Geophysical Research Letters®

Key Points:

- A Linear Inverse Model is evaluated using last millennium model simulations for Arctic climate prediction
- The Linear Inverse Model successfully predicts Arctic conditions when the same model simulation is used for training and validation
- Linear Inverse Model forecast skill is proportional to the spatial correlation of variance in the validation and training data

Monthly Arctic Sea-Ice Prediction With a Linear Inverse Model

M. Kathleen Brennan[1] [ID]**, Gregory J. Hakim[1], and Edward Blanchard-Wrigglesworth[1]** [ID]

[1]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA

Abstract We evaluate Linear Inverse Models (LIMs) trained on last millennium model data to predict Arctic sea-ice concentration, thickness, and other atmospheric and oceanic variables on monthly timescales. We find that more than 500 years of training data and 100 years of validation data are needed to reliably estimate LIM forecast skill. The best LIM has skill up to 8 months lead time and outperforms an autoregressive model of order one (AR1) forecast at all locations, with particularly large outperformance near the ice edge. However, for out-of-sample validation tests using data from various different model simulations and reanalysis products, they underperform an AR1 model due to differences in the location of the sea-ice edge from the training data. We present a metric for predicting LIM forecast skill, based on the spatial correlation of the variance in the training and validation data sets.

Plain Language Summary Rapid changes in sea-ice concentration in recent decades have produced new navigational challenges and hazards in the region, elevating the importance of seasonal sea-ice forecasts. Arctic sea ice has been shown to contain inherent predictability on seasonal timescales, yet current predictions generally show poor skill. Here, we employ a statistical technique referred to as Linear Inverse Modeling, which uses linearized dynamical modes estimated from a training data set to predict sea-ice conditions on monthly timescales. We find a Linear Inverse Model is able to outperform a baseline statistical model throughout the Arctic when initialized on the data derived from the same model.

1. Introduction

Large climatic changes observed in the Arctic in recent decades have made prediction of Arctic sea ice on monthly to seasonal timescales increasingly relevant (Jung et al., 2016). Younger, thinner ice in the Arctic (e.g., R. Lindsay & Schweiger, 2015; Maslanik et al., 2007) has allowed for new shipping routes (e.g., Smith & Stephenson, 2013) and a growing prevalence of industrial and tourist (Hall & Saarinen, 2010) activities in the region. Furthermore, fast changing sea-ice conditions present hazards for local communities as conditions become increasingly challenging to anticipate and predict (Eicken, 2013).

Arctic sea ice has been shown to contain inherent predictability with persistence of sea-ice concentration (SIC) up to 5 months (Blanchard-Wrigglesworth, Armour, et al., 2011; Lemke et al., 1980; R. W. Lindsay et al., 2008), and strong area–thickness coupling provides predictability for even longer timescales (Blanchard-Wrigglesworth, Bitz, et al., 2011). Blanchard-Wrigglesworth and Bushuk (2019) find that predictability in models is likely robust within a constant climate mean state, but these relationships might not be stationary in a warming climate (Bonan & Blanchard-Wrigglesworth, 2020; Holland & Stroeve, 2011; Holland et al., 2019). Beyond persistence, other major contributors to Arctic sea-ice predictability include dynamical advection of sea-ice anomalies by mean Arctic circulation patterns (Guemas et al., 2016), atmospheric temperature variability (Olonscheck et al., 2019), and ocean heat flux (Bitz et al., 2005; Bushuk et al., 2019; Yeager et al., 2015).

There has been a growing effort to predict Arctic sea ice on subseasonal to seasonal timescales, and the Sea Ice Outlook (SIO, Stroeve et al., 2014) represents a substantial sea-ice-research community effort to develop and improve Arctic sea-ice prediction. Starting in 2008, the SIO has accepted forecasts of September sea-ice conditions from groups around the world, and though much progress has been made, there remains considerable room for improvement. Stroeve et al. (2014) evaluated the skill of SIO forecasts from 2008 to 2013 and found that, overall, the skill, regardless of method, was generally poor when sea-ice conditions depart from the long-term trend. Moreover, statistical models tend to outperform dynamical models (Stroeve et al., 2014) likely due in part to uncertainty in model physics (Blanchard-Wrigglesworth et al., 2015) and initial conditions (Blanchard-Wrigglesworth et al., 2017).

Most statistical, data intensive, approaches for predicting Arctic sea ice involve predicting pan-Arctic or regional quantities using linear regression techniques (e.g., Drobot et al., 2006; R. W. Lindsay et al., 2008; Petty et al., 2017; Tivy et al., 2007). R. W. Lindsay et al. (2008) combine observed atmospheric indices and sea-ice coverage with ocean and ice thickness fields derived from Pan-Arctic Ice-Ocean Modeling and Assimilation System (PIOMAS) in order to develop an empirical linear model. They find substantially more skill than previous linear regression approaches that include only SIC and atmospheric variables (Drobot et al., 2006). More recently, other data-driven techniques have been applied to Arctic sea-ice prediction on seasonal timescales. For example, Andersson et al. (2021) use convolutional neural networks trained on both climate model and observational data to predict 6 months of SIC fields that are able to outperform a dynamical model for seasonal forecasts. Hogg et al. (2020) apply a Koopman Mode Decomposition to satellite observations of SIC to make future predictions of SIC in both hemispheres.

Here, we employ a Linear Inverse Model (LIM) to predict Arctic sea ice on monthly timescales. A LIM is ideal for this application as it uses linearized dynamical modes estimated from a training data set, harnessing the main contributors to Arctic predictability. The LIM framework assumes stationary statistics and that fluctuations about a mean state can be modeled by linear dynamics plus stochastic noise. Yuan et al. (2016) used a linear Markov model trained on sea-ice, oceanic, and atmospheric variables from reanalysis data to predict Arctic SIC on monthly timescales, similar to the LIM approach here; however, they did not detrend the data and only use 35 years (1979–2013) for training and validation. LIMs have become a useful tool for predicting and probing the dynamics of the tropical atmospheric and oceanic variables on seasonal timescales (e.g., Alexander et al., 2008; Cavanaugh et al., 2015; Dias et al., 2019; Henderson et al., 2020; Huddart et al., 2017; Newman et al., 2009; Penland, 1996; Penland & Matrosova, 1998; Penland & Sardeshmukh, 1995; Shin et al., 2021; Winkler et al., 2001). Perkins and Hakim (2020) built a multivariate LIM for forecasting global climate states on annual timescales. We build on this previous work with a focus on predicting Arctic sea-ice and other climate fields on monthly timescales.

