediction periods.

To investigate the robustness of our results with respect to the different training periods, we compare the projections of training periods 1962–79 (T1), 1971–89 (T2), and 1981–99 (T3) to the period 1946–65 for the A2 scenario. In Fig. 9a the mean of the GCM biases for the different training periods and their average are shown. The dispersal of the bias among the models is similar for all the training periods evaluated, and the averaged biases are very close together. The training period 1981–99 has the highest averaged model bias, which may be related to the fact that IPCC GCM runs do not reproduce well the decrease of sea ice extent within this period (Stroeve et al. 2007). The average precision of all models is on par for nearly all training periods; it is slightly worse for 1971–89 (results not shown). However, these differences do not strongly influence the projection results: Figure 9b exemplarily shows the distribution of the common mean for year 2054. In our model we assume the same bias for the training and projection periods. The bias correction is largest for the training period 1981–99; therefore, those projections lie below the other projections. The width of the density of the common mean is largest for the training period 1962–79, which hints at a larger spread of the model outputs during that period. The use of a longer training period (gray lines) does not lead to a coarser density of the common mean estimate. Nevertheless, a longer training period allows for a better estimation of the constant parameters such as model biases, which are meant to represent average values. In Fig. 9c the evolution of the common means are depicted. We find comparable trend tendencies within the projection period for all the training periods assessed. All in all the projection results show to be robust. This is due to the little differences in model bias and precision for the training periods chosen.

In Fig. 10 the BMC projections are depicted in comparison to the total global cumulative CO_2 emissions (cf. Nakicenovic and Swart 2000). A direct link between global emission amounts and the ice index, which represents dynamics on the comparatively small oceanic Arctic region, might not exist. However, for the first projection period, the emission paths for both scenarios overlap significantly, whereas for the second prediction period already a clear separation appears visible. This fits well with the evolution of the ice index for both scenarios. Furthermore, the potential stabilization of the ice accumulation decrease for the B1 scenario and the second prediction period is accompanied by a decelerated increase of the cumulative CO_2 emissions: In the B1 scenario an actual reduction of the global annual CO_2 emissions is achieved from 2050 on, whereas this annual contribution never ceases to increase for the A2 scenario. This seems to have an effect on the SAT

