Regional September Sea Ice Forecasting with Complex Networks and Gaussian Processes

WILLIAM GREGORY AND MICHEL TSAMADOS

Centre for Polar Observation and Modelling, Earth Sciences, University College London, London, United Kingdom

JULIENNE STROEVE

Centre for Polar Observation and Modelling, Earth Sciences, University College London, London, United Kingdom, and National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado

PETER SOLLICH

Department of Mathematics, King's College London, London, United Kingdom, and Institute for Theoretical Physics, Georg-August-University Göttingen, Göttingen, Germany

ABSTRACT

Reliable predictions of the Arctic sea ice cover are becoming of paramount importance for Arctic communities and industry stakeholders. In this study pan-Arctic and regional September mean sea ice extents are forecast with lead times of up to 3 months using a complex network statistical approach. This method exploits relationships within climate time series data by constructing regions of spatiotemporal homogeneity (i.e., nodes), and subsequently deriving teleconnection links between them. Here the nodes and links of the networks are generated from monthly mean sea ice concentration fields in June, July, and August; hence, individual networks are constructed for each respective month. Network information is then utilized within a linear Gaussian process regression forecast model, a Bayesian inference technique, in order to generate predictions of sea ice extent. Pan-Arctic forecasts capture a significant amount of the variability in the satellite observations of September sea ice extent, with detrended predictive skills of 0.53, 0.62, and 0.81 at 3-, 2-, and 1-month lead times, respectively. Regional forecasts are also performed for nine Arctic regions. On average, the highest predictive skill is achieved in the Canadian Archipelago, Beaufort, Chukchi, East Siberian, Laptev, and Kara Seas, although the ability to accurately predict many of these regions appears to be changing over time.

1. Introduction

Passive microwave satellite observations since 1979 have shown a consistent decline in pan-Arctic sea ice extent (SIE) for all months of the year, in direct accordance with a warming planet as a result of anthropogenic CO_2 emissions (Notz and Stroeve 2016). Across all months, September SIE has shown the fastest decline (Overland and Wang 2013; Serreze and Stroeve 2015) as well as the largest interannual variability (Stroeve and Notz 2018). Significant anomalies such as the 1996 maximum and the 2007 and 2012

minima contribute largely to this variability and make seasonal predictions difficult. The extreme 2007 minimum in particular led to the initiation of the Sea Ice Outlook (SIO) project by the Study of Environmental Arctic Change (SEARCH), with the initial aim of providing annual summaries of the expected pan-Arctic September SIE. In 2014 this was formally turned in to the Sea Ice Prediction Network through U.S. interagency funding, and expanded to include regional forecasts of September sea ice concentration (SIC), sea ice probability (SIP), first ice-free date, and first ice-advance date. Starting each May, the SIO solicits predictions of the September sea ice cover from the sea ice community based on forecasts made on the first day of June, July, and August. Approaches

Corresponding author: William Gregory, william.gregory.17@ ucl.ac.uk

vary from fully coupled ice–ocean–atmosphere or ice–ocean dynamical models, to statistical models and heuristic and/or best guesses. The ability of these forecast models to accurately capture the variability of Arctic sea ice, with lead times of days to months, is becoming increasingly important from an ecological, economical and societal standpoint. The opening of the Northern sea route for example has the potential to reduce the shipping route between Europe and Asia by 5000 nautical miles in comparison to the Suez Canal (Lee and Song 2014), hence advanced timing of this opening, on the order of months, can make for cost-effective planning (Eguíluz et al. 2016).

While fully coupled climate models are intuitively the model of choice for sea ice forecasting, with a number of studies reporting high detrended prediction skill (Chevallier et al. 2013; Sigmond et al. 2013; Wang et al. 2013; Msadek et al. 2014; Peterson et al. 2015; Bushuk et al. 2017), on average past summaries of SIO submissions (Stroeve et al. 2014; Hamilton and Stroeve 2016) have shown that generally dynamical models do not significantly outperform their statistical model counterparts. In fact the majority of models participating in the SIO (both statistical and dynamical) only perform relatively successfully when sea ice anomalies occur close to the long-term trend. This leaves a window open to explore statistical models, which have in some cases shown promise in exploiting sources of predictability with up to 4-month lead time (Lindsay et al. 2008; Schröder et al. 2014; Yuan et al. 2016; Stroeve et al. 2016; Petty et al. 2017; Ionita et al. 2019).

This paper presents a statistical approach that exploits spatiotemporal variability within the sea ice record via construction of complex networks based on monthly mean fields of SIC from June, July, and August. These networks are then utilized within a Gaussian process regression (GPR) model in order to predict both pan-Arctic and regional September monthly mean SIE, for all years between 1985 and 2019. In the context of real-time forecasts, this would correspond to forecasts being made on 1 July (June data), 1 August (July data), and 1 September (August data), with forecast validation made on 1 October. Hence forecast lead times extend from 1 to 3 months.

Networks provide a relatively new perspective with which to extract significant statistical relationships from the multitude of entangled interactions within the climate system, and are a useful complement to standard methods of analyzing patterns of climate variability (Tsonis and Roebber 2004; Donges et al. 2009; Guez et al. 2012; Radebach et al. 2013; Fountalis et al. 2014; Boers et al. 2014; Donges et al. 2015;

Dijkstra et al. 2019). These entangled interactions are more commonly referred to as climatological teleconnections, which in themselves have been studied for much of the last century (Walker and Bliss 1932; Wallace and Gutzler 1981; Glantz et al. 1991), and also in recent years for the similar purpose of Arctic sea ice prediction (Yuan et al. 2016; Comeau et al. 2019). GPR is then a natural complement to complex networks. The construction of kernel functions for random graphs (networks) has been studied extensively (Kondor and Lafferty 2002; Smola and Kondor 2003; Rasmussen and Williams 2006; Vishwanathan et al. 2010; Spielman 2010; Urry and Sollich 2013), and provides the ability to integrate the network information directly into the regression model in the form of Gaussian prior and likelihood distributions (see section 3b). To our knowledge, this is the first time this methodology has been implemented for the purpose of sea ice forecasting.

The paper is structured as follows: section 2 provides an overview of the data used for this study. Section 3 describes the methodology of both the generation of complex networks and a GPR model. Results are presented in section 4, followed by a discussion (section 5) and conclusions (section 6).

2. Data

For the generation of the complex networks, monthly mean SIC fields between 1979 and 2018 were extracted from the National Snow and Ice Data Center (NSIDC) based on the NASA Team sea ice algorithm applied to passive microwave brightness temperatures (Cavalieri et al. 1996). Multiple satellites comprise this data record: *Nimbus-7* SSMR (1979–87), the DMSP *F-8*, *F-11*, and *F-13* SSM/Is (1987–2007), and finally the DMSP *F-18* SSM/I (2007–17). To extend beyond 2018, these data are combined with the 2019 near-real-time daily SIC fields from NSIDC (Maslanik and Stroeve 1999), which are averaged to generate the corresponding monthly fields. Datasets are provided on a polar stereographic 25 km × 25 km grid, which are regridded here to a polar stereographic 100 km × 100 km grid prior to the calculation of the cell-level network (see section 3a) for computational reasons. Gridcell area information (used to generate area-weighted time series, see section 3a) was also extracted from NSIDC's pixel area tools library. Finally, it should be noted that the time-varying polar hole in the SIC data is filled for the purpose of this study. This is achieved by simply filling the hole with the mean SIC value at 0.5° below the hole latitude, for each respective year.

To predict SIE on a regional level, the Arctic is separated into nine geographic areas that surround the Arctic coastlines and encompass the most likely maritime navigation routes (Fig. 1)—the data mask for these areas is also available from NSIDC (Fetterer et al. 2010). Here SIE is defined as the sum of gridcell areas for which SIC values are greater than or equal to 15%. A monthly mean value of extent is produced for each year by calculating the SIE for each day in September and then taking an average over all days.

3. Methods

a. Complex networks

A network can be seen as groups of nodes (vertices) whereby each node i may be connected to any other node in the network j via links (edges)—the notation hereafter follows that i and j are used to index different network nodes, whereas p and q are used to index different SIC data grid cells. The reader is also referred to appendix A where a full list of variable definitions is given.

In the common example of social networks, people are represented by nodes and their relationships by links. The weight of the link then depends on the strength of the relationship existing between the two people (nodes). In the case of the climate network, a network of N nodes can correspond to time series data $X(t) = \{\chi_i(t)\}_{i=1}^{N}$ representing n observations $\{t_1, \ldots, t_n\}$ at N fixed geographical locations, and the links can represent statistical interdependencies between any pair of nodes $\chi_i(t)$ and $\chi_j(t)$.

