Noname manuscript No.
(will be inserted by the editor)

Predicting regional and pan-Arctic sea ice anomalies with kernel analog forecasting

Darin Comeau · Dimitrios Giannakis · Zhizhen Zhao · Andrew J. Majda

Received: date / Accepted: date

Abstract Predicting the Arctic sea ice extent is a notoriously difficult forecasting problem, even from lead times as short as one month. Motivated by Arctic intraannual variability phenomena such as sea surface temperature reemergence and sea ice anomaly reemergence, we use a prediction approach for sea ice anomalies based on analog forecasting. Traditional analog forecasting relies on identifying a single analog in a historical record, usually by minimizing Euclidean distance, and forming a forecast from the analog's historical trajectory. We use an ensemble of analogs for our forecasts, where the ensemble weights are determined by a dynamics-adapted similarity kernel, which takes into account the nonlinear geometry on the underlying data manifold. We apply this method for forecasting regional and pan-Arctic sea ice concentration and volume anomalies from multicentury climate model data, and in many cases find improvement over the damped persistence forecast. Moreover the patterns of predictive skill we see by region and season are consistent with different types of sea ice anomaly reemergence.

D. Comeau
Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY.
E-mail: dcomeau@lanl.gov *Present address:* of D. Comeau: Climate, Ocean, and Sea Ice Modeling Group, Computational Physics and Methods Group (CCS-2), Los Alamos National Laboratory, Los Alamos, NM.

D. Giannakis
Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY.

Z. Zhao
Department of Electrical and Computer Engineering, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL.

A. Majda
Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY.

1 Introduction

Predicting the climate state of the Arctic, particularly with regards to sea ice extent, has been a subject of increased recent interest in part driven by record-breaking minimums in September sea ice extent in 2007 and again in 2012. As new areas of the Arctic become accessible, this has increasingly become an important practical problem in addition to a scientific one, e.g., navigating shipping routes (Smith and Stephenson, 2013). Many different approaches have been used recently to address Arctic sea ice prediction, including statistical frameworks (Lindsay et al, 2008; Wang et al, 2015), through inherent predictability within general circulation models (GCMs) (Blanchard-Wrigglesworth et al, 2011a,b; Chevallier et al, 2013; Tietsche et al, 2013, 2014; Day et al, 2014), and using dynamical models to predict observations (Zhang et al, 2008; Sigmond et al, 2013; Wang et al, 2013). These methods have varying degrees of success in predicting pan-Arctic and regional sea ice area or extent (area with at least 15% sea ice concentration) and to a lesser degree sea ice volume. Indeed, in sea ice prediction, current generation numerical models and data assimilation systems have little additional skill beyond simple persistence or damped persistence forecasts (Blanchard-Wrigglesworth et al, 2015).

Following the 2007 September sea ice extent minimum, the Study of Environmental Arctic Change (SEARCH) began soliciting forecasts of September sea ice extent for the Sea Ice Outlook (SIO), which since 2013 has been handled by the Sea Ice Prediction Network (SIPN). They have found that year to year variability, rather than methods, dominate the ensemble's success, and that extreme years are in general less predictable (Stroeve et al, 2014). The forecasts, given at one to three lead month times, had particular difficulty with the record low extent of September 2012 and the subsequent September 2013, which saw a partial recovery in sea ice extent. A more recent study of SIO model forecasts by

Blanchard-Wrigglesworth et al (2015) highlighted the importance of model uncertainty on predictability by performing an initial condition perturbation experiment and finding wide spread among models' response.