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

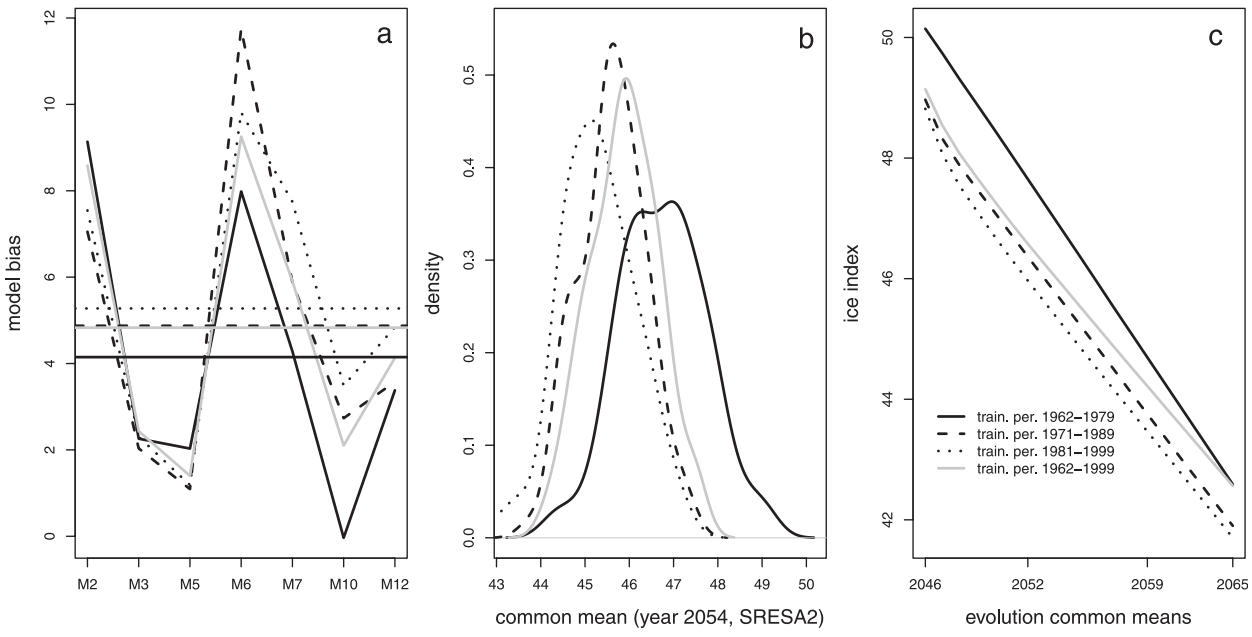

FIG. 9. Comparison of projections to period 2046–65 of the A2 scenario with training periods 1962–79 (black lines), 1971–89 (black dashed lines), 1981–99 (black dotted lines), and with the entire historical training period (gray lines). Shown are (a) model biases and the average model bias, (b) the density of the common mean at year 2054, and (c) the evolution over the common means over time.

and therefore on the ice accumulation over the oceanic Arctic.

5. Conclusions

In this paper we present a Bayesian method to enhance projections. Information from a multimodel ensemble is combined within a statistical framework. The parameters of the statistical model are estimated by regarding observations and multimodel outputs as random variables, which float (with a potential bias) around a common mean. This allows assess to model-specific deficiencies, namely, variability and bias. It is advantageous that all parameters are estimated together, which reduces estimation errors. Although we start from diffuse priors, informative posterior distributions are derived for all parameters. The methodology is applied to an ice index representing the ice accumulation over the oceanic Arctic region during cold seasons. Under the A2 scenario, we find a continuing decrease of the ice index, whereas for the B1 scenario, stabilization appears visible by the end of the twenty-first century. The stabilization hints at the retention of a minimum of ice rebuilding capacity in the freezing season for this scenario, which is important for questions related to adaptation and resilience (cf. Chapin et al. 2006; Laidler 2006).

Information from the entire training and prediction period is utilized for the BMC projections, whereas for simple averaging, only GCM outputs at one time step are relevant. Furthermore, here the models are not combined additively. In this way we do not have to adjust the length of the training period to the amount of information available. The Bayesian combination of multimodel ensembles were shown to potentially ameliorate projection skills in comparison to single-model projections or to the average of ensemble projections. A possible extension is the integration of expert knowledge on the GCM outputs by means of the priors.

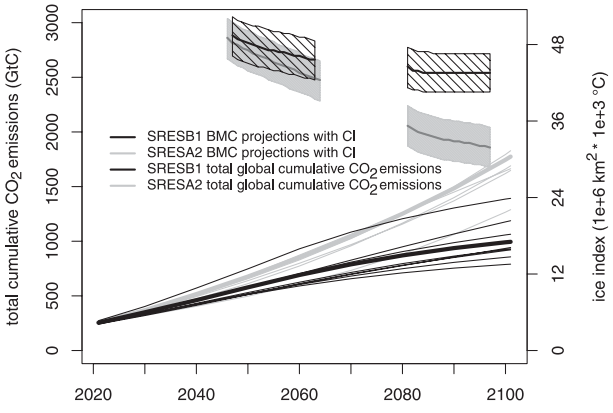

FIG. 10. BMC projections of the ice index into future periods 2046–65 and 2082–99 for the A2 (dark gray line and gray-shaded credibility intervals) and B1 (black line and black-shaded credibility intervals) scenarios with the axis on the right-hand side. For comparison total global cumulative CO_2 emissions for the A2 (gray lines) and B1 (black lines) scenario groups are shown [Source: data tables, appendix VII of Nakicenovic and Swart (2000)].