We focus on answering two main questions: (a) over what time period is a LIM useful for predicting Arctic sea-ice and climate conditions on monthly timescales? (b) What conditions are required for the LIM to make skillful predictions? To address the first question, in Section 3.1, we use last millennium simulations to train a LIM and assess the sample size needed for robust estimates of LIM skill. Given the absence of strong external forcing during the last millennium, and large availability of data, we are able to optimize the parameters of the LIM for predicting Arctic sea-ice coverage and assess its skill. We then address the second question in Section 3.2 by initializing the LIM with data from different model simulations and reanalysis data and find that the LIM fails to beat an autoregressive model of order one (AR1) forecast due to differences in the sea ice from the training data. We present a metric for predicting when the stationary statistics assumed for the LIM are not satisfied.

2. Methods

2.1. Linear Inverse Modeling

An LIM is an empirically determined linearization of a statistically stationary dynamical system about its mean state (Penland, 1996; Penland & Matrosova, 1994; Penland & Sardeshmukh, 1995). The tendency of a multivariate state vector $\mathbf{x}$ can be represented as

$$\frac{d\mathbf{x}}{dt} = \mathbf{L}\mathbf{x} + \zeta \tag{1}$$

(Equation 2 in Penland & Sardeshmukh, 1995). Here, $\mathbf{L}$ is the deterministic dynamical operator that propagates the state in time and ζ represents the unpredictable dynamics as uncorrelated white noise forcing in time with state–space correlations. Integrating Equation 1 with respect to time gives the forecast from time t to $t + \tau$,

$$\mathbf{x}(t + \tau) = \exp(\mathbf{L}\tau)\mathbf{x}(t) + \mathbf{n}. \tag{2}$$

Here, $\mathbf{n}$ is a random error vector from the integration of the white noise (ζ). Since $\mathbf{L}$ is a matrix, it has an eigenvalue decomposition such that $\mathbf{L}\mathbf{u}_m = \lambda_m\mathbf{u}_m$, where $\mathbf{u}_m$ are the eigenmodes of $\mathbf{L}$ and λ_m are the corresponding eigenvalues. Stationary statistics require that the eigenmodes of $\mathbf{L}$ ($\mathbf{u}_m$) are all damped (eigenvalues have negative real parts). Since the eigenmodes are not orthogonal, interference results in transient growth despite the decay of each eigenmode (e.g., Farrell, 1982).

We use standard inverse methods to determine $\mathbf{L}$ empirically as described in Penland (1989) by defining a matrix $\mathbf{G}_\tau = \exp(\mathbf{L}\tau)$, which is determined through τ-lag covariances,

$$\mathbf{G}_\tau = \mathbf{C}_\tau \mathbf{C}_o^{-1}. \tag{3}$$

Here, $\mathbf{C}_\tau = \langle \mathbf{x}(t + \tau)\mathbf{x}^T(t) \rangle$ represents the sample covariance of $\mathbf{x}$ for a time lag of τ and $\mathbf{C}_o = \langle \mathbf{x}(0)\mathbf{x}^T(0) \rangle$. Assuming that the state and the error are uncorrelated, we can use Equation 2 to propagate the state covariance:

$$\text{cov}(\mathbf{x}(t + \tau), \mathbf{x}(t + \tau)) = \mathbf{G}_\tau \text{cov}(\mathbf{x}(t), \mathbf{x}(t))\mathbf{G}_\tau^T + \mathbf{N}_\tau. \tag{4}$$

From Equations 3 and 4 we also have

$$\mathbf{N}_\tau = \mathbf{C}_o - \mathbf{G}_\tau \mathbf{C}_o \mathbf{G}_\tau^T. \tag{5}$$

Given $\mathbf{G}_\tau$ and $\mathbf{N}_\tau$, we solve for the state $\mathbf{x}$ at time τ using Equation 2 and the covariance using Equation 4. Note that when validating the LIM mean forecast, we take the expectation of Equation 2, and the noise term vanishes.

2.2. LIM Training Procedure

As mentioned in Section 2.1, the LIM operator $\mathbf{L}$ is assumed to be stationary about a mean value, and damped (it has negative eigenvalues), so we linearly detrend the training data and remove the climatological mean for each variable. Given that the trend in time is different across the seasonal cycle (particularly for Arctic sea-ice variables, e.g., Serreze et al., 2007), we remove the linear trend and mean for each month individually at each grid point. We limit the domain to north of 40°N in order to optimize for Arctic prediction.

Given the large number of degrees of freedom in the state vector, we truncate the state using an area-weighted empirical orthogonal function (EOF) decomposition before calculating the lagged covariance. The leading 50 EOFs for each variable are retained and the resulting state projection normalized by the square-root of the total temporal variance such that the sum of the variance over the truncated state for each variable sums to 1. Once truncated and normalized, all LIM variables are stacked in a matrix:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_{TAS} \\ \mathbf{x}_{PSL} \\ \cdots \\ \mathbf{x}_{SIC} \end{bmatrix}. \tag{6}$$

With this truncated state, we determine $\mathbf{G}_\tau$ using Equation 4 for $\tau = 1$ month.

Arctic sea-ice coverage exhibits a large seasonal cycle; thus, we anticipate the need to potentially build a LIM for different months separately. To investigate this, we train LIMs for single-month transitions (e.g., forecasting only February from January) as well as LIMs trained on all months (with seasonal cycle and trends removed). We find that single-month LIMs are prone to have positive eigenvalues for $\mathbf{L}$, indicating that the stationary assumptions of the LIM are not met. We hypothesize that this is due to a change in the location of variability for sea ice as the location of the sea-ice edge changes; that is, the statistics are not stationary. In order for the LIM to produce anomalies in a future month for locations where there is not much variability in the initial month, positive eigenvalues result. As a result, we use the all-month LIMs for the rest of the paper.

We perform both *in-* and *out-of-sample* validation, which refer to the time period used for validation relative to the training period. For in-sample validation, the time period used for validation is also used to train the LIM, and for out-of-sample validation, the validation time period is not used in training. We will use *intramodel* validation to indicate when data originating from the same model run is used for both training and validation of the LIM and *cross-model* validation to indicate when the LIM is trained and initialized using data originating from different model simulations and reanalysis data.