Accurately predicting aspects of Arctic sea ice is made difficult by a number of factors, notably that the mean state of the Arctic is changing (Stroeve et al, 2014; Guemas et al, 2014). Variability of observed sea ice area has increased, in part due to thinner ice being more variable (in both thickness and extent), which is hypothesized to have lower predictability (Blanchard-Wrigglesworth et al, 2011a; Holland et al, 2011, 2008). Since the satellite record began in 1979, all months have a downward trend in sea ice extent, the largest being for September (Stroeve et al, 2012). Moreover, as thicker multiyear ice is replaced by thinner, younger ice, the trends steepen (Stroeve et al, 2012). Ice thickness data is seen as offering key predictive information for sea ice area / extent (Bushuk et al, 2017; Blanchard-Wrigglesworth and Bitz, 2014; Chevallier and Salas-Mélia, 2012; Lindsay et al, 2008; Wang et al, 2013), but such observational data sets do not yet exist in uniform spatial and temporal coverage, although some observations are available from the CryoSat-2 satellite (Tilling, 2016), and the Pan-Arctic Ice Ocean Modeling and Assimilation System (Zhang and Rothrock, 2003), which produces sea ice volume data by assimilating observations of sea ice concentration with a sea ice thickness model. Ice age, in particular area of ice of a certain age, is also seen as an important predictor, also of which there is no reliable observational record (Stroeve et al, 2012). Statistical predictions based on historical relationships in particular have difficulty in predicting sea ice due in a changing Arctic mean state (Holland and Stroeve, 2011; Stroeve et al, 2014).

There are both dynamic and thermodynamic elements that factor into sea ice predictability. Locally, ice thickness predictability in the Arctic is dominated by dynamic, rather than thermodynamic properties (Blanchard-Wrigglesworth and Bitz, 2014; Tietsche et al, 2014). On the other hand, limits on September sea ice extent are primarily thermodynamic (related to amount of open water formation in melt season), whereas dynamic induced anomalies have smaller influence, except in a thin ice regime (Holland et al, 2011). Improvement in melt-pond parameterizations in the sea ice model Community Ice CodE (CICE) (Holland et al, 2012) has yielded skill in predicting September sea ice extent (Schröder et al, 2014), demonstrating potential predictive yield in improving process models.

Chaotic atmosphere variability also places an inherent, and perhaps dominant, limit on sea ice predictability, both through dynamic effects with redistribution of sea ice (Day et al, 2014; Holland et al, 2011; Blanchard-Wrigglesworth et al, 2011b; Ogi et al, 2010), and thermodynamic effects on ocean conditions (Tietsche et al, 2016). While correlations between the Arctic Oscillation (AO) and summer ice extent historically have been high (Rigor et al, 2002; Rigor and Wallace, 2004) these correlations have become weaker in recent years as the sign of the AO has changed while sea ice extent has continued to decline, suggesting that this may not be a very predictive atmospheric metric for sea ice (Holland and Stroeve, 2011; Ogi et al, 2010). Nevertheless, summer atmospheric conditions have a strong impact sea ice extent, particularly for thinning sea ice, and may contain more predictive skill than sea ice thickness in terms of predicting the September ice extent minimum (Stroeve et al, 2014). The ocean temperature at depth has also been found to be an important predictor factor for sea ice extent (Lindsay et al, 2008).

The problem of sea ice prediction becomes both of more practical use, while becoming more difficult, as we move from the pan-Arctic to regional scale, where local ice advection across regional boundaries and small scale influences on sea ice processes become important (Blanchard-Wrigglesworth and Bitz, 2014). Certain regions have been found to be more predictable than others; in terms of extent, basins adjacent to the Atlantic (Labrador to Barents) and other seasonal ice zones are more predictable than central Arctic basins (Day et al, 2014; Lindsay et al, 2008; Koenigk and Mikolajewicz, 2009), whereas for thickness, persistence in the central Arctic basins leads to higher predictability (Day et al, 2014) than other regions. In addition to the September sea ice extent metric, there has been increased focus on predicting regional sea ice advance and retreat dates (e.g. Sigmond et al (2016)), which are now included as part of the SIO solicitation. Seasonality also plays a strong role in predictability, e.g. SST conditions in the North Atlantic can lead to high predictability of winter sea ice extent (Yeager et al, 2015), whereas the summer melt season provides a barrier to predictability, limiting the skill of forecasts initialized in late spring (Day et al, 2014).

Another aspect of sea ice variability with a seasonal dependence is sea ice anomaly reemergence, a phenomenon where anomalies at one time reappear several months later, made evident by high lagged correlations following low correlations at shorter time lags. This has been found in both models and observations (Blanchard-Wrigglesworth et al, 2011a). Reemergence phenomena fall into two categories; one where anomalies from a melt season reemerge in the subsequent growth season, typically found in marginal ice zones, governed by ocean and large-scale atmospheric conditions, and another where anomalies from a growth season reemerge in the subsequent melt season, typically found in central Arctic regions that exhibit perennial sea ice, driven by sea ice thickness (Blanchard-Wrigglesworth et al, 2011a; Bushuk et al, 2014; Bushuk and Giannakis, 2015; Bushuk et al, 2015; Bushuk and Giannakis, 2017). These observed phenomena provide a promising source of sea ice predictabil-

ity, which Day et al (2014) found to be robust in several GCMs.