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

Acknowledgments. This work has been realized within the EU Marie Curie research network Network for Ice Sheet and Climate Evolution (NICE, available online at nice.ipsl.jussieu.fr; Contract 036127). In addition, Philippe Naveau's work has been supported by the EU-FP7 ACQWA project (www.acqwa.ch; Contract 212250), by the PEPER-GIS project (www.gisclimat.fr/projet/peper), and by the ANR-MOPERA project. We wish to thank both projects for their financial support. Furthermore, we are grateful to Michel Kolasinski for the preprocessing of the data and for helpful comments. The technique to create the ice index, which has been evaluated in this study, has been developed in the context of the EU project DAMOCLES (available online at www.damocles-eu.org), and we especially thank J.-C. Gascard for its provision.

APPENDIX

MCMC Routine for Model with Time-Dependent Parameters

A MCMC routine is used to generate pseudorandom drawings from the posterior distribution of the state-space components $\mathbf{\Theta}$ (cf. Harvey et al. 2007). To obtain the joint posterior $[\mathbf{\Theta}|\mathbf{D}]$, we divide a Gibb's sampling procedure in several blocks: First, the state-space parameters $\mu_t^{(i)}$, $d_t^{(i)}$, $\nu_t^{(i)}$, and $\delta_t^{(i)}$ are obtained by Kalman filtering and smoothing; then, marginal conditional likelihoods of the stationary parameters are evaluated. Next, for the parameter ϕ, a Metropolis–Hastings step is integrated in the Gibb's procedure. The algorithm converges to the joint posterior distribution of the parameters (cf. Chib and Greenberg 1996).

The vector of initial values for the stationary parameters $\mathbf{\Theta}_{\text{stat}}^{(0)} = (c_T^{(0)}, c_P^{(0)}, \gamma^{(0)}, \phi^{(0)}, k_1^{(0)}, k_2^{(0)}, \lambda_1^{(0)}, \ldots, \lambda_M^{(0)}, b_1^{(0)}, \ldots, b_M^{(0)}, \lambda_\mu^{(0)}, \lambda_\nu^{(0)}, \lambda_d^{(0)}, \lambda_\delta^{(0)})$ is provided. Then, the MCMC routine is run to produce a sequence of draws $\mathbf{\Theta}^{(i)}$, $i = 1, \ldots, R$. We utilize a burn in the period of 5000 steps, set $R = 100$, and keep the parameter values only every 50th iteration to avoid correlations between the $\mathbf{\Theta}^{(i-1)}$ and $\mathbf{\Theta}^{(i)}$. In each iteration step, we proceed as follows:

1) To draw $\mu^{(i)}$, $\nu^{(i)}$, $d^{(i)}$, and $\delta^{(i)}$ out of $[\mu^{(i)}, d^{(i)}, \nu^{(i)}, \delta^{(i)}|\mathbf{\Theta}_{\text{stat}}, \mathbf{D}]$ given the most recent iterate of $\mathbf{\Theta}_{\text{stat}}^{(i-1)}$, a Kalman smoother proposed by Dethlefsen and Lundbye-Christensen (2005) is employed (see also Durbin and Koopman 2002). We restrict the trend components to have mean 0; a potential deviation from this is captured by the intercepts c_T and c_P.
2) We sample $\mathbf{\Theta}_{\text{stat}}$ in block from $[\mathbf{\Theta}_{\text{stat}}^{(i)}|\mu^{(i)}, d^{(i)}, \nu^{(i)}, \delta^{(i)}, \mathbf{D}]$ by using the Gibb's sampler. The constant parameters have independent priors and their posterior distribution is given by Eq. (13). We only use information of the training period to estimate the precision of the models λ_i and their bias b_i. Therefore, we obtain

$$\lambda_i : U(0,c) \Rightarrow G\left\{\frac{T+2}{2}, \frac{\sum_{t-1}^{T}[X_{i_t} - (c_T + \mu_t + d_t + b_i)]^2}{2}\right\}$$

and

$$b_i : U(-\infty,\infty) \Rightarrow N\left\{\frac{\sum_{t=1}^{T}[X_{i_t} - (c_T + \mu_t + d_t)]}{T}, (\lambda_i T)^{-1}\right\}. \tag{A1}$$