2.3. Data Sources

For training the LIM, we use monthly averaged data from the Community Earth System Model version 1 (CESM1) Last Millennium Ensemble (LME, Otto-Bliesner et al., 2016). We train and validate our LIM on a single ensemble

member spanning years 850–1850 CE. Model fields used to test their affect on LIM forecast skill are: 2-m air temperature (TAS), sea level pressure (PSL), 500 hPa geopotential height (ZG500), sea surface temperature (SST), sea-ice thickness (SIT), and SIC. These variables were selected because they have all been associated with sea-ice variability, and we hypothesize that including each of these variables may contribute positively to the LIM's ability to predict SIC. Specifically, Arctic sea ice has been shown to exhibit strong coupling with surface air temperature, particularly on longer timescales (e.g., Armour et al., 2011; Gregory et al., 2002; Mahlstein & Knutti, 2012; Olonscheck et al., 2019). PSL is associated with surface winds that can drive sea-ice motion and variability (e.g., Rigor et al., 2002) and, similarly, ZG500 is associated with the large-scale dynamical state of the atmosphere which can also contribute to sea-ice variability. Furthermore, persistence in SIC is strongly influenced by both SSTs and SIT (e.g., Blanchard-Wrigglesworth, Armour, et al., 2011; Bushuk et al., 2015).

We also initialize and train the LIM using other models and data products simulating various time periods in order to test the sensitivity of the LIM to different mean states and covariance structures across variables. During the last millennium, we use monthly averaged data from both the Community Climate System Model version 4 (CCSM4, Landrum et al., 2013) and Max Planck Institute (MPI) for meteorology last millennium simulations (Jungclaus et al., 2012), which were run from 850 to 1850 CE. Both of these simulations are part of the Coupled Model Intercomparison Project, phase 5 (Taylor et al., 2012), Paleoclimate Modeling Intercomparison Project phase 3. During the historical period, we use monthly averaged data from the Coupled Model Intercomparison Project, phase 6 (CMIP6, Eyring et al., 2016) Community Earth System Model version 2 (CESM2, Danabasoglu et al., 2020), MPI Earth System Model version 1.2 (low resolution [Mauritsen et al., 2019]), and Geophysical Fluid Dynamics Laboratory Earth System Model version 4.1 (GFDL, Dunne et al., 2020) simulations, which were run from 1850 to 2014 CE. We also validate the LIM using the European Center for Medium-Range Weather Forecasts reanalysis product (ERA5, Hersbach et al., 2020), which spans 1979–2020 CE (ERA5 does not include a SIT variable). For simulations of the future, we use one ensemble member from the CESM1 Large Ensemble (Kay et al., 2015) which simulates 1920–2100 CE. All variables from all sources are regridded onto the native ocean and atmosphere grids from the CESM1 LME simulations.

2.4. Skill Metrics

To validate our predictions, we use the squared correlation coefficient (R^2 value), coefficient of efficiency (CE), and root mean squared error (RMSE). As defined below, the correlation coefficient (R) measures relative phasing of two time series.

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(v_i - \bar{v})}{\sqrt{\sum_i^n (x_i - \bar{x})^2}\sqrt{\sum_i^n (v_i - \bar{v})^2}}. \tag{7}$$

Here, v is the verification data and x is the state being evaluated (the forecasted value). The square of R describes the percentage of the variance in v that is linearly explained by x. The CE (Nash & Sutcliffe, 1970), like the correlation coefficient, not only measures the relative phasing of two data sets but also includes bias in the mean and variance:

$$CE = 1 - \frac{\sum_i^n (v_i - x_i)^2}{\sum_i^n (v_i - \bar{v})^2}. \tag{8}$$

CE has an upper bound of one if the time series are identical and is unbounded in the negative direction. A CE value of zero occurs when the sum of squared errors is equal to the variance in the verification data (a climatological forecast, where $x_i = \bar{v}$, would result in a CE of zero), and negative CE values result from a bias between the two data sets, either in the mean or amplitude of the variability.

To assess the pan-Arctic skill of SIC, which is reported as the percentage of a grid cell covered in sea ice, we use total Arctic RMSE. For forecasts of other quantities such as SIT, we use Arctic mean RMSE. To calculate the RMSE we use,

$$RMSE = \sqrt{\frac{\sum_i^n (x_i - v_i)^2}{n}}, \tag{9}$$

where i represents a given forecast, v is the verification data, and x is the forecast being evaluated. The RMSE is calculated for each grid cell. For SIC, the RMSE is multiplied by the grid cell area and the sum is taken across the ocean domain (north of 40°N, land not included), which we refer to as total Arctic RMSE. For other variables, the area-weighted Arctic mean RMSE is taken across the Arctic (north of 40°N), which we refer to at the Arctic mean RMSE. LIM forecasts are projected back from EOF space and compared with the verification data (not truncated) in latitude–longitude space.

3. Results

Given that the LIM assumes stationary statistics, we expect the LIM to produce skillful forecasts during the last millennium when there is less anthropogenic forcing than the Instrumental Era or future projections. We anticipate that the LIM will also perform well for intramodel validation given that the covariances in the LIM match those in the target state. In Section 3.1, we will use intramodel experiments to optimize the LIM parameters, and in Section 3.2, we will use cross-model validation to quantify under what time periods the LIM produces skillful forecasts. Forecasts from an AR1 model, trained on the EOF truncated data, serve as a baseline reference for comparison with the LIM forecasts. Here, an AR1 forecast (W) for a given month t is defined as $W_t = \alpha_1 W_{t-1}$, where α_1 is the 1 month lag correlation of the system.

3.1. LIM Optimization and Intramodel Performance

To start, we train and validate the LIM during the last millennium using intramodel validation with a CESM1 LME simulation, which provides ample data to test the sensitivity of the LIM to various parameters. In particular, we test the sensitivity of LIM performance to the following parameters: number of training and validation years, number of EOFs included in the truncation, and variables included in training.