The process of generating complex networks from monthly mean SIC data (for June, July, and August) follows similar definitions to previous studies involving climate networks (Donges et al. 2009, 2011; Fountalis et al. 2014, 2018). We summarize the process here in three separate steps:

1) A cell-level network (essentially a correlation matrix) is first computed from the linearly detrended anomaly time series of each SIC grid cell $x_p(t)$ via the linear Pearson cross-correlation metric at zero lag, for all possible pairs of p and q:

$$r_{pq} = \frac{\sum\limits_{k=1}^{n} x_p(t_k) x_q(t_k)}{\sqrt{\left[\sum\limits_{k=1}^{n} x_p^2(t_k)\right]\left[\sum\limits_{k=1}^{n} x_q^2(t_k)\right]}}. \quad (1)$$

2) A clustering algorithm is then used in order to group grid cells into geographic areas, which effectively
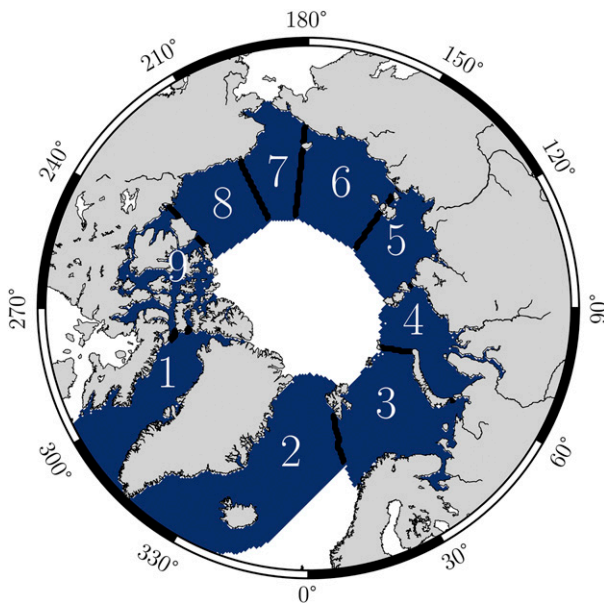

FIG. 1. Regional Arctic areas used for September SIE Forecasting (from NSIDC). Shown are 1) Baffin Bay, 2) Greenland Sea, 3) Barents Sea, 4) Kara Sea, 5) Laptev Sea, 6) East Siberian Sea, 7) Chukchi Sea, 8) Beaufort Sea, and 9) Canadian Archipelago.

represent regions of homogeneity in terms of SIC over the length of the time series. The clustering algorithm uses a greedy approach whereby individual cells continuously search for neighbors to group with providing the correlation between them exceeds a certain threshold. In the fashion of a greedy algorithm, this form of local clustering will be heuristically optimal for the global network in the sense that small-scale homogeneities identified with local clustering will scale to large-scale homogeneity across all network nodes.

For grid cells to form an area A, three criteria must be adhered to

- $|A_i| \geq 2$,
- $\forall (p, q) \in A_i$, a connected path must exist from p to q,
- $\dfrac{\sum\limits_{p \neq q \in A_i} r_{pq}}{|A_i|(|A_i| - 1)} > \tau$,

where $|A_i|$ represents the number of grid cells in area A_i. The third condition represents the fact that the average correlation of all grid cells in each area must be greater than a predefined threshold factor τ. Here the threshold factor is defined as the mean of all positive correlation values (from the cell-level network) which are significant at the 1% level, determined by a one-sided t test:

$$t \, \text{score} = r_{pq} \sqrt{\frac{n-2}{1-r_{pq}^2}}. \qquad (2)$$

The number of areas within the climate network is also minimized. This is achieved by a similar process to the initial generation of areas; except rather than grid cells searching for neighbors, areas instead search for neighboring areas to determine whether, if merged, the average correlation between all the grid cells in a newly merged area exceeds τ. The neighboring area with the greatest average correlation above τ after merging is then chosen. These areas now form the final nodes of the complex network.

3) Links between nodes can be seen as a proxy for teleconnections. The network here is fully connected, such that each node A_i shares a weighted link with every other node of the network. The links are then created by first generating the cumulative anomaly time series for each network node. This is taken as the sum of the area-weighted detrended time series of all grid cells within each respective network node:

$$\chi_i(t) = \sum_{p \in A_i} x_p(t) \sqrt{\psi_p}. \qquad (3)$$

The cells are weighted by the square root of their geographic area ψ (km^2) to account for the fact the data exist on a polar stereographic grid and hence grid cells are not equal in area. The link weights w_{ij} are then calculated as the temporal covariance between two network node anomaly time series:

$$w_{ij} = \frac{1}{n} \sum_{k=1}^{n} [\chi_i(t_k) - \langle \chi_i(t) \rangle][\chi_j(t_k) - \langle \chi_j(t) \rangle]. \qquad (4)$$

Note that the weights are not normalized, such that the weight of the link is proportional to the power of the anomaly time series $\chi_i(t)$ and $\chi_j(t)$ in nodes A_i and A_j, rather than just their correlation value (Fountalis et al. 2014). It is also worth noting that the strength (or weighted degree) of a given network node λ_i is defined here as the sum of the absolute value of all its associated link weights (Fig. 2):

$$\lambda_i = \sum_{j \neq i} |w_{ij}|. \qquad (5)$$

As we use these SIC networks to make predictions of September SIE for all years between 1985 and 2019, it is worth commenting on how the network structure varies over time as new networks are created for each forecast year. Figure 2 provides a visual illustration of this for

August SIC data as an example. The network structure appears to vary significantly across the 34-yr period shown here. Further investigation is required as to whether this variation is due to actual changes in sea ice behavior or the fact that the length of the time series n used to construct each network is not constant. For example, the network labeled 1986 is constructed with $n = 8$, whereas the network labeled 2019 is constructed with $n = 41$ (time series begin at 1979). This does coincide with significant differences in the threshold factor τ, and subsequently the number of network nodes, as shown in Fig. 3. Networks generated with shorter time series show a much larger value of τ which means that SIC grid cells will only cluster with neighboring cells if the correlation between them is significantly large. As this is only likely to occur for perhaps the first few neighboring cells, the resultant network hosts a very large number of nodes which contain a very small number of grid cells. Indeed it may be necessary to impose a predefined threshold of τ for years with shorter time series, although this goes beyond the scope of this study.

b. Gaussian process regression

This section outlines the model setup of a GPR model specific to this study; however, an accompanying theory is provided in appendix B, for the interested reader.

GPR is a Bayesian inference technique commonly used to learn about functional relationships between inputs **X** and outputs **y**, and to subsequently make predictions based on this learning (training). This functional relationship is typically expressed in the form:

$$\mathbf{y} = f(\mathbf{X}) + \varepsilon, \; \varepsilon \sim \mathcal{N}(0, \sigma^2), \qquad (6)$$