The timescales of Arctic sea ice predictability vary across studies, depending on the measure of forecast skill and the target month of prediction (among other factors), but generally fall in the 3–6 month range (Blanchard-Wrigglesworth et al, 2015; Guemas et al, 2014; Tietsche et al, 2014; Chevallier and Salas-Mélia, 2012), with potential predictability extending to 2-3 years (Tietsche et al, 2014; Blanchard-Wrigglesworth et al, 2011b). While Lindsay et al (2008) found that most predictive information in the ice-ocean system is lost for lead times greater than 3 months, Blanchard-Wrigglesworth et al (2011a) found pan-Arctic sea ice area predictable for 1–2 years, and sea ice volume up to 3–4 years, in a perfect model framework. It has been found that predicting the state of sea ice in the spring is particularly difficult, with most of the predictive skill coming from fall persistence (Wang et al, 2013; Holland et al, 2011), and that March sea ice extent is largely uncorrelated with the following September sea ice extent (detrended) (Blanchard-Wrigglesworth et al, 2011a; Stroeve et al, 2014).

While many forecast methods use expensive dynamical models, there is promise in utilizing nonparametric statistical techniques. In particular, analog forecasting is an idea dating back to Lorenz (1969), where a prediction is made by identifying an appropriate historical analog to a given initial state, and using the analog's trajectory in the historical record to make a forecast of the present state. While this is attractive as a fully non-parametric, data-driven approach, a drawback of traditional analog forecasting is that it relies upon a single analog, usually identified by Euclidean distance, possibly introducing highly discontinuous behavior into the forecasting scheme. This can be improved upon by selecting an ensemble of analogs, and taking a weighted average of the associated trajectories. Another potential drawback is that a large number of historical data may be needed in order to identify appropriate analogs (Van den Dool, 1994). Analog forecasting has been used in numerous climate applications (Drosdowsky, 1994; Xavier and Goswami, 2007; Alessandrini et al, 2015; Atencia and Zawadzki, 2015), the latter of which employed an ensemble approach. Given there are sources of sea ice predictability from the ocean, atmosphere, and sea ice itself (Guemas et al, 2014), a data-driven approach such as analog forecasting may be able to exploit complex coupled-system dynamics encoded in GCM data and provide skill in such a prediction problem.

In Zhao and Giannakis (2016) this idea was extended upon by assigning ensemble weights derived from a dynamics-adapted kernel, constructed in such a way as to give preferential weight to states with similar dynamics, referred to as kernel analog forecasting (KAF). Modes of variability intrinsic to the data analysis, as eigenfunctions of the kernel operator, are extracted with clean timescale separation and inherent predictability through the encoding of dynamic information in the analysis. KAF has been used in forecasting modes representing the Pacific Decadal Oscillation (PDO) and North Pacific Gyre Oscillation (NPGO) (Zhao and Giannakis, 2016), in which cases it was shown to be more skillful than parametric regression forecasting methods (Comeau et al, 2017). More recently, KAF has been used in forecasting variability in the tropics by the Madden-Julian oscillation and the boreal summer intraseasonal oscillation (Alexander et al, 2017).

While KAF exhibits predictive skill in these extracted modes intrinsic to the data analysis, that being eigenfunctions of the kernel operator, it is also desirable to have skill in forecasting quantities that are independent of the analysis approach and can be physically observed, e.g. Arctic sea ice extent anomalies (Comeau et al, 2017). The aim of this study is to extend upon Comeau et al (2017) and assess the skill of KAF in predicting Arctic sea ice anomalies on various spatial and temporal scales in order to identify where and when we may (or may not) have predictive skill. Since, as with every statistical technique, the utility of KAF depends upon the availability of an appropriately rich training record, we examine predictability in a long control run of a coupled GCM to establish a baseline of KAF predictive skill in predicting the internal dynamics attributed to natural variability. This allows us to estimate the upper limits of skill with this method in a statistically robust manner. We consider various combinations of predictor variables including sea ice concentration (SIC), sea surface temperature (SST), sea ice thickness (SIT), and sea level pressure (SLP) data to assess which variables contain the most useful predictive information. The important consideration of statistical prediction in the presence of a changing climate is beyond the scope of this work, though in Sect. 5 we make suggestions of how to address this issue.