To investigate the number of training and validation years necessary for converged LIM performance metrics, we perform two experiments. First, we fix the training period from 850 to 1650 CE (800 years) and vary the number of validation years from 10 to 200 years segments between 1651 and 1850 CE. All possible nonoverlapping validation segments of each length are used. Next, we fix the validation period from 1751 to 1850 CE (100 years) and vary the number of training years ranging from 100 to 900 years for all possible nonoverlapping segments between 850 and 1750 CE. For 1 month forecasts, the LIM total Arctic RMSE asymptotes at large sample size to 1.5×10^6 km² as compared to 1.6×10^6 km² for the AR1 forecasts (Figure S1 in Supporting Information S1). This indicates that more than 500 years of monthly training data and 100 years of validation data are needed to produce reliable skill metrics. For the remaining experiments, we will train the LIM between 850 and 1650 CE unless otherwise noted.

Next, we vary the number of EOFs included in the truncation during the training procedure (Section 2.4) from 5 to 250 for each variable and perform both in-sample (851–1050 CE) and out-of-sample (1651–1850 CE) validation. The percent of variance explained as a function of the number of EOFs retained for each of the six variables in the CESM1 LME simulation data is shown in Figure S2 in Supporting Information S1. For all experiments, and 10 or more EOFs for each variable, the LIM out-of-sample forecasts outperform an AR1 forecast for lead times up to around 5–6 months (not shown). While LIM forecast skill decreases monotonically in number of EOFs for the in-sample experiments, skill for the out-of-sample experiments levels off around 40–80 EOFs depending on lead time (Figure S3 in Supporting Information S1). This indicates overfitting for larger numbers of EOFs, and to avoid that we truncate all subsequent experiments to 50 EOFs per variable.

We now investigate how including different variables contribute to LIM predictability of Arctic SIC at different lead times. We consider the role of atmospheric (TAS, PSL, and ZG500), oceanic (SST), and SIT on SIC predictions. Nine different LIMs are trained using: only SIC, SIC plus TAS, SIC plus PSL, SIC plus ZG500, SIC plus SST, SIC plus SIT, SIC plus SST plus SIT, SIC plus TAS plus SST plus SIT, and a LIM trained using all six variables. For each LIM, we use 800 training years (851–1650 CE), 200 validation years (1651–1850 CE), and truncate to 50 EOFs per variable. Moreover, we evaluate LIM performance relative to a LIM trained on SIC alone (Figure 1). Most of the SIC forecast skill comes from SST and SIT variables (dark blue dashed line in Figure 1), as well as TAS (purple dashed line in Figure 1); however, all variables contribute favorably to forecast skill of SIC. TAS and SST contribute the most to SIC forecast skill on 1–3 months lead times, while SIT along with TAS and SST contribute the largest skill increase on 4–6 months lead times. PSL and ZG500
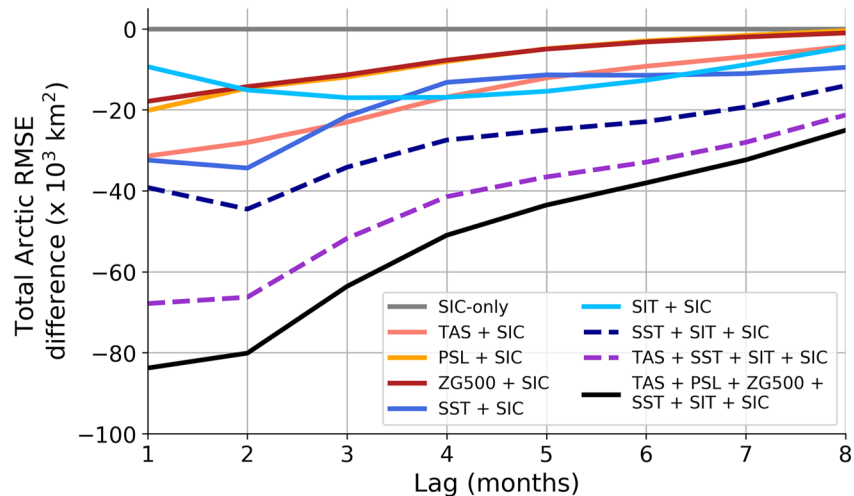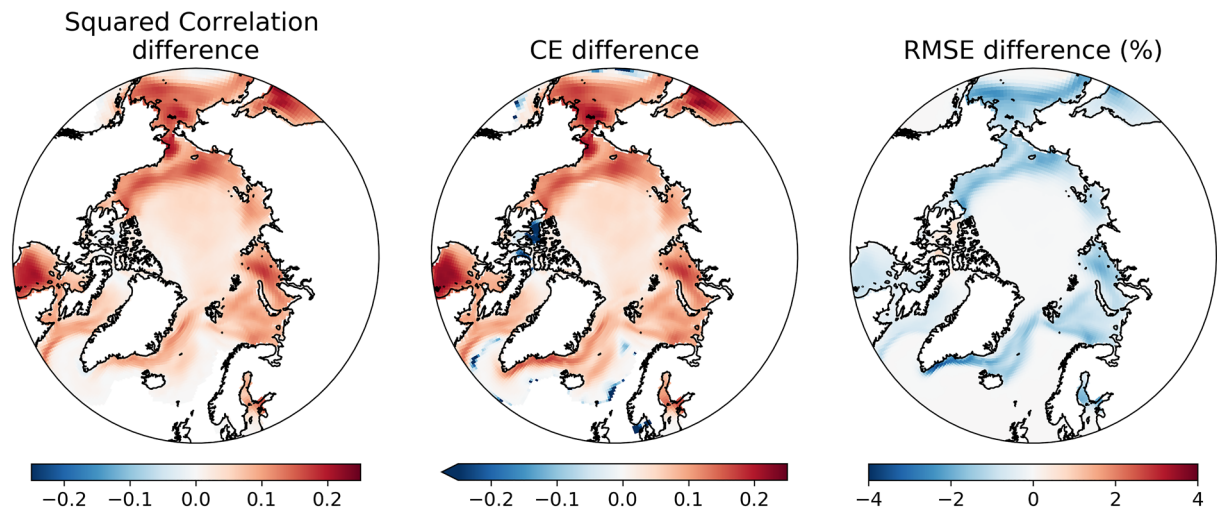
Figure 1. Forecast skill for Arctic sea-ice concentration (SIC) as a function of lead time for nine different Linear Inverse Models (LIMs) containing different combinations of variables: only SIC; 2-m air temperature (TAS), sea level pressure (PSL), 500 hPa geopotential height (ZG500), sea surface temperature (SST), and sea-ice thickness (SIT) each plus SIC; and all variables. Total Arctic root mean squared error (RMSE) is plotted relative to a LIM containing only SIC. Each LIM was trained using a CESM1 LME simulation from 850 to 1650 CE and validated between 1651 and 1850 CE.

show very similar contributions to the forecast skill in SIC at all lead times, though adding both variables increases forecast performance more than adding either variable individually (not shown). Given these results, we include all six variables in LIM training except when validating on reanalysis data, which does not include SIT.