where ε represents independent identically distributed Gaussian noise with zero mean and variance σ^2. In our case **y** corresponds to n observations of September SIE anomalies, and **X** corresponds to n observations of the cumulative SIC anomalies in (3) from N network nodes—to keep the notation uncluttered we use **X** to denote **X**(t). Note that the n observations used for training extend from (and include) 1979 to the year preceding the forecast year. For example if the forecast year is 2019 then $n = 40$, hence a new SIC network and GPR model are constructed for each forecast year. The aim of the model training is therefore to investigate the unknown relationship between the nodes of a given SIC network and the target September SIE. Here we assume that this relationship is linear, such that $f(\mathbf{X})$ can be represented in the form $\mathbf{X}\boldsymbol{\beta}$, and subsequently our implementation becomes a *Bayesian linear regression*. As such $f(\mathbf{X})$ is a
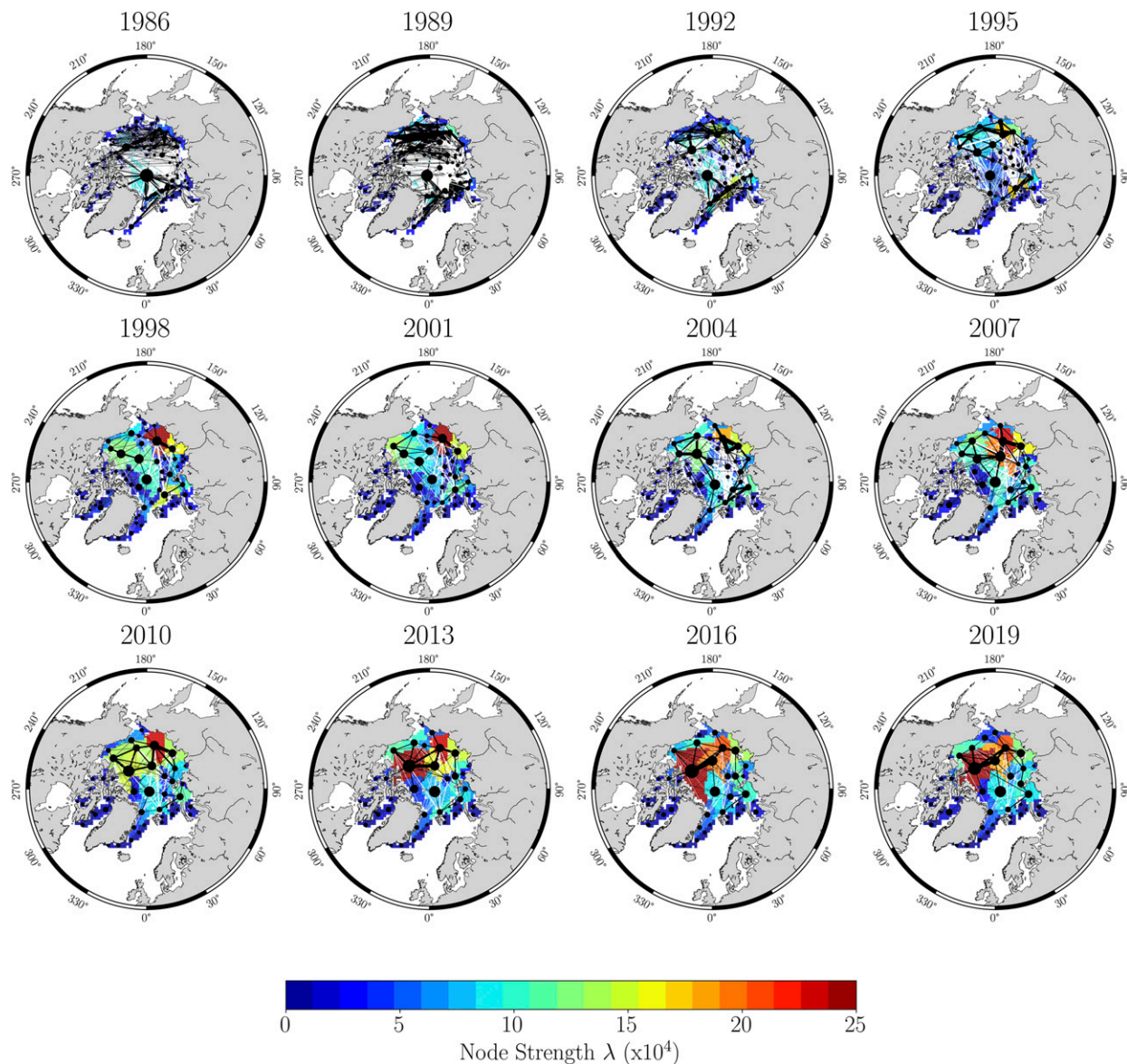
FIG. 2. Network nodes and teleconnection links derived from linearly detrended monthly mean August SIC data. The length of the time series used to create each network extends from 1979 to the years marked above each image. Each network area node A_i can be represented by a single (black) node point, whose size is proportional to $\sum_{p \in A_i} \sqrt{\psi_p}$. The thickness of each of the links here is proportional to the link weight w_{ij} between each pair of nodes, where black(white) links signify positive(negative) covariance. Only links that are significant at the 10% level are shown here to aid visualization. See (5) for the definition of node strength.

Gaussian process (see appendix B), and can be completely specified by its mean (which we take to be zero) and covariance function:

$$f(\mathbf{X}) \sim \mathcal{N}(0, K). \qquad (7)$$

The covariance function K can be computed between training inputs $\mathbf{X}$, test inputs $\mathbf{Z}$, or a combination of both. Test inputs are used to make predictions after the model training and are given here as the nth $+ 1$ entries that $\mathbf{X}$

would have if it were continued by another year (i.e., SIC anomalies of the year being forecast). We can therefore define a series of covariance functions for all training and test inputs as follows:

$$\begin{aligned} K(\mathbf{X}, \mathbf{X}) &= \mathbf{X}\mathbf{\Sigma}\mathbf{X}^{\mathrm{T}} + \sigma^2 \mathbf{I}, \\ K(\mathbf{Z}, \mathbf{Z}) &= \mathbf{Z}\mathbf{\Sigma}\mathbf{Z}^{\mathrm{T}} + \sigma^2, \\ K(\mathbf{X}, \mathbf{Z}) &= \mathbf{X}\mathbf{\Sigma}\mathbf{Z}^{\mathrm{T}}, \\ K(\mathbf{Z}, \mathbf{X}) &= \mathbf{Z}\mathbf{\Sigma}\mathbf{X}^{\mathrm{T}}. \end{aligned} \qquad (8)$$
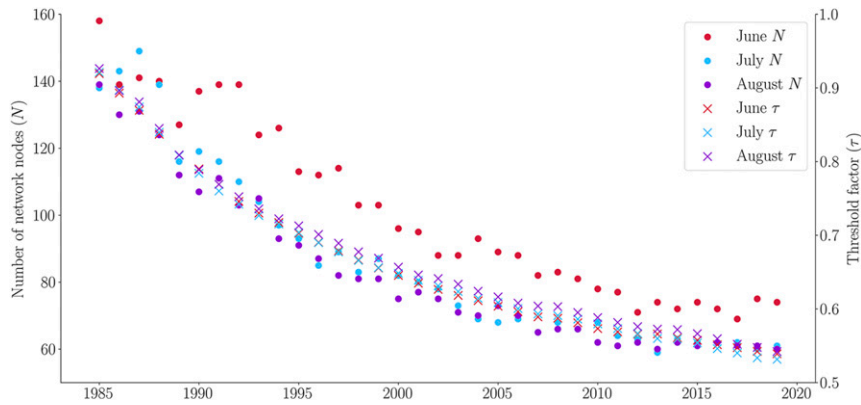
FIG. 3. For networks based on linearly detrended monthly mean SIC data from June, July, and August, the variation in the number of network nodes N and also the threshold factor τ are shown here for all years between 1985 and 2019.

Here, $\mathbf{\Sigma}$ is given as a random walk matrix (Urry and Sollich 2013; see references therein) that takes the form $\mathbf{\Sigma} = \alpha \exp(\ell \mathbf{M})$. The term $\mathbf{M}$ is then a symmetric transition matrix with off-diagonal elements containing the absolute teleconnection weights w_{ij} between nodes of the SIC network (4), and diagonal elements containing the negative strength of a given node (5):

$$\mathbf{M}_{ij} = |w_{ij}|, \quad \text{for} \quad i \neq j; \quad \mathbf{M}_{ii} = -\lambda_i = -\sum_{j \neq i} |w_{ij}| \qquad (9)$$

The remaining undefined variables (ℓ, α, σ^2)—commonly referred to as hyperparameters—are considered as free parameters and play an important role in the minimization of overfitting during the model training (see appendix B for more details). We optimize these hyperparameters here by the widely used empirical numerical optimization approach of type II maximum likelihood [see (B7)]. This is achieved by first defining $\tilde{\sigma}^2 = \sigma^2/\alpha$ and working with the set of hyperparameters $\theta = (\ell, \alpha, \tilde{\sigma}^2)$. This way we can optimize α analytically to obtain $\alpha = n^{-1} \mathbf{y}^{\mathrm{T}} \tilde{\mathbf{K}}^{-1} \mathbf{y}$, where $\tilde{\mathbf{K}} = \mathbf{X} \exp(\ell \mathbf{M}) \mathbf{X}^{\mathrm{T}} + \tilde{\sigma}^2 \mathbf{I}$ (Sollich 2005), such that only ℓ and $\tilde{\sigma}^2$ need to be optimized numerically.

With the optimal hyperparameters found, the GPR model is then fully determined and forecasts of September SIE anomalies $\overline{f}$ and their corresponding variance σ_f^2 (forecasts are Gaussian) are given by

$$\overline{f} = K(\mathbf{Z}, \mathbf{X}) K(\mathbf{X}, \mathbf{X})^{-1} \mathbf{y}$$
$$\sigma_f^2 = K(\mathbf{Z}, \mathbf{Z}) - K(\mathbf{Z}, \mathbf{X}) K(\mathbf{X}, \mathbf{X})^{-1} K(\mathbf{X}, \mathbf{Z}). \qquad (10)$$

4. Results

a. Pan-Arctic September SIE forecast

Forecasts of pan-Arctic September SIE are performed for all years between 1985 and 2019, based on

monthly mean SIC fields from June, July, and August. Each forecast is performed by GPR as explained above, with $\mathbf{X}$ representing linearly detrended SIC data (for all network nodes) taken in the relevant month (June, July, or August), and $\mathbf{y}$ the corresponding pan-Arctic September SIE. As explained in the previous section, for each forecast a new network of area nodes and hence a GPR model is constructed in order to ensure predictors capture the variability of the sea ice record over the period leading up to the year being forecast. Accordingly, optimal hyperparameters for the GPR are also determined separately for each forecast.