The marginal posterior distribution of the change in precision γ is given by

$$\gamma : U(0,c) \Rightarrow G\left\{\frac{\text{MP}+2}{2}, \frac{\sum_{i=1}^{M}\sum_{t=G+1}^{G+P}\lambda_i[Y_{i_t} - (c_P + \nu_t + \delta_t + b_i)]^2}{2}\right\}. \tag{A2}$$

For the precisions of the state-space model,

$$\lambda_\mu : U(0,c) \Rightarrow G\left[\frac{T+1}{2}, \frac{\sum_{t=2}^{T}(\mu_t - \phi\mu_{t-1})^2}{2}\right],$$

$$\lambda_d : U(0,c) \Rightarrow G\left[\frac{T+1}{2}, \frac{\sum_{t=2}^{T}(d_t - d_{t-1} - k_1)^2}{2}\right],$$

$$\lambda_\nu : U(0,c) \Rightarrow G\left[\frac{P+1}{2}, \frac{\sum_{t=G+2}^{G+P}(\nu_t - \phi\nu_{t-1})^2}{2}\right], \quad \text{and}$$

$$\lambda_\delta : U(0,c) \Rightarrow G\left[\frac{P+1}{2}, \frac{\sum_{t=G+2}^{G+P}(\delta_t - \delta_{t-1} - k_2)^2}{2}\right] \tag{A3}$$

holds, and the posterior marginal distributions of k_1 and k_2 are given by

$$k_1 : U(-\infty,\infty) \Rightarrow N\left\{\frac{\sum_{t=2}^{T}(d_t - d_{t-1})}{T-1}, [(T-1)\lambda_d]^{-1}\right\}$$

and

$$k_2 : U(-\infty,\infty) \Rightarrow N\left\{\frac{\sum_{t=G+2}^{G+P}(\delta_t - \delta_{t-1})}{P-1}, [(P-1)\lambda_\delta]^{-1}\right\}. \tag{A4}$$

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

3) The autoregressive parameter ϕ is drawn from

$$[\phi|k_1,k_2,b_1,\ldots,b_M,c_T,c_P,\gamma,\lambda_1,\ldots,\lambda_M,\lambda_\mu,\lambda_d,\lambda_\nu,\lambda_\delta,\mu,d,\nu,\delta,\mathbf{D}] \propto [\mathbf{D}|\boldsymbol{\Theta}][\mu,d,\nu,\delta|\boldsymbol{\Theta}_{\text{stat}}][\phi] \propto [\mu,d,\nu,\delta|\boldsymbol{\Theta}_{\text{stat}}][\phi]. \quad \text{(A5)}$$

As posterior marginal distribution, we obtain

$$\phi{:}U(-1,1) \Rightarrow N\left(\frac{\lambda_\mu\sum_{t=2}^{T}\mu_t\mu_{t-1}+\lambda_\nu\sum_{t=G+2}^{G+P}\nu_t\nu_{t-1}}{\lambda_\mu\sum_{t=2}^{T}\mu_{t-1}^2+\lambda_\nu\sum_{t=G+2}^{G+P}\nu_{t-1}^2}, \left(\lambda_\mu\sum_{t=2}^{T}\mu_{t-1}^2+\lambda_\nu\sum_{t=G+2}^{G+P}\nu_{t-1}^2\right)^{-1}\right), \quad \text{(A6)}$$

whereas ϕ is sampled from a truncated normal. The state-space variables depend on ϕ, and the relation is more complex than for the other parameters. Therefore, we include a Metropolis–Hastings step to sample ϕ: a proposal value ϕ^* is obtained from a symmetric proposal distribution, and Eq. (A6) is evaluated for $\phi^{(i-1)}$ and ϕ^*. If $[\phi^*|\cdot] > [\phi^{(i-1)}|\cdot]$, the proposal ϕ^* is accepted, that is, $\phi^{(i)} = \phi^*$; otherwise, ϕ^* is accepted with probability $[\phi^*|\cdot]/[\phi^{(i-1)}|\cdot]$.