Performance of the LIM during out-of-sample validation involves training on all months between 850 and 1650 CE and validation between 1651 and 1850 CE using the CESM1 LME simulation. The LIM shows skill above an AR1 forecast for all variables at 1 month lead times (Figure S4 in Supporting Information S1). For mass fields (PSL and ZG500), such skill is short-lived and disappears beyond 2 months lead time. For sea-ice variables (SIT and SIC) and TAS, LIM outperforms an AR1 forecast through 8 months lead times and 7 months lead times for SST. Generally, the shorter-lived predictability of PSL and ZG500 makes physical sense given the relatively short-lived nature of pressure anomalies in the atmosphere relative to SST anomalies, which tend to be more persistent. The longer lived predictability of SIT is also consistent with previous work showing that SIT anomalies persist across seasons (Balan-Sarojini et al., 2021) and up to a year (Blanchard-Wrigglesworth, Armour, et al., 2011).

In terms of the spatial distribution of skill, we find positive CE values everywhere except for small regions near the ice edge where sea-ice variability is only present during a small number of months in the validation data set (Figure S5 in Supporting Information S1). RMSE is generally greatest near the sea-ice edge, with largest values near the Fram Strait into the Barents Sea. Figure 2 shows the difference in spatial skill of 1-month forecasts between the LIM and AR1 forecasts. The LIM outperforms an AR1 forecast in nearly all regions, increasing the correlation and CE value by up to approximately 0.24. RMSE is also reduced or remains the same everywhere with a maximum reduction of approximately 3.8%. To place this value in context, we note that the maximum mean monthly standard deviation in SIC across all grid cells in the CESM1 LME simulation (1850–2005 CE) is approximately 30%. The reduction in RMSE, relative to this standard deviation, is highest (0.28 sigma) in the Pacific sector, Hudson Bay, and Kara Sea (see Figure S6 in Supporting Information S1).

These intramodel, out-of-sample results indicate that the LIM can not only predict SIC well on seasonal timescales, but other important oceanic and atmospheric variables of the Arctic climate state. We have optimized the LIM to predict SIC, but it maintains skill in other variables for long lead times (up to 8 months, Figure S4 in Supporting Information S1). Furthermore, the LIM is able to predict SIC on 1 month lead times throughout the Arctic region, showing better skill than an AR1 forecast everywhere. These results indicate that the LIM is a useful tool for efficient emulation of coupled model simulations.

Figure 2. Squared correlation coefficient, coefficient of efficiency (CE), and root mean squared error (RMSE) difference between 1 month forecasts of Arctic sea-ice concentration from a Linear Inverse Model (LIM) and autoregressive model of order one at each grid cell. Both models were trained using a monthly CESM1 LME simulation from 850 to 1650 CE and validated between 1651 and 1850 CE.

3.2. Cross-Model Validation

To investigate our second research question and determine under what conditions the LIM works well for predicting Arctic sea-ice conditions, we perform cross-model validation. For these experiments, we initialize the LIM using data that originate from a different model than the LIM training data (CESM1 LME simulation from 850 to 1650 CE). Generally, we find that the LIM fails to outperform an AR1 forecast (trained on the same model data as the LIM) when initialized with models outside CESM1 and outside the mean state conditions of the last millennium. Forecast skill as a function of lead time for the cross-model experiments are shown in Figure S7 in Supporting Information S1. For all cross-model validation experiments, the LIM outperforms climatology on 1 month timescales and up to 4 months when using CESM1 historical simulations. Only experiments performed using historical or last millennium simulations from CCSM4 and CESM1 outperform an AR1 forecast (Figure S7 in Supporting Information S1).

To summarize and evaluate LIM performance when trained using CESM1 LME and validated using different data sets, we compare the variability in the validation data relative to the training data set. Specifically, we calculate the spatial correlation between the training and validation patterns of variability to represent spatial differences. Figure 3 shows this spatial correlation versus the total Arctic RMSE for different LIM forecasts relative to
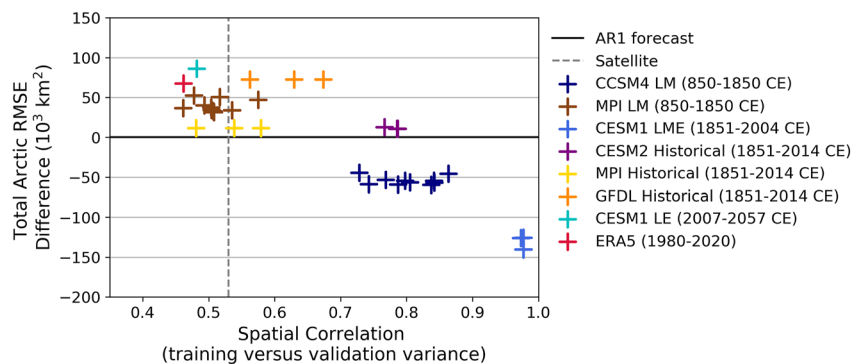
Figure 3. Total Arctic root mean squared error (RMSE) for 1 month Linear Inverse Model (LIM) forecasts validated on various different model simulations (different colors) versus the spatial correlation between the variance (across time) of each validation data set and the LIM training data set. The total Arctic RMSE difference is defined by the LIM forecast minus that from autoregressive model of order one (AR1). For all forecasts, the LIM and AR1 models are trained using a CESM1 LME simulation from 850 to 1650 CE. Validation periods were chosen as 100 years nonoverlapping segments for last millennium simulations (LM) and all 50 years nonoverlapping segments for historical simulations.

an AR1 forecast of SIC. The vertical axis in Figure 3 (and Figure S7 in Supporting Information S1) indicates that the LIM is able to outperform an AR1 forecast when initialized with CCSM4 last millennium simulations (navy blue), but not with MPI last millennium simulations (brown). Similarly when initialized with CMIP6 historical simulations (1851–2014 CE) from MPI, GFDL, and CESM2 as well as a future scenario from the CESM2 LE, the LIM was unable to outperform an AR1 forecast. Finally, we tried to validate the LIM using ERA5 reanalysis data from 1980 to 2020 CE and this LIM also failed to outperform an AR1 forecast (SIT not included).