Before presenting the results, we give the definition of detrended forecast skill S as the explained variance in the target variable $\mathbf{y}$ (observed detrended pan-Arctic September SIE) relative to the detrended forecast result $\overline{f}$ from (10). Explicitly, $1 - S$ is the ratio of the mean squared prediction error (MSE) to the variance of these targets around their mean $\overline{y}$:

$$S = 1 - \frac{\frac{1}{n}\sum_{k=1}^{n} (y_k - \overline{f}_k)^2}{\frac{1}{n}\sum_{k=1}^{n} (y_k - \overline{y})^2}. \qquad (11)$$

Furthermore, as we will make comparisons to other studies, we give the definition of the anomaly correlation coefficient (ACC) as the linear Pearson cross-correlation metric between $\overline{f}$ and $\mathbf{y}$ from (1), which we restate here:

$$\mathrm{ACC}_{\overline{f}\mathbf{y}} = \frac{\sum_{k=1}^{n} \overline{f}_k y_k}{\sqrt{\left(\sum_{k=1}^{n} \overline{f}_k^2\right)\left(\sum_{k=1}^{n} y_k^2\right)}}. \qquad (12)$$
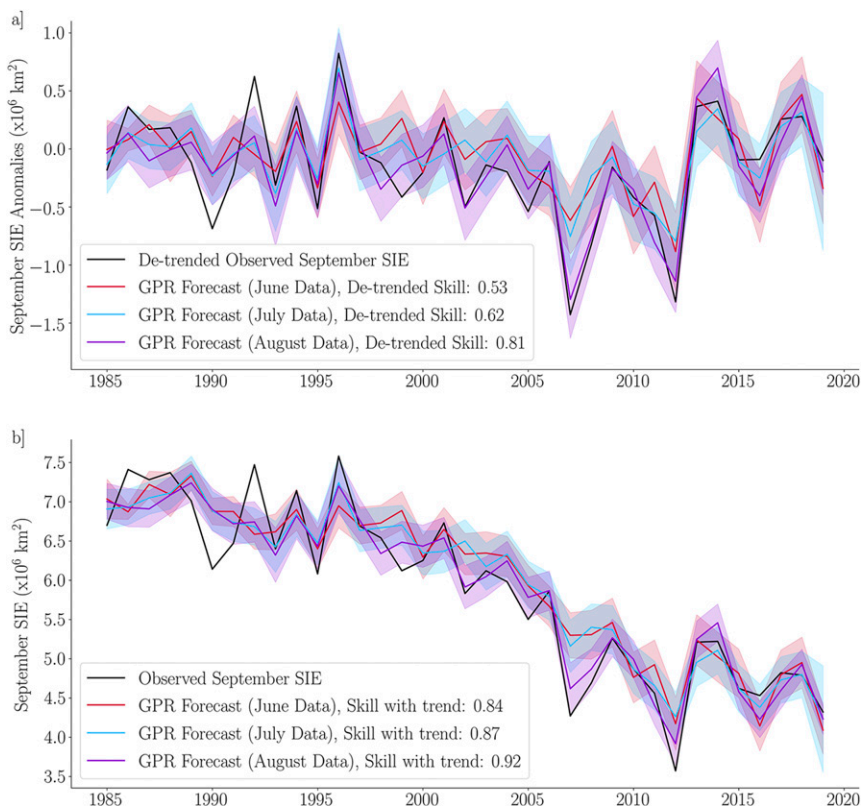
FIG. 4. Predictions of pan-Arctic September SIE based on networks of monthly mean SIC data from June, July, and August SIC (3-, 2-, and 1-month lead time, respectively). (a) Detrended forecast $\bar{f}$ showing predictive skills from (11) of 0.53, 0.62, and 0.81 for forecasts based on June, July, and August SIC networks, respectively. (b) Forecasts with trend, with skills of 0.84, 0.87, and 0.92, respectively. The shaded areas in both plots represent the forecast standard deviation σ_f.

Figure 4 displays the results of the pan-Arctic September SIE forecasts. The detrended result shows remarkable skill of 0.53 at 3-month lead time using a single-layer predictor (in the sense that GPR can be viewed as a single-layer SIC network), and increasing skill thereafter of 0.62 and 0.81 at 2 and 1 months, respectively. Furthermore it is apparent from Fig. 4 that each of the three forecasts appear to have performed better in the latter section of the time series (i.e., after approximately 1992). Therefore if we calculate the detrended skill between 1993 and 2019 we see an increase in skill for all lead times; from (0.53, 0.62, 0.81) to (0.58, 0.66, 0.88), respectively.

The skill value of anomaly persistence between 1985 and 2019 is given here as 0.18, 0.52, and 0.88 at 3-, 2-, and 1-month lead times, respectively; hence, the GPR model outperforms persistence at 3 and 2 months, although at 1 month the skill of persistence is greater. Comparing our forecast results to recent statistical studies, Lindsay et al. (2008), Schröder et al. (2014), Petty et al. (2017), and Ionita et al.

(2019) each presented skill scores for statistical forecasts of pan-Arctic September SIE. Skill values for each respective study were given as [0.63, 0.81, and 0.96 (with trend) at 3, 2, and 1 months, respectively], [0.41 (detrended) at 3–4 months], [0.45, 0.49, and 0.59 (detrended) at 3, 2, and 1 months, respectively], and [0.78 and 0.81 (detrended) at 3 and 2 months, respectively]. Similarly, Yuan et al. (2016) reported detrended ACC scores of [0.75, 0.82, and 0.90 at 3, 2, and 1 months, respectively]; the detrended ACC scores for our forecasts are similar at (0.77, 0.83, 0.90) for the same respective lead times.

Looking to dynamical forecast comparisons, Sigmond et al. (2013), Wang et al. (2013), Msadek et al. (2014), and Bushuk et al. (2017) each presented detrended ACC scores for various coupled model forecasts of pan-Arctic sea ice. Approximate results for each respective study were given as (0.55, 0.75, 0.90), (0.55, 0.65, 0.75), (0.60, 0.70, 0.81), and (0.50, 0.60, 0.70) for lead times of 3, 2, and 1 months, respectively.

It should be noted that forecast skills presented above are not directly comparable as the period of analysis is not consistent between all studies. However, given that our skill results outperform the majority of the studies above, and that our analysis is conducted over a longer time period should give an indication as to the exceptional performance of the combined complex network–GPR approach, for the presented lead times.

b. Regional September SIE forecasts

Regional forecasts of September SIE are also performed to assess the model's performance beyond the single pan-Arctic metric and to be more in-line with stakeholder needs. As stated previously, nine geographic regions have been chosen (Fig. 1). These exclude the Bering Sea, Sea of Okhotsk, and Hudson Bay, as these regions are ice free in September. In a similar vein, the Central Arctic has also been excluded as this is generally the location of the core ice pack, such that predictions there would be of limited use for stakeholders.

Similar to section 4a, a new SIC network is constructed for each forecast year and the hyperparameters for the GPR are optimized separately for each forecast year and region.

Figure 5 displays the results of the regional September SIE forecasts. Across all lead times, predictions are generally most successful in regions toward the Pacific sector, with highest detrended skill achieved in the Canadian Archipelago, Beaufort, Chukchi, East Siberian, and Laptev Seas. Beyond 1-month lead time the GPR model consistently achieves poor prediction skill in Baffin Bay, and both Greenland and Barents Seas, with Baffin Bay showing poor predictability for all lead times. Predictability appears to suffer in regions which have undergone significant changes in behavior over the satellite record. For example, in the Kara Sea, the GPR model achieves very high detrended skill between 1985 and 2006 (0.52, 0.59, and 0.68 at 3-, 2-, and 1-month lead time, respectively); however, between 2007 and 2019 this skill drops significantly (−1.25, −0.44, and 0.35 for the same respective lead times), ultimately having a negative impact on the overall score. This change in predictability is likely due to a sudden change in the interannual variability of SIE in the Kara Sea after 2007 (see Fig. 5b), such a change, which would be extremely difficult for a statistical model to adapt to. The opposite can be said for other regions, in that the dominant window of predictability lies in the latter half of the time series record (i.e., after 2007). Regions such as the Canadian Archipelago, Beaufort,

Chukchi, and East Siberian Seas show significantly higher detrended prediction skill between 2007 and 2019 than between 1985 and 2006 (Fig. 6). See also Table 1 for a summary of the detrended prediction skills for each region, and between different time windows.

Looking again to comparison studies of regional sea ice forecasting using statistical models, Lindsay et al. (2008) made regional predictions of the September mean SIE at 3-month lead time. Reported skill values (with trend) were given as: 0.77, 0.74, 0.18, 0.08, and −0.67 for the Barents, Kara, Laptev, East Siberian, and Beaufort Seas, respectively. Our skill values (also with trend) for the same respective regions at 3-month lead time are given as 0.04, 0.54, 0.45, 0.66, and 0.65. Ionita et al. (2019) made similar predictions in the East Siberian Sea, with detrended skill of 0.69, and 0.78 at 3- and 2-month lead time, respectively. Our skill is given as 0.48 and 0.47 at 3- and 2-month lead time, respectively.