REFERENCES

Anderson, D. L., 1961: Growth rate of sea ice. *J. Glaciol.,* **3,** 1170–1172.

Baettig, M. B., M. Wild, and D. M. Imboden, 2007: A climate change index: Where climate change may be most prominent in the 21st century. *Geophys. Res. Lett.,* **34,** L01705, doi:10.1029/2006GL028159.

Berliner, L. M., and Y. Kim, 2008: Bayesian design and analysis for superensemble-based climate forecasting. *J. Climate,* **21,** 1891–1910.

Bloomfield, P., 1992: Trends in global temperature. *Climatic Change,* **21,** 1–16.

Bromwich, D. H., R. L. Fogt, K. I. Hodges, and J. E. Walsh, 2007: A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions. *J. Geophys. Res.,* **112,** D10111, doi:10.1029/2006JD007859.

Cavalieri, D. J., and C. L. Parkinson, 2008: Antarctic sea ice variability and trends, 1979–2006. *J. Geophys. Res.,* **113,** C07004, doi:10.1029/2007JC004564.

Chapin, F. S., III, and Coauthors, 2006: Building resilience and adaptation to manage Arctic change. *Ambio,* **35,** 198–202.

Charbit, S., D. Paillard, and G. Ramstein, 2008: Amount of CO_2 emissions irreversibly leading to the total melting of Greenland. *Geophys. Res. Lett.,* **35,** L12503, doi:10.1029/2008GL033472.

Chib, S., and E. Greenberg, 1996: Markov chain Monte Carlo simulation methods in econometrics. *Econ. Theory,* **12,** 409–431.

Coelho, C. A. S., S. Pezzulli, M. Balmaseda, F. J. Doblas-Reyes, and D. B. Stephenson, 2004: Forecast calibration and combination: A simple Bayesian approach for ENSO. *J. Climate,* **17,** 1504–1516.

Cohn, T. A., and H. F. Lins, 2005: Nature's style: Naturally trendy. *Geophys. Res. Lett.,* **32,** L23402, doi:10.1029/2005GL024476.

Congdon, P., 2003: *Applied Bayesian Modelling*. John Wiley & Sons, 457 pp.

Davison, A. C., 2003: *Statistical Models*. Cambridge University Press, 726 pp.

De Jong, P., and N. Shephard, 1995: The simulation smoother for time series models. *Biometrika,* **82,** 339–350.

DelSole, T., 2007: A Bayesian framework for multimodel regression. *J. Climate,* **20,** 2810–2826.

Dethlefsen, C., and S. Lundbye-Christensen, 2005: Formulating state space models in R with focus on longitudinal regression models. Aalborg University Research Rep. Series R-2005-21, 12 pp.

Döscher, R., K. Wyser, H. E. M. Meier, M. Qian, and R. Redler, 2009: Quantifying Arctic contributions to climate predictability in a regional coupled ocean-ice-atmosphere model. *Climate Dyn.,* 1157–1176, doi:10.1007/s00382-009-0567-y.

Durbin, J., and S. J. Koopman, 2002: A simple and efficient simulation smoother for state space time series analysis. *Biometrika,* **89,** 603–616, doi:10.1093/biomet/89.3.603.

Gelman, A., 2006: Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.,* **1,** 515–534.

——, J. B. Carlin, H. S. Stern, and D. B. Rubin, Eds., 2004: *Bayesian Data Analysis*. 2nd ed. Chapman & Hall/CRC, 668 pp.

Gneiting, T., A. E. Raftery, A. H. Westveld III, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.,* **133,** 1098–1118.

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.

Harvey, A. C., T. M. Trimbur, and H. K. Van Dijk, 2007: Trends and cycles in economic time series: A Bayesian approach. *J. Econ.,* **140,** 618–649, doi:10.1016/j.jeconom.2006.07.006.

Hassol, S. J., 2004: *Impacts of a Warming Arctic: Arctic Climate Impact Assessment*. Cambridge University Press, 139 pp.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting,* **15,** 559–570.