Figure 3 also indicates that, overall, this spatial correlation is a good predictor of LIM performance. Thus, the locations of the variability in the training and validation data sets have to be sufficiently colocated in order for the LIM to outperform an AR1 forecast. Applying least squares regression to the data shown in Figure 3 yields an R^2-value of 0.77. The x-intercept of this best fit line is 0.66, indicating that a spatial correlation between the training and validation data variance must exceed that value in order to likely outperform an AR1 forecast. A similar analysis was also done for SIT (see Figure S8 in Supporting Information S1) and the x-intercept of the best fit line is 0.55 (R^2-value of 0.72). These results confirm that LIMs are most useful when applied to time periods exhibiting stationary statistics and that the LIM inherits model biases from the training data set.

4. Conclusions and Discussion

Overall, we find that a LIM trained using a CESM1 LME simulation performs well as an emulator, yielding skillful forecasts of Arctic sea ice and other climate variables to 8 months lead, when validated on data from the model used for training, including out-of-sample validation in time. We find that at least 500 years of monthly data are needed for training and 100 or more years of monthly validation data to reliably quantify skill. We also find that TAS and SST contribute most to Arctic SIC prediction on 1–3 months timescales, while SIT becomes more important on 4–6 months timescales.

When validated on out-of-sample intramodel data, the LIM performs well not only for SIC prediction but for all of the six variables included. The predictions outperform an AR1 forecast for all variables except for PSL and ZG500 for lead times beyond 2 months. This highlights the ability of the LIM to capture a broader picture of the Arctic system despite being optimized for predicting sea-ice coverage. The LIM outperforms an AR1 forecast throughout the Arctic and shows the most skill in regions near the ice edge with the most variability in SIC.

We test the ability of the LIM to predict conditions when initialized with other last millennium simulations, historical simulations, reanalysis data, and a future scenario, although the LIMs produce skillful forecasts to 1–4 months lead time, none outperform an AR1 forecast (except the CCSM4 last millennium simulation). This result is likely due in part to model bias and in part to changes in climate statistics depending on the scenario of interest. Previous work has found a large range of persistence estimates across models and a general overestimation of persistence in models compared to observations (Blanchard-Wrigglesworth & Bushuk, 2019; Giesse et al., 2021), which may contribute to the lack of skill we see during cross-model validation. We find that the spatial correlation between the variability (in time) of the training versus validation data can be used as a good predictor (R^2-value of 0.77) of whether the LIM will outperform an AR1 forecast, with a spatial correlation greater than 0.66 between the training and validation needed for LIM forecasts to outperform AR1.

Given this criteria, one can pursue training and validation data sets that best fit a period of interest. If a LIM were to be trained for present-day forecasting, more sample training data are needed than are available from satellite and reanalysis products. Furthermore, a training data set's variability would need to correlate well with satellite observations. The training data set we use here, CESM1 LME, has a spatial correlation of 0.57 with satellite observations between 1980 and 2015 CE which does not pass the threshold determined here. This value drops to 0.42 between 2000 and 2015 CE.

As described here, the LIM is best used as a model emulator as it seems to inherit model biases from the training data set. However, given that the LIM can predict a diverse set of variables across the climate system at minimal computational cost, it could also be a useful tool for probing coupled model dynamics in the Arctic.

Data Availability Statement

All data used to train and validate models for this work were derived from other sources. CESM1 LME simulations are available through Otto-Bliesner et al. (2016), CCSM4 last millennium simulations through Landrum et al. (2013), and MPI last millennium simulations from Jungclaus et al. (2012). Historical simulations from

CESM2 are available through Danabasoglu et al. (2020), MPI historical simulations through Mauritsen et al. (2019), and GFDL historical simulations from Dunne et al. (2020). ERA5 reanalysis data are available from Hersbach et al. (2020) and CESM1 LE simulations through Kay et al. (2015).

References

Alexander, M. A., Matrosova, L., Penland, C., Scott, J. D., & Chang, P. (2008). Forecasting Pacific SSTs: Linear Inverse Model predictions of the PDO. *Journal of Climate*, 21(2), 385–402. https://doi.org/10.1175/2007JCLI1849.1

Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., et al. (2021). Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nature Communications*, 12(1), 5124. https://doi.org/10.1038/s41467-021-25257-4

Armour, K. C., Eisenman, I., Blanchard-Wrigglesworth, E., McCusker, K. E., & Bitz, C. M. (2011). The reversibility of sea ice loss in a state-of-the-art climate model. *Geophysical Research Letters*, 38, L16705. https://doi.org/10.1029/2011GL048739

Balan-Sarojini, B., Tietsche, S., Mayer, M., Balmaseda, M., Zuo, H., de Rosnay, P., et al. (2021). Year-round impact of winter sea ice thickness observations on seasonal forecasts. *The Cryosphere*, 15(1), 325–344. https://doi.org/10.5194/tc-15-325-2021

Bitz, C. M., Holland, M. M., Hunke, E. C., & Moritz, R. E. (2005). Maintenance of the sea-ice edge. *Journal of Climate*, 18(15), 2903–2921. https://doi.org/10.1175/JCLI3428.1

Blanchard-Wrigglesworth, E., Armour, K. C., Bitz, C. M., & DeWeaver, E. (2011). Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations. *Journal of Climate*, 24(1), 231–250. https://doi.org/10.1175/2010JCLI3775.1

Blanchard-Wrigglesworth, E., Barthélemy, A., Chevallier, M., Cullather, R., Fučkar, N., Massonnet, F., et al. (2017). Multi-model seasonal forecast of Arctic sea-ice: Forecast uncertainty at pan-Arctic and regional scales. *Climate Dynamics*, 49(4), 1399–1410. https://doi.org/10.1007/s00382-016-3388-9

Blanchard-Wrigglesworth, E., Bitz, C. M., & Holland, M. M. (2011). Influence of initial conditions and climate forcing on predicting Arctic sea ice. *Geophysical Research Letters*, 38, L18503. https://doi.org/10.1029/2011GL048807