In terms of regional dynamical forecasts, Bushuk et al. (2017) reported detrended ACC scores for all Arctic regions in a coupled model forecast. Here, other than the Greenland Sea at 3-month lead time, our equivalent detrended ACC scores exceed theirs in all regions for each of the three respective lead times.

Note that similar to the pan-Arctic skill comparison, these skill metrics are not completely comparable due to different analysis periods.

5. Discussion

Several studies have linked sea ice persistence to high forecast skill of the pan-Arctic SIE metric with lead times ranging from 1 to 3 months (Drobot et al. 2006; Lindsay et al. 2008; Petty et al. 2017; Ionita et al. 2019). In this study, forecasts of pan-Arctic SIE outperform the majority of both statistical and dynamical forecast models presented in the previous section despite the poor performance between 1985 and 1993, which is likely due to lack of observations in the GPR model training. Indeed heuristic "rules-of-thumb" are available for deciding the minimum number of observations required to effectively model a given statistical problem (Bishop 2006), although here we could pragmatically say that this minimum number should equate to the point after which the model is able to make accurate and precise future predictions. Nevertheless with increasing demand for reliable forecasting of useful sea ice components, which is the focal point of the Phase 2 of the Sea Ice Prediction Network (SIPN2), arguably
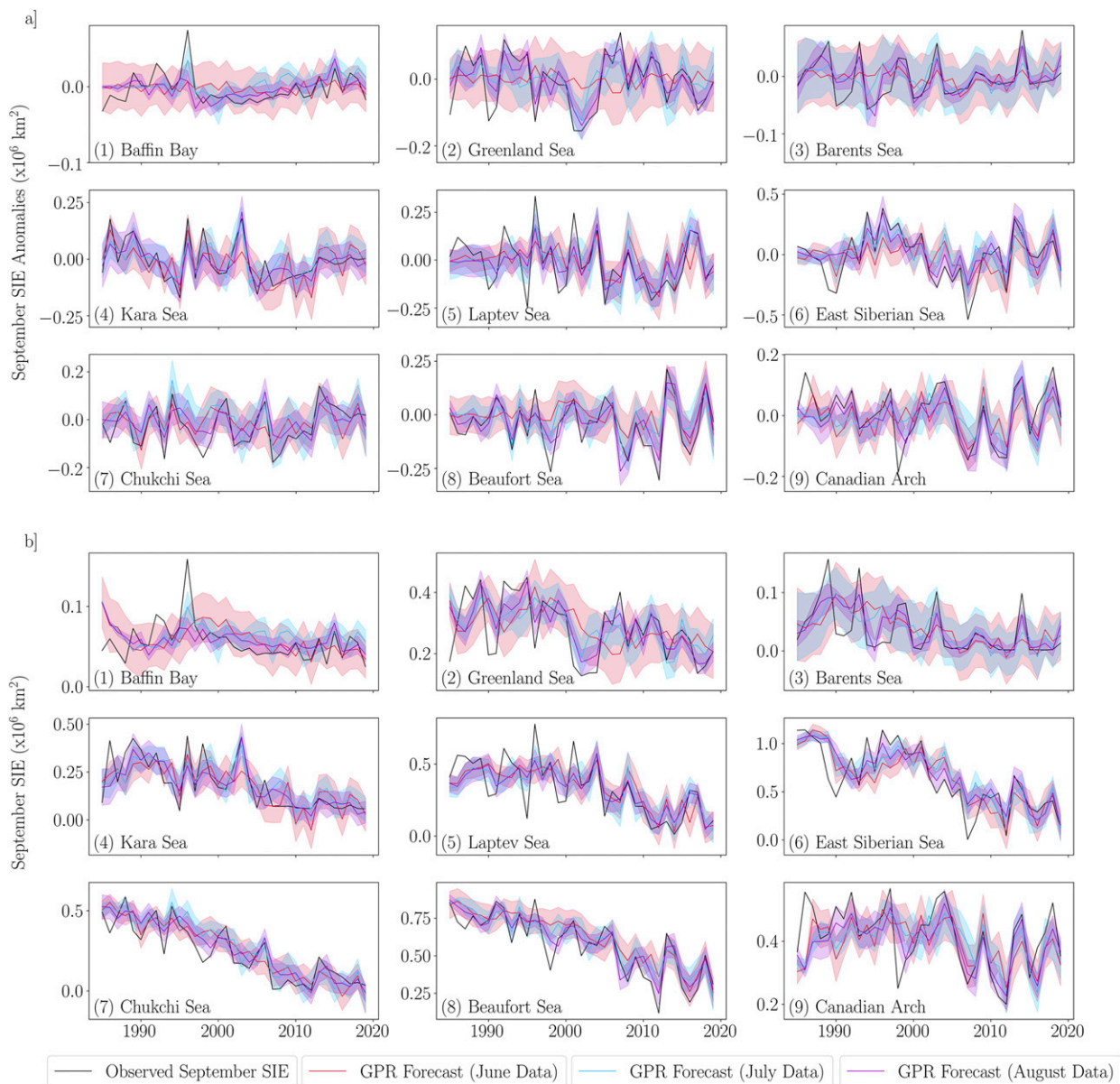
FIG. 5. Predictions of regional September SIE based on networks of monthly mean SIC data from June, July, and August SIC (3-, 2-, and 1-month lead time, respectively). (a) Detrended forecast $\bar{f}$. (b) Forecasts with trend. The shaded areas in all plots represent the forecast standard deviation σ_f.

regional forecasts of SIE and alternative metrics hold more importance than those of a single pan-Arctic extent. Therefore while skillful forecasts have been achieved for pan-Arctic SIE, this must be transferable to other regions to be useful for stakeholders, industry and Arctic communities alike. Here regional forecasts of SIE have proven to be relatively successful at 3-month lead time. Regions with high detrended forecast skill (Canadian Archipelago and Beaufort, Chukchi, East Siberian, Laptev, and Kara

Seas) coincide with areas which have been shown to contribute most to the decline in September SIE from 1979 (Onarheim et al. 2018). Similarly, regions with relatively poor detrended forecast skill (Barents Sea, Greenland Sea, and Baffin Bay) are those which show little variability in their September extents and in fact have been shown to contribute most to the decline in March SIE from 1979 (Onarheim et al. 2018). In any case, future work must move to accurate predictions of spatiotemporal trends and derivation of metrics
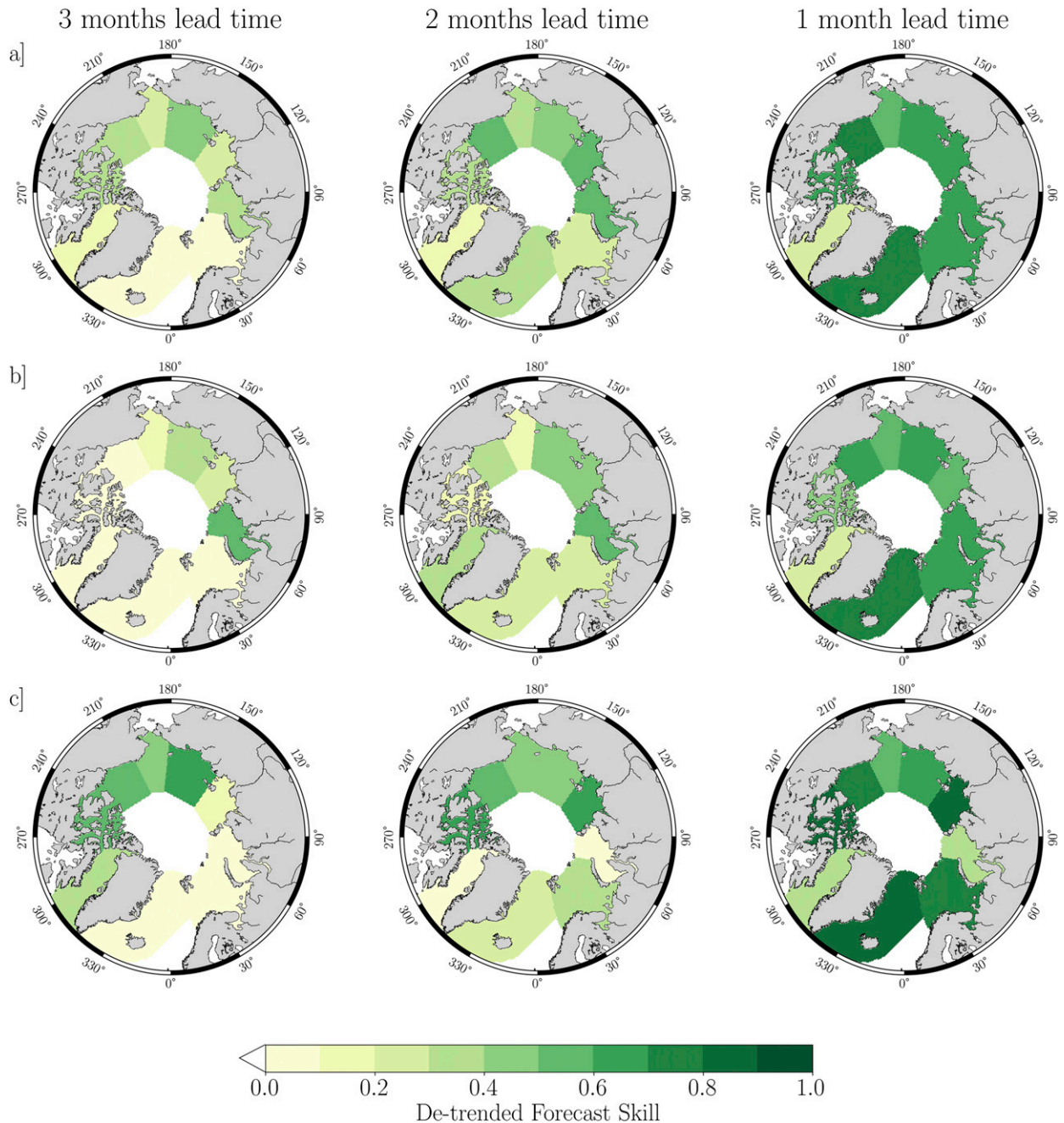
FIG. 6. Skill by Arctic region for predictions of detrended September SIE based on networks of monthly mean SIC data from June, July, and August (3-, 2-, and 1-month lead time, respectively). (a) Detrended skill calculated between 1985 and 2019. (b) Detrended skill calculated between 1985 and 2006. (c) Detrended skill calculated between 2007 and 2019. See Table 1 for values.