The rest of this paper is structured as follows: The KAF method is described in Sect. 2. The data and experimental setup is described in Sect. 3, with the associated results in Sect. 4. Discussion and concluding remarks are given in Sect. 5.

2 Methods

The KAF method (Zhao and Giannakis, 2016; Comeau et al, 2017; Alexander et al, 2017), is designed to address the difficult task of prediction using massive data sets sampled from a complex nonlinear dynamical system in a very large state space. The motivating idea is to encode information from the underlying dynamics of the system into a kernel function, an exponentially decaying pairwise measure of similarity that plays an analogous role to the covariance operator in principal components analysis. At the outset, during the training phase we have access to an n sample size time-ordered

training data set $\{x_1,\ldots,x_n\}$ and the corresponding values $\{f_1,\ldots,f_n\}$ of a prediction observable. In our applications, the target observable is the aggregate sea ice anomaly in extent or volume over some region, and the training data are monthly averaged gridded climate variables. The main steps in KAF, outlined in detail below, are 1) perform Takens embedding of the data, 2) evaluate a dynamics-adapted similarity kernel on the embedded data, and 3) use weights from this kernel to make a forecast of an observable via out-of-sample extension formed by a weighted iterated sum.

2.1 Takens embedding

The first step in our analysis is to construct a new state variable through time-lagged embedding. For an embedding window of length q, which will depend on the time scale of our observable of interest, and a spatiotemporal series $z_1, z_2, \ldots z_n$ with $z_i \in \mathbb{R}^d$, where d is the number of spatial (grid) points and i is the time index, we form a data set of x_i in lagged-embedded space (also called Takens embedding space) by

$$x_i = \left(z_i, z_{i-1}, \ldots, z_{i-(q-1)}\right) \in \mathbb{R}^{qd}.$$

The utility of this embedding is that it recovers the topology of the attractor of the underlying dynamical system through partial observations (the z_i's) (Packard et al, 1980; Takens, 1981; Broomhead and King, 1986; Sauer et al, 1991; Robinson, 2005; Deyle and Sugihara, 2011). The choice of the embedding window q should be chosen long enough to capture the time-scales of interest, but not so long as to reduce the discriminating power of the kernel in determining locality.

2.2 Dynamics-adapted kernels

The collection of data points $\{x_1,\ldots,x_n\}$ can be thought of as lying on a manifold nonlinearly embedded in data space $\mathbb{R}^{qd}$. We will endow this manifold with a geometry (i.e., a means of measuring distances and angles) through a kernel function of data similarity. The kernel function we use for that purpose is from the Nonlinear Laplacian Spectral Analysis (NLSA) algorithm (Giannakis and Majda, 2012a,b, 2013, 2014). The kernel incorporates additional dynamic information by using phase velocities $\xi_i = \|x_i - x_{i-1}\|$, thus giving higher weight to regions of data space where the data is changing rapidly (see Giannakis (2015) for a geometrical interpretation), and is:

$$k(x_i,x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\varepsilon \xi_i \xi_j}\right),$$

where ε is a scale parameter. We modify this to include multiple variables $x_i = \left(x_i^{(1)}, x_i^{(2)}\right)$ (Bushuk et al, 2014), possibly of different physical units, embedding windows, or grid points, which for two variables is

$$k(x_i,x_j) = \exp\left(-\frac{\|x_i^{(1)} - x_j^{(1)}\|^2}{\varepsilon^{(1)} \xi_i^{(1)} \xi_j^{(1)}} - \frac{\|x_i^{(2)} - x_j^{(2)}\|^2}{\varepsilon^{(2)} \xi_i^{(2)} \xi_j^{(2)}}\right), \quad (1)$$

and extended to more than two variables in a similar manner. While in principle different scale parameters ε may be used for different variables to assign relative weights between variables within the kernel, in this analysis we use the same scale parameter for all variables. The exponential decay of the kernel in Eq. (1) means that very dissimilar states will carry negligible weight in our construction of historical analogs. In practice we enhance this localizing behavior further by setting small entries of k to zero. The number of weights retained is then the ensemble size of historical analogs used to make a forecast. Finally, we next form row-normalized kernels,

$$P(x_i,x_j) = \frac{k(x_i,x_j)}{\sum_{l=1}^n k(x_i,x_l)}, \quad (2)$$

which forms a row-stochastic matrix P that allows us to interpret each row as an empirical probability distribution of the second argument that depends on the data point in the first argument.