Blanchard-Wrigglesworth, E., & Bushuk, M. (2019). Robustness of Arctic sea-ice predictability in GCMs. *Climate Dynamics*, 52(9–10), 5555–5566. https://doi.org/10.1007/s00382-018-4461-3

Blanchard-Wrigglesworth, E., Cullather, R. I., Wang, W., Zhang, J., & Bitz, C. M. (2015). Model forecast skill and sensitivity to initial conditions in the seasonal Sea Ice Outlook. *Geophysical Research Letters*, 42, 8042–8048. https://doi.org/10.1002/2015GL065860

Bonan, D. B., & Blanchard-Wrigglesworth, E. (2020). Nonstationary teleconnection between the Pacific Ocean and Arctic sea ice. *Geophysical Research Letters*, 47, e2019GL085666. https://doi.org/10.1029/2019GL085666

Bushuk, M., Giannakis, D., & Majda, A. J. (2015). Arctic sea ice reemergence: The role of large-scale oceanic and atmospheric variability. *Journal of Climate*, 28(14), 5477–5509. https://doi.org/10.1175/JCLI-D-14-00354.1

Bushuk, M., Yang, X., Winton, M., Msadek, R., Harrison, M., Rosati, A., & Gudgel, R. (2019). The value of sustained ocean observations for sea ice predictions in the Barents Sea. *Journal of Climate*, 32(20), 7017–7035. https://doi.org/10.1175/JCLI-D-19-0179.1

Cavanaugh, N. R., Allen, T., Subramanian, A., Mapes, B., Seo, H., & Miller, A. J. (2015). The skill of atmospheric Linear Inverse Models in hindcasting the Madden–Julian Oscillation. *Climate Dynamics*, 44(3–4), 897–906. https://doi.org/10.1007/s00382-014-2181-x

Danabasoglu, G., Lamarque, J., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001916. https://doi.org/10.1029/2019MS001916

Dias, D. F., Subramanian, A., Zanna, L., & Miller, A. J. (2019). Remote and local influences in forecasting Pacific SST: A Linear Inverse Model and a multimodel ensemble study. *Climate Dynamics*, 52(5–6), 3183–3201. https://doi.org/10.1007/s00382-018-4323-z

Drobot, S. D., Maslanik, J. A., & Fowler, C. (2006). A long-range forecast of Arctic summer sea-ice minimum extent. *Geophysical Research Letters*, 33, L10501. https://doi.org/10.1029/2006GL026216

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., et al. (2020). The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS002015. https://doi.org/10.1029/2019MS002015

Eicken, H. (2013). Arctic sea ice needs better forecasts. *Nature*, 497(7450), 431–433. https://doi.org/10.1038/497431a

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Farrell, B. F. (1982). The initial growth of disturbances in a baroclinic flow. *Journal of the Atmospheric Sciences*, 39(8), 1663–1686. https://doi.org/10.1175/1520-0469(1982)039<1663:TIGODI>2.0.CO;2

Giesse, C., Notz, D., & Baehr, J. (2021). On the origin of discrepancies between observed and simulated memory of Arctic sea ice. *Geophysical Research Letters*, 48, e2020GL091784. https://doi.org/10.1029/2020GL091784

Gregory, J. M., Stott, P. A., Cresswell, D. J., Rayner, N. A., Gordon, C., & Sexton, D. M. H. (2002). Recent and future changes in Arctic sea ice simulated by the HadCM3 AOGCM. *Geophysical Research Letters*, 29(24), 2175. https://doi.org/10.1029/2001GL014575

Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J. J., Déqué, M., Doblas-Reyes, F. J., et al. (2016). A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, 142(695), 546–561. https://doi.org/10.1002/qj.2401

Hall, C. M., & Saarinen, J. (2010). Polar tourism: Definitions and dimensions. *Scandinavian Journal of Hospitality and Tourism*, 10(4), 448–467. https://doi.org/10.1080/15022250.2010.521686

Henderson, S. A., Vimont, D. J., & Newman, M. (2020). The critical role of non-normality in partitioning tropical and extratropical contributions to PNA growth. *Journal of Climate*, 33(14), 6273–6295. https://doi.org/10.1175/JCLI-D-19-0555.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hogg, J., Fonoberova, M., & Mezić, I. (2020). Exponentially decaying modes and long-term prediction of sea ice concentration using Koopman Mode Decomposition. *Scientific Reports*, 10(1), 16313. https://doi.org/10.1038/s41598-020-73211-z

Holland, M. M., Landrum, L., Bailey, D., & Vavrus, S. (2019). Changing seasonal predictability of Arctic summer sea ice area in a warming climate. *Journal of Climate*, 32(16), 4963–4979. https://doi.org/10.1175/JCLI-D-19-0034.1

Holland, M. M., & Stroeve, J. (2011). Changing seasonal sea ice predictor relationships in a changing Arctic climate. *Geophysical Research Letters*, 38, L18501. https://doi.org/10.1029/2011GL049303

Huddart, B., Subramanian, A., Zanna, L., & Palmer, T. (2017). Seasonal and decadal forecasts of Atlantic sea surface temperatures using a Linear Inverse Model. *Climate Dynamics*, 49(5–6), 1833–1845. https://doi.org/10.1007/s00382-016-3375-1

Jung, T., Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., et al. (2016). Advancing polar prediction capabilities on daily to seasonal time scales. *Bulletin of the American Meteorological Society*, *97*(9), 1631–1647. https://doi.org/10.1175/BAMS-D-14-00246.1

Jungclaus, J., Giorgetta, M., Reick, C., Legutke, S., Brovkin, V., Crueger, T., et al. (2012). *CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-P model: The past1000 experiment, served by ESGF*. World Data Center for Climate (WDCC) at DKRZ. https://doi.org/10.1594/WDCC/CMIP5.MXEPpk

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

Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom, N., & Teng, H. (2013). Last millennium climate and its variability in CCSM4. *Journal of Climate*, *26*(4), 1085–1111. https://doi.org/10.1175/JCLI-D-11-00326.1

Lemke, P., Trinkl, E. W., & Hasselmann, K. (1980). Stochastic dynamic analysis of polar sea ice variability. *Journal of Physical Oceanography*, *10*(12), 2100–2120. https://doi.org/10.1175/1520-0485(1980)010⟨2100:SDAOPS⟩2.0.CO;2