such as the location of the sea ice edge, and sea ice probability (SIP)—in line with stakeholder requirements and the SIPN2 mission statement. On this note, we have tested increasing the present methodology to 4-month lead time (May SIC data); however, the resultant detrended prediction skill is very poor for both the pan-Arctic and regional cases. This may be in part related to the spring predictability barrier that has been observed in model studies (Bushuk et al. 2017; Bonan et al. 2019); however, we have not investigated this in any detail as of yet. Irrespective of this, the methodology presented here can be advanced in order

TABLE 1. Detrended regional forecast skill values for predictions of September SIE based on monthly mean SIC data from: June, July, and August. Skill values from (11) are calculated between three periods: 1985–2019, 1985–2006, and 2007–19. Detrended ACC scores from (1) are also given for the period 1985–2019. Labels 1–9 correspond to area labels in Fig. 1.

Skill: 1985–2019, 1985–2006, 2007–2019 (ACC: 1985–2019)	June data	July data	August data
Pan-Arctic	0.53, 0.32, 0.65 (0.77)	0.62, 0.43, 0.73 (0.83)	0.81, 0.63, 0.91 (0.90)
Baffin Bay (1)	0.11, 0.07, 0.30 (0.35)	0.15, 0.38, −0.95 (0.51)	0.25, 0.24, 0.30 (0.51)
Greenland Sea (2)	−0.17, −0.09, −0.51 (−0.23)	0.30, 0.29, 0.30 (0.56)	0.74, 0.72, 0.82 (0.88)
Barents Sea (3)	0.10, 0.10, 0.07 (0.34)	0.29, 0.27, 0.31 (0.54)	0.69, 0.67, 0.77 (0.83)
Kara Sea (4)	0.40, 0.52, −1.25 (0.67)	0.53, 0.59, −0.44 (0.74)	0.67, 0.68, 0.35 (0.83)
Laptev Sea (5)	0.25, 0.22, 0.16 (0.52)	0.53, 0.46, 0.62 (0.73)	0.62, 0.52, 0.84 (0.80)
East Siberian Sea (6)	0.48, 0.34, 0.64 (0.70)	0.47, 0.48, 0.44 (0.70)	0.63, 0.62, 0.63 (0.82)
Chukchi Sea (7)	0.26, 0.12, 0.46 (0.54)	0.31, 0.18, 0.48 (0.65)	0.56, 0.53, 0.59 (0.79)
Beaufort Sea (8)	0.34, −0.07, 0.57 (0.71)	0.51, 0.36, 0.59 (0.77)	0.71, 0.63, 0.76 (0.86)
Canadian Archipelago (9)	0.33, 0.03, 0.57 (0.59)	0.40, 0.12, 0.62 (0.65)	0.66, 0.47, 0.82 (0.84)

to achieve improved sea ice predictability. In terms of the GPR model the prior covariance function can be adapted such that we move away from a strictly linear system and instead allow for fully nonparametric and nonlinear fits (Sollich and Halees 2002; Girard et al. 2003; Rasmussen and Williams 2006; Sarkar et al. 2019). Furthermore one could provide an alternative Gaussian prior than the typical uniform zero mean that is currently given to all network nodes. In the alternative case each node is assigned a different prior as necessary.

In terms of the network framework, this also has the ability to grow in terms of complexity. Currently the implementation is rather simple with a single-layer network of SIC being used as a predictor. With the incorporation of multilayer networks (Boccaletti et al. 2014) comes the ability to incorporate multiple climate variables within the regression forecasts. This may in turn lead to increasing the forecast lead time and skill.

6. Conclusions

Forecast submissions to the SIO project have shown that statistical models still have the ability to be competitive with their dynamical model counterparts, within the interannual time frame (Stroeve et al. 2014; Hamilton and Stroeve 2016). Although with apparent increasing temporal variability in the sea ice record (Holland et al. 2010) these statistical models need to be sufficiently complex in order to remain competitive. Not only this, the sea ice record shows significant spatial variability that can be seen across the observed SIE records (Fig. 5). Complex networks are a method that exploits such spatiotemporal variability for purposes ranging from improving understanding of climatological teleconnections, to time series forecasting; as presented here. Forecasts of pan-Arctic September SIE have shown to be successful

within a GPR forecast model and result in competitive detrended skill values with the literature to date: 0.53, 0.62, and 0.81 at 3-, 2-, and 1-month lead time, respectively. Regional forecasts made across the same time frame also show competitive detrended skill in the Canadian Archipelago, Beaufort, Chukchi, East Siberian, Laptev and Kara Seas, with skill values typically greater than 0.3. Improving forecast skill beyond 3 months requires development of the complex network and GPR methodology. The predictability of summer sea ice conditions using only concentration has been shown to diminish beyond 3 months (Drobot et al. 2006; Lindsay et al. 2008; Petty et al. 2017; Ionita et al. 2019). Extending the forecast window to 6 months or even to 1 year would require the incorporation of multiple climate variables (hence multilayered networks) to capture sources of predictability such as ocean heat advection effects on summer sea ice conditions (Serreze and Stroeve 2015) and possibly sea ice reemergence patterns (Blanchard-Wrigglesworth et al. 2011; Bushuk et al. 2014), among many others.

APPENDIX A

List of Variables

Table A1 contains a list of variables that appear in this paper.

TABLE A1. List of variables.

Symbol	Description
$x_p(t)$	The linearly detrended SIC anomaly data of the pth grid cell
A_i	The ith network node (area)
$\chi_i(t)$	The linearly detrended cumulative SIC anomaly data of the ith network node
r_{pq}	Linear Pearson correlation between linearly detrended SIC grid cells p and q
τ	The threshold factor required for grid cells to cluster to form a network node (area)
w_{ij}	Weighted link (temporal covariance) between network nodes i and j
λ_i	Strength of a given network node i
ψ_p	Area (km^2) of the pth SIC data grid cell
N	Number of network nodes
n	Length of SIC or SIE time series
T	Matrix transpose
$\mathbf{y}$	Linearly detrended SIE time series (pan-Arctic or regional)
$\mathbf{X}$	$n \times N$ design matrix containing linearly detrended cumulative SIC anomaly time series for all network nodes, also $\mathbf{X}(t)$
$\mathbf{Z}$	Vector of detrended network SIC anomalies used for prediction
$f(\mathbf{X})$	The function defining the Gaussian process
$\boldsymbol{\beta}$	Parameters of the GP function $f(\mathbf{X})$
ε	Gaussian noise, with variance σ^2
$(\ell, \alpha, \tilde{\sigma}^2)$	Hyperperameters
K	Covariance kernel of the Gaussian process function
$\boldsymbol{\Phi}$	Substituted variable for either $\mathbf{X}$ or $\mathbf{Z}$
$\boldsymbol{\Sigma}$	Covariance matrix of the Gaussian prior
$\mathbf{M}$	Stochastic matrix containing weighted links of all network nodes
$\mathbf{I}$	Identity matrix
$\bar{f}$	Mean of the predictive distribution (September SIE forecast value)
σ_f^2	Variance of the predictive distribution (error on the September SIE forecast)
S	Forecast skill
$\bar{y}$	Mean of the detrended September SIE time series
ω	Regularization coefficient of ridge regression