2.3 Out-of-sample extension via Laplacian pyramids

Our approach of assigning a value for a function f defined on a training data set X to a new test value $y \notin X$ will be through an out-of-sample extension technique known as Laplacian pyramids (Rabin and Coifman, 2012). In our context, the training data will be a spatiotemporal data set comprised of (lagged-embedded) state vectors x_i of gridded state variables (e.g. some combination of SIC, SST, SLP, and SIT), $f_i = f(x_i)$ is the function that gives us the sea ice cover anomaly of the state x_i, and y will be a new state vector (in lagged-embedded space), from which we would like to make a forecast of future sea ice cover anomalies.

We define a family of kernels P_l by modifying the NLSA kernels k in Eq. (2) to have scale parameter of the form $\sigma_0/2^l$ rather than ε, which we denote k_l, and then P_l is the row-sum normalized k_l, as in Eq. (2). This forms a multiscale family of kernels with increasing dyadic resolution in l with a shape parameter σ_0. A function f is approximated in a multiscale manner as an iterated weighted sum by $f \approx s_0 + s_1 + s_2 + \cdots$, where the first level s_0 and difference d_1 is defined by

$$s_0(x_k) = \sum_{i=1}^{n} P_0(x_i, x_k) f(x_i), \qquad d_1(x_i) = f(x_i) - s_0(x_i),$$

and the lth level decomposition s_l is defined iteratively:

$$s_l(x_k) = \sum_{i=1}^{n} P_l(x_i, x_k) d_l(x_i), \qquad d_l(x_i) = f(x_i) - \sum_{i=0}^{l-1} s_i(x_i).$$

For the choice of kernels k_l, increasing l can lead to overfitting, which we mitigate by zeroing out the diagonals of the kernels (Fernández et al, 2013). We set the truncation level for the iterations at level L once the approximation error begins to increase in l. Given a new data point y, we extend f by

$$\bar{s}_0(y) = \sum_{i=1}^{n} P_0(y, x_i) f(x_i), \qquad \bar{s}_l(y) = \sum_{i=1}^{n} P_l(y, x_i) d_l(x_i),$$

for $L \geq 1$, and assign f the value

$$\bar{f}(y) = \sum_{l=0}^{L} \bar{s}_l(y). \tag{3}$$

That is, we use the kernels P_l to evaluate the similarity of y to points x_i in the training data, and use this measure of similarity to form a weighted average of $f(x_i)$ values to define $\bar{f}(y)$. Note that there will be some reconstruction error between the out-of-sample extension value $\bar{f}(y)$ and the ground truth $f(y)$, which in general is not known.

2.4 Kernel Analog Forecast

Recall that in traditional analog forecasting, a forecast is made by identifying a single historical analog that is most similar to the given initial state, and using the historical analog's trajectory as the forecast. As mentioned in Sect. 2.2, it is convenient to think of rows of normalized kernels as empirical probability distributions in the second argument, $p_y(x) = P(y, x)$. In this setting, we can then consider taking weighted sums as taking an expectation ($\mathbb{E}$) over a probability distribution. As an example the traditional analog forecast $\bar{f}_{TAF}$ for a lead time τ can be written as

$$\bar{f}_{TAF}(y, \tau) = \mathbb{E}_{p_y} S_\tau f = \sum_{i=1}^{n} p_y(x_i) f(x_{i+\tau}) = f(x_{j+\tau}),$$

where $p_y = \delta_{iy}$ is the Kronecker delta distribution and $S_\tau f(x_i) = f(x_{i+\tau})$ is the shift operator. Note that $S_\tau f(x_i)$ can be evaluated since we know the time-shifted value of f exactly over the training data set.