Lindsay, R., & Schweiger, A. (2015). Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. *The Cryosphere*, *9*(1), 269–283. https://doi.org/10.5194/tc-9-269-2015

Lindsay, R. W., Zhang, J., Schweiger, A. J., & Steele, M. A. (2008). Seasonal predictions of ice extent in the Arctic Ocean. *Journal of Geophysical Research*, *113*, C02023. https://doi.org/10.1029/2007JC004259

Mahlstein, I., & Knutti, R. (2012). September Arctic sea ice predicted to disappear near 2°C global warming above present. *Journal of Geophysical Research*, *117*, D06104. https://doi.org/10.1029/2011JD016709

Maslanik, J. A., Fowler, C., Stroeve, J., Drobot, S., Zwally, J., Yi, D., & Emery, W. (2007). A younger, thinner Arctic ice cover: Increased potential for rapid, extensive sea-ice loss. *Geophysical Research Letters*, *34*, L24501. https://doi.org/10.1029/2007GL032043

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*, 998–1038. https://doi.org/10.1029/2018MS001400

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Newman, M., Sardeshmukh, P. D., & Penland, C. (2009). How important is air–sea coupling in ENSO and MJO evolution? *Journal of Climate*, *22*(11), 2958–2977. https://doi.org/10.1175/2008JCLI2659.1

Olonscheck, D., Mauritsen, T., & Notz, D. (2019). Arctic sea-ice variability is primarily driven by atmospheric temperature fluctuations. *Nature Geoscience*, *12*(6), 430–434. https://doi.org/10.1038/s41561-019-0363-1

Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., et al. (2016). Climate variability and change since 850 CE: An ensemble approach with the Community Earth System Model. *Bulletin of the American Meteorological Society*, *97*(5), 735–754. https://doi.org/10.1175/BAMS-D-14-00233.1

Penland, C. (1989). Random forcing and forecasting using principal oscillation pattern analysis. *Monthly Weather Review*, *117*(10), 2165–2185. https://doi.org/10.1175/1520-0493(1989)117⟨2165:RFAFUP⟩2.0.CO;2

Penland, C. (1996). A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D: Nonlinear Phenomena*, *98*(2–4), 534–558. https://doi.org/10.1016/0167-2789(96)00124-8

Penland, C., & Matrosova, L. (1994). A balance condition for stochastic numerical models with application to the El Niño–Southern Oscillation. *Journal of Climate*, *7*(9), 1352–1372. https://doi.org/10.1175/1520-0442(1994)007⟨1352:ABCFSN⟩2.0.CO;2

Penland, C., & Matrosova, L. (1998). Prediction of tropical Atlantic sea surface temperatures using Linear Inverse Modeling. *Journal of Climate*, *11*(3), 483–496. https://doi.org/10.1175/1520-0442(1998)011⟨0483:POTASS⟩2.0.CO;2

Penland, C., & Sardeshmukh, P. D. (1995). The optimal growth of tropical sea surface temperature anomalies. *Journal of Climate*, *8*(8), 1999–2024. https://doi.org/10.1175/1520-0442(1995)008⟨1999:TOGOTS⟩2.0.CO;2

Perkins, W. A., & Hakim, G. (2020). Linear Inverse Modeling for coupled atmosphere–ocean ensemble climate prediction. *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001778. https://doi.org/10.1029/2019MS001778

Petty, A. A., Schröder, D., Stroeve, J. C., Markus, T., Miller, J., Kurtz, N. T., et al. (2017). Skillful spring forecasts of September Arctic sea ice extent using passive microwave sea ice observations. *Earth's Future*, *5*, 254–263. https://doi.org/10.1002/2016EF000495

Rigor, I. G., Wallace, J. M., & Colony, R. L. (2002). Response of sea ice to the Arctic Oscillation. *Journal of Climate*, *15*(18), 2648–2663. https://doi.org/10.1175/1520-0442(2002)015⟨2648:ROSITT⟩2.0.CO;2

Serreze, M. C., Holland, M. M., & Stroeve, J. (2007). Perspectives on the Arctic's shrinking sea-ice cover. *Science*, *315*(5818), 1533–1536. https://doi.org/10.1126/science.1139426

Shin, S.-I., Sardeshmukh, P. D., Newman, M., Penland, C., & Alexander, M. A. (2021). Impact of annual cycle on ENSO variability and predictability. *Journal of Climate*, *34*(1), 171–193. https://doi.org/10.1175/JCLI-D-20-0291.1

Smith, L. C., & Stephenson, S. R. (2013). New trans-Arctic shipping routes navigable by midcentury. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(13), E1191–E1195. https://doi.org/10.1073/pnas.1214212110

Stroeve, J., Hamilton, L. C., Bitz, C. M., & Blanchard-Wrigglesworth, E. (2014). Predicting September sea ice: Ensemble skill of the SEARCH Sea Ice Outlook 2008–2013. *Geophysical Research Letters*, *41*, 2411–2418. https://doi.org/10.1002/2014GL059388

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). *An overview of CMIP5 and the experiment design*. American Meteorological Society. http://doi.org/10.1175/BAMS-D-11-00094.1

Tivy, A., Alt, B., Howell, S., Wilson, K., & Yackel, J. (2007). Long-range prediction of the shipping season in Hudson Bay: A statistical approach. *Weather and Forecasting*, *22*(5), 1063–1075. https://doi.org/10.1175/WAF1038.1

Winkler, C. R., Newman, M., & Sardeshmukh, P. D. (2001). A linear model of wintertime low-frequency variability. Part I: Formulation and forecast skill. *Journal of Climate*, *14*(24), 4474–4494. https://doi.org/10.1175/1520-0442(2001)014⟨4474:ALMOWL⟩2.0.CO;2

Yeager, S. G., Karspeck, A. R., & Danabasoglu, G. (2015). Predicted slowdown in the rate of Atlantic sea ice loss. *Geophysical Research Letters*, *42*, 10704–10713. https://doi.org/10.1002/2015GL065364

Yuan, X., Chen, D., Li, C., Wang, L., & Wang, W. (2016). Arctic sea ice seasonal prediction by a linear Markov model. *Journal of Climate*, *29*(22), 8151–8173. https://doi.org/10.1175/JCLI-D-15-0858.1