APPENDIX B

Gaussian Process Regression Theory

In Bayesian linear regression a functional relationship $f(\mathbf{X})$ which relates inputs $\mathbf{X}$ and outputs $\mathbf{y}$ can be expressed in the form $\mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ corresponds to a set of regression parameters. In this case, the aim of the model training is to derive a posterior distribution for the regression parameters $\boldsymbol{\beta}$ rather than point estimates as is done in classical statistical regressions. This is achieved through Bayes's Law by computing the likelihood function $p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta})$, which can be regarded as the probability of the observations for a given set of model

parameters $\boldsymbol{\beta}$, and combining this with a prior distribution over the model parameters themselves $p(\boldsymbol{\beta})$:

$$p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \frac{p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta})p(\boldsymbol{\beta})}{p(\mathbf{y}|\mathbf{X})}, \tag{B1}$$

where $p(\mathbf{y}|\mathbf{X})$ corresponds to the normalizing term, the marginal likelihood:

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta})p(\boldsymbol{\beta}) \, d\boldsymbol{\beta}. \tag{B2}$$

The Gaussian prior is a powerful tool in Bayesian inference. It incorporates belief about the behavior of $f(\mathbf{X})$ prior to any data analysis and plays an important role in the prevention of overfitting. Consider specifically a generic Gaussian prior $p(\boldsymbol{\beta})$ with zero mean and covariance matrix $\boldsymbol{\Sigma}$, such that $\boldsymbol{\beta} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$. This prior over $\boldsymbol{\beta}$ also provides a Gaussian prior over functions $f(\mathbf{X})$, and so our method fits into the broader scope of Gaussian process regression (GPR). Furthermore, as this prior over $f(\mathbf{X})$ has zero mean, it is fully specified by its covariance between the function values for two arbitrary inputs:

$$f(\boldsymbol{\Phi})f(\boldsymbol{\Phi}') = \boldsymbol{\Phi}^T \boldsymbol{\Sigma} \boldsymbol{\Phi}', \tag{B3}$$

where $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}'$ can be training inputs (columns of $\mathbf{X}^T$) or test inputs $\mathbf{Z}$. The prior covariance matrix between the function values for all training inputs is then simply $K(\mathbf{X}, \mathbf{X}) = \mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^T + \sigma^2\mathbf{I}$.

We can also see how, for this prior over $\boldsymbol{\beta}$, one obtains a posterior over $\boldsymbol{\beta}$ that is also Gaussian:

$$p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) \propto \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]$$
$$\times \exp\left(-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}\right), \tag{B4}$$

with mean $\tilde{\mu} = (\mathbf{X}^T\mathbf{X} + \sigma^2\boldsymbol{\Sigma}^{-1})^{-1}\mathbf{X}^T\mathbf{y}$. Here we notice that the maximization of the logarithm of (B4):

$$\ln p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$
$$-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta} + \text{const}, \tag{B5}$$

is equivalent to the minimization of the regularized least squares error function of *ridge regression* (Hoerl and Kennard 1970):

$$E_{\text{ridge}} = \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \frac{\omega}{2}\boldsymbol{\beta}^T\boldsymbol{\beta}, \tag{B6}$$

the analytical solution of which is given as $\boldsymbol{\beta}_{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \omega\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$. In ridge regression, it is ω which

minimizes overfitting by controlling the effectiveness of the quadratic penalty term $\omega\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}/2$. For large values of ω, $\boldsymbol{\beta}$ becomes very small; thus reducing the number of effective regression parameters. In our Bayesian approach the penalty term becomes $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}/(2\sigma^2)$; the effectiveness of which is controlled by the model hyperparameters, in our case (ℓ, α, $\tilde{\sigma}^2 = \sigma^2/\alpha$). A way to optimize these hyperparameters, and hence minimize overfitting, is through the empirical numerical optimization approach of type II maximum likelihood, which maximizes the normalized log of (B2):

$$\frac{1}{n}\ln p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2n}\mathbf{y}^{\mathrm{T}}K(\mathbf{X}, \mathbf{X})^{-1}\mathbf{y} - \frac{1}{2n}\ln|K(\mathbf{X}, \mathbf{X})|$$
$$- \frac{1}{2}\ln(2\pi), \qquad (B7)$$

Intuitively, this procedure chooses hyperparameters in such a way as to make the probability of the observations ($\mathbf{X}$, $\mathbf{y}$) under the model prior as large as possible, across all possible sets of hyperparameters. The maximization can also be performed using the gradients:

$$\frac{\partial \ln p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{2}\mathbf{y}^{\mathrm{T}}K(\mathbf{X}, \mathbf{X})^{-1}\frac{\partial K(\mathbf{X}, \mathbf{X})}{\partial \boldsymbol{\theta}}K(\mathbf{X}, \mathbf{X})^{-1}\mathbf{y}$$
$$- \frac{1}{2}\mathrm{tr}\left[K(\mathbf{X}, \mathbf{X})^{-1}\frac{\partial K(\mathbf{X}, \mathbf{X})}{\partial \boldsymbol{\theta}}\right]. \qquad (B8)$$

Once the optimal hyperparameters have been derived the GPR model is fully determined. Generating predictions then corresponds to averaging over all the regression parameters $\boldsymbol{\beta}$ which have been weighted by their posterior probability, for the given test inputs $\mathbf{Z}$, as given by (10).

REFERENCES

Bishop, C. M., 2006: *Pattern Recognition and Machine Learning*. Springer, 738 pp.

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

Boccaletti, S., and Coauthors, 2014: The structure and dynamics of multilayer networks. *Phys. Rep.*, **544**, 1–122, https://doi.org/10.1016/j.physrep.2014.07.001.

Boers, N., B. Bookhagen, H. M. J. Barbosa, N. Marwan, J. Kurths, and J. A. Marengo, 2014: Prediction of extreme floods in the eastern Central Andes based on a complex networks approach. *Nat. Commun.*, **5**, 5199, https://doi.org/10.1038/ncomms6199.

Bonan, D., M. Bushuk, and M. Winton, 2019: A spring barrier for regional predictions of summer Arctic sea ice. *Geophys. Res. Lett.*, **46**, 5937–5947, https://doi.org/10.1029/2019GL082947.

Bushuk, M., D. Giannakis, and A. J. Majda, 2014: Reemergence mechanisms for North Pacific sea ice revealed through nonlinear Laplacian spectral analysis. *J. Climate*, **27**, 6265–6287, https://doi.org/10.1175/JCLI-D-13-00256.1.

——, R. Msadek, M. Winton, G. A. Vecchi, R. Gudgel, A. Rosati, and X. Yang, 2017: Skillful regional prediction of Arctic sea ice on seasonal timescales. *Geophys. Res. Lett.*, **44**, 4953–4964, https://doi.org/10.1002/2017GL073155.

Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. J. Zwally, 1996: Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1. National Snow and Ice Data Center, accessed 11 July 2018, https://doi.org/10.5067/8GQ8LZQVL0VL.

Chevallier, M., D. Salas y Mélia, A. Voldoire, M. Déqué, and G. Garric, 2013: Seasonal forecasts of the pan-Arctic sea ice extent using a GCM-based seasonal prediction system. *J. Climate*, **26**, 6092–6104, https://doi.org/10.1175/JCLI-D-12-00612.1.

Comeau, D., D. Giannakis, Z. Zhao, and A. J. Majda, 2019: Predicting regional and pan-Arctic sea ice anomalies with kernel analog forecasting. *Climate Dyn.*, **52**, 5507–5525, https://doi.org/10.1007/s00382-018-4459-x.

Dijkstra, H. A., E. Hernández-García, C. Masoller, and M. Barreiro, 2019: *Networks in Climate*. Cambridge University Press, 282 pp.

Donges, J. F., Y. Zou, N. Marwan, and J. Kurths, 2009: Complex networks in climate dynamics: Comparing linear and nonlinear network construction methods. *Eur. Phys. J.: Spec. Top.*, **174**, 157–179, https://doi.org/10.1140/epjst/e2009-01098-2.

——, H. C. Schultz, N. Marwan, Y. Zou, and J. Kurths, 2011: Investigating the topology of interacting networks. *Eur. Phys. J.*, **84B**, 635–651, https://doi.org/10.1140/epjb/e2011-10795-8.

——, I. Petrova, A. Loew, N. Marwan, and J. Kurths, 2015: How complex climate networks complement eigen techniques for the statistical analysis of climatological data. *Climate Dyn.*, **45**, 2407–2424, https://doi.org/10.1007/s00382-015-2479-3.

Drobot, S. D., J. A. Maslanik, and C. Fowler, 2006: A long-range forecast of Arctic summer sea-ice minimum extent. *Geophys. Res. Lett.*, **33**, L10501, https://doi.org/10.1029/2006GL026216.

Eguíluz, V. M., J. Fernández-Gracia, X. Irigoien, and C. M. Duarte, 2016: A quantitative assessment of Arctic shipping in 2010–2014. *Sci. Rep.*, **6**, 30682, https://doi.org/10.1038/srep30682.