Johannessen, O. M., and Coauthors, 2004: Arctic climate change: Observed and modelled temperature and sea-ice variability. *Tellus,* **56A,** 328–341.

Kallache, M., H. W. Rust, and J. Kropp, 2005: Trend assessment: Applications for hydrology and climate. *Nonlinear Processes Geophys.,* **12,** 201–210.

Kaufman, D. S., and Coauthors, 2009: Recent warming reverses long-term Arctic cooling. *Science,* **325,** 1236–1239, doi:10.1126/science.1173983.

Koop, G., and H. K. Van Dijk, 2000: Testing for integration using evolving trend and seasonals models: A Bayesian approach. *J. Econ.,* **97,** 261–291.

Kwok, R., and D. A. Rothrock, 2009: Decline in Arctic sea ice thickness from submarine and ICESat records: 1958–2008. *Geophys. Res. Lett.,* **36,** L15501, doi:10.1029/2009GL039035.

Laidler, G. J., 2006: Inuit and scientific perspectives on the relationship between sea ice and climate change: The ideal

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC

complement? *Climatic Change,* **78,** 407–444, doi:10.1007/s10584-006-9064-z.

Laxon, S., N. Peacock, and D. Smith, 2003: High interannual variability of sea ice thickness in the Arctic region. *Nature,* **425,** 947–950, doi:10.1038/nature02050.

Maksimovich, E., and J. C. Gascard, cited 2010: Winter near-surface warming in NCEP/NCAR reanalysis I. Reanalysis error or arctic warming? Earlier melt onset and later freeze-up? [Available online at http://www.damocles-eu.org/artman2/uploads/1/poster_Sopot_task_5.2.1.pdf.]

Maykut, G. A., 1986: The surface heat and mass balance. *The Geophysics of Sea Ice,* N. Untersteiner, Ed., NATO ASI Series B, Vol. 146, Plenum Press, 395–463.

Meinhold, R. J., and N. D. Singpurwalla, 1983: Understanding the Kalman filter. *Amer. Stat.,* **37,** 123–127.

Min, S.-K., and A. Hense, 2006: A Bayesian assessment of climate change using multimodel ensembles. Part I: Global mean surface temperature. *J. Climate,* **19,** 3237–3256.

——, ——, H. Paeth, and W.-T. Kwon, 2004: A Bayesian decision method for climate change signal analysis. *Meteor. Z.,* **13,** 421–436.

Nakicenovic, N., and R. Swart, Eds., 2000: *Special Report on Emissions Scenarios*. Cambridge University Press, 599 pp.

Parry, M. L., O. F. Canziani, J. P. Palutikof, P. J. van der Linden, and C. E. Hanson, Eds., 2007: *Climate Change 2007: Impacts, Adaptation and Vulnerability*. Cambridge University Press, 976 pp.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.,* **133,** 1155–1174.

Rougier, J., 2007: Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change,* **81,** 247–264, doi:10.1007/s10584-006-9156-9.

Ruzmaikin, A., and J. Feynman, 2002: Solar influence on a major mode of atmospheric variability. *J. Geophys. Res.,* **107,** 4209, doi:10.1029/2001JD001239.

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.

Tebaldi, C., R. L. Smith, D. Nychka, and L. O. Mearns, 2005: Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles. *J. Climate,* **18,** 1524–1540.

Tol, R. S. J., and A. F. De Vos, 1998: A Bayesian statistical analysis of the enhanced greenhouse effect. *Climatic Change,* **38,** 87–112.

van Loon, M., and Coauthors, 2007: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble. *Atmos. Environ.,* **41,** 2083–2097.

West, M., 1997: Time series decomposition. *Biometrika,* **84,** 489–494.

Zhang, X., and J. E. Walsh, 2006: Toward a seasonally ice-covered Arctic ocean: Scenarios from the IPCC AR4 model simulations. *J. Climate,* **19,** 1730–1747.

Unauthenticated | Downloaded 04/17/21 05:55 AM UTC