Fetterer, F., M. Savoie, S. Helfrich, and P. Clemente-Colón, 2010: Multisensor Analyzed Sea Ice Extent—Northern Hemisphere (MASIE-NH), version 1. National Snow and Ice Data Center, accessed 5 June 2018, https://doi.org/10.7265/N5GT5K3K.

Fountalis, I., A. Bracco, and C. Dovrolis, 2014: Spatio-temporal network analysis for studying climate patterns. *Climate Dyn.*, **42**, 879–899, https://doi.org/10.1007/s00382-013-1729-5.

——, C. Dovrolis, A. Bracco, B. Dilkina, and S. Keilholz, 2018: δ-MAPS: From spatio-temporal data to a weighted and lagged network between functional domains. *Appl. Network Sci.*, **3**, 21, https://doi.org/10.1007/S41109-018-0078-z.

Girard, A., C. E. Rasmussen, J. Q. Candela, and R. Murray-Smith, 2003: Gaussian process priors with uncertain inputs—Application to multiple-step ahead time series forecasting. *Advances in Neural Information Processing Systems*, S. Becker et al., Eds., MIT Press, 545–552.

Glantz, M. H., R. W. Katz, and N. Nicholls, 1991: *Teleconnections Linking Worldwide Climate Anomalies*. Cambridge University Press, 535 pp.

Guez, O., A. Gozolchiani, K. Yamasaki, Y. Berezin, S. Brenner, and S. Havlin, 2012: Climate network structure evolves with North Atlantic Oscillation phases. *EPL*, **98**, 38006, https://doi.org/10.1209/0295-5075/98/38006.

Hamilton, L. C., and J. Stroeve, 2016: 400 predictions: The SEARCH Sea Ice Outlook 2008–2015. *Polar Geogr.*, **39**, 274–287, https://doi.org/10.1080/1088937X.2016.1234518.

Hoerl, A. E., and R. W. Kennard, 1970: Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–67, https://doi.org/10.1080/00401706.1970.10488634.

Holland, M. M., D. A. Bailey, and S. Vavrus, 2010: Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model, version 3. *Climate Dyn.*, **36**, 1239–1253, https://doi.org/10.1007/s00382-010-0792-4.

Ionita, M., K. Grosfeld, P. Scholz, R. Treffeisen, and G. Lohmann, 2019: September Arctic sea ice minimum prediction—A skillful new statistical approach. *Earth Syst. Dyn.*, **10**, 189–203, https://doi.org/10.5194/esd-10-189-2019.

Kondor, R. I., and J. Lafferty, 2002: Diffusion kernels on graphs and other discrete structures. *Proc. 19th Int. Conf. on Machine Learning*, San Francisco, CA, Morgan Kaufmann, 315–322.

Lee, S.-W., and J.-M. Song, 2014: Economic possibilities of shipping though Northern Sea route. *Asian J. Shipp. Logist.*, **30**, 415–430, https://doi.org/10.1016/j.ajsl.2014.12.009.

Lindsay, R., J. Zhang, A. Schweiger, and M. Steele, 2008: Seasonal predictions of ice extent in the Arctic Ocean. *J. Geophys. Res.*, **113**, C02023, https://doi.org/10.1029/2007JC004259.

Maslanik, J., and J. Stroeve, 1999: Near-real-time DMSP SSMIS daily polar gridded sea ice concentrations, version 1 (updated daily). National Snow and Ice Data Center, accessed 27 November 2019, https://doi.org/10.5067/U8C09DWVX9LM.

Msadek, R., G. A. Vecchi, M. Winton, and R. Gudgel, 2014: Importance of initial conditions in seasonal predictions of Arctic sea ice extent. *Geophys. Res. Lett.*, **41**, 5208–5215, https://doi.org/10.1002/2014GL060799.

Notz, D., and J. Stroeve, 2016: Observed Arctic sea-ice loss directly follows anthropogenic CO_2 emission. *Science*, **354**, 747–750, https://doi.org/10.1126/science.aag2345.

Onarheim, I. H., T. Eldevik, L. H. Smedsrud, and J. C. Stroeve, 2018: Seasonal and regional manifestation of Arctic sea ice loss. *J. Climate*, **31**, 4917–4932, https://doi.org/10.1175/JCLI-D-17-0427.1.

Overland, J. E., and M. Wang, 2013: When will the summer Arctic be nearly sea ice free? *Geophys. Res. Lett.*, **40**, 2097–2101, https://doi.org/10.1002/grl.50316.

Peterson, K. A., A. Arribas, H. Hewitt, A. Keen, D. Lea, and A. McLaren, 2015: Assessing the forecast skill of Arctic sea ice extent in the GloSea4 seasonal prediction system. *Climate Dyn.*, **44**, 147–162, https://doi.org/10.1007/s00382-014-2190-9.

Petty, A. A., D. Schröder, J. C. Stroeve, T. Markus, J. Miller, N. T. Kurtz, D. L. Feltham, and D. Flocco, 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.

Radebach, A., R. V. Donner, J. Runge, J. F. Donges, and J. Kurths, 2013: Disentangling different types of El Niño episodes by evolving climate network analysis. *Phys. Rev. E-Stat. Nonlinear Soft Matter Phys.*, **88**, 052807, https://doi.org/10.1103/PhysRevE.88.052807.

Rasmussen, C. E., and C. K. I. Williams, 2006: *Gaussian Processes for Machine Learning*. MIT Press, 272 pp.

Sarkar, D., M. A. Osborne, and T. A. Adcock, 2019: Spatiotemporal prediction of tidal currents using Gaussian processes. *J. Geophys. Res. Oceans*, **124**, 2697–2715, https://doi.org/10.1029/2018JC014471.

Schröder, D., D. L. Feltham, D. Flocco, and M. Tsamados, 2014: September Arctic sea-ice minimum predicted by spring melt-pond fraction. *Nat. Climate Change*, **4**, 353–357, https://doi.org/10.1038/nclimate2203.

Serreze, M. C., and J. Stroeve, 2015: Arctic sea ice trends, variability and implications for seasonal ice forecasting. *Philos. Trans. Roy. Soc.*, **373A**, 20140159, https://doi.org/10.1098/RSTA.2014.0159.

Sigmond, M., J. Fyfe, G. Flato, V. Kharin, and W. Merryfield, 2013: Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system. *Geophys. Res. Lett.*, **40**, 529–534, https://doi.org/10.1002/grl.50129.

Smola, A. J., and R. Kondor, 2003: Kernels and regularization on graphs. *Learning Theory and Kernel Machines*, B. Schölkopf and M. K. Warmuth, Eds., Springer, 144–158.

Sollich, P., 2005: Can Gaussian process regression be made robust against model mismatch? *Deterministic and Statistical Methods in Machine Learning*, J. Winkler, Eds., Springer, 199–210.

——, and A. Halees, 2002: Learning curves for Gaussian process regression: Approximations and bounds. *Neural Comput.*, **14**, 1393–1428, https://doi.org/10.1162/089976602753712990.

Spielman, D. A., 2010: Algorithms, graph theory, and linear equations in Laplacian matrices. *Proc. Int. Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures,* Hyderabad, Telangana, World Scientific, 2698–2722.

Stroeve, J., and D. Notz, 2018: Changing state of Arctic sea ice across all seasons. *Environ. Res. Lett.*, **13**, 103001, https://doi.org/10.1088/1748-9326/aade56.

——, L. C. Hamilton, C. M. Bitz, and E. Blanchard-Wrigglesworth, 2014: Predicting September sea ice: Ensemble skill of the SEARCH Sea Ice Outlook 2008–2013. *Geophys. Res. Lett.*, **41**, 2411–2418, https://doi.org/10.1002/2014GL059388.

——, A. D. Crawford, and S. Stammerjohn, 2016: Using timing of ice retreat to predict timing of fall freeze-up in the Arctic. *Geophys. Res. Lett.*, **43**, 6332–6340, https://doi.org/10.1002/2016GL069314.

Tsonis, A. A., and P. J. Roebber, 2004: The architecture of the climate network. *Physica A*, **333**, 497–504, https://doi.org/10.1016/j.physa.2003.10.045.

Urry, M. J., and P. Sollich, 2013: Random walk kernels and learning curves for Gaussian process regression on random graphs. *J. Mach. Learn. Res.*, **14**, 1801–1835.

Vishwanathan, S. V. N., N. N. Schraudolph, R. Kondor, and K. M. Borgwardt, 2010: Graph kernels. *J. Mach. Learn. Res.*, **11**, 1201–1242.

Walker, G., and E. W. Bliss, 1932: World weather V. *Mem. Roy. Meteor. Soc.*, **4**, 53–84.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812, https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

Wang, W., M. Chen, and A. Kumar, 2013: Seasonal prediction of Arctic sea ice extent from a coupled dynamical forecast system. *Mon. Wea. Rev.*, **141**, 1375–1394, https://doi.org/10.1175/MWR-D-12-00057.1.

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