Given a new state y, we define our prediction $\bar{f}(y, \tau)$ for lead time τ, via Laplacian pyramids, by

$$\bar{f}(y, \tau) = \mathbb{E}_{p_{y,0}} S_\tau f + \sum_{l=1}^{L} \mathbb{E}_{p_{y,l}} S_\tau d_l, \tag{4}$$

where $p_{y,l}(x) = P_l(y, x)$ corresponds to the probability distribution from the kernel at scale l. Note that when $\tau = 0$, Eq. (4) reduces to the Laplacian pyramid out-of-sample extension expression for $\bar{f}(y)$ in Eq. (3).

The reconstruction error from the out-of-sample extension manifests itself in the fidelity of the forecasts as the error at time lag 0. While in our applications, knowing the climate state y allows us to compute the observable $f(y)$ exactly at time lag 0, we need the out-of-sample extension to compute the predicted observable $\bar{f}(y, \tau)$ at any time lag $\tau > 0$. Hence we must contend with the initial reconstruction error, which is the difference between $f(y)$ and $\bar{f}(y, 0)$. This will impact forecasts for all time lags.

2.5 Error Metrics

For the purposes of defining the error metrics for predictions, let $F(y_j, \tau)$ be a general prediction of an observable f of state y_j at lead time τ, with $f(y_{j+\tau})$ being the true value. We evaluate the performance of predictions with two aggregate error metrics, the root-mean-square error (RMSE) and pattern correlation (PC), defined as

$$\mathrm{RMSE}(\tau) = \sqrt{\frac{1}{n'} \sum_{j=1}^{n'} \left(F(y_j, \tau) - f(y_{j+\tau})\right)^2},$$

$$\mathrm{PC}(\tau) = \frac{1}{n'} \sum_{j=1}^{n'} \frac{\left(F(y_j, \tau) - \tilde{F}(y, \tau)\right)\left(f(y_{j+\tau}) - \tilde{f}(y, \tau)\right)}{\sigma_{\tilde{F}(y,\tau)} \sigma_{\tilde{f}(y,\tau)}},$$

where

$$\tilde{F}(y, \tau) = \frac{1}{n'} \sum_{j=1}^{n'} F(y_j, \tau), \qquad \tilde{f}(y, \tau) = \frac{1}{n'} \sum_{j=1}^{n'} f(y_{j+\tau}),$$

$$\sigma^2_{\tilde{F}(y,\tau)} = \frac{1}{n'} \sum_{j=1}^{n'} \left(F(y_j, \tau) - \tilde{F}(y, \tau)\right)^2,$$

$$\sigma^2_{\tilde{f}(y,\tau)} = \frac{1}{n'} \sum_{j=1}^{n'} \left(f(y_{j+\tau}) - \tilde{f}(y, \tau)\right)^2.$$

The averaging is over predictions formed from using testing data of length n' (second portion of the data set) as initial conditions. KAF error metrics are evaluated with the predictions $F(y, \tau) = \bar{f}(y, \tau)$ (as defined in Eq. (4)). We use error metrics for the damped persistence forecast $F(y_j, \tau) =$

$\beta^\tau f(y_j)$, where β is the lag-1 autocorrelation coefficient of f as our benchmark, and use a threshold of 0.5 in pattern correlation score, below which predictive skill is considered low (Germe et al, 2014). Given our interest in high (> 0.5) pattern correlation and a large number of samples (≈ 4500), the correlations considered are statistically significant. RMSE scores are normalized by the truth (NRMSE) in the figures that follow.

3 Datasets

We use monthly averaged CCSM4 (Gent et al, 2011) model data from a pre-industrial control simulation (b40.1980), where 800 years of the simulation are split into a training dataset and a test dataset, 400 years each. The sea ice component is CICE4 (Hunke and Lipscomb, 2008), the ocean component is POP (Smith et al, 2010), and the atmosphere component is CAM4 (Neale et al, 2010). Our default experimental setup is to include SIC, SST, and SLP fields, and we will later explore the role of SIT as an additional predictor variable. We consider the entire Arctic, as well as the following regions: Beaufort Sea, Chukchi Sea, East Siberian Sea, Laptev Sea, Kara Sea, Barents Sea, Greenland Sea, Baffin Bay, Labrador Sea, Bering Sea, and Sea of Okhotsk. While the ice and ocean state variables are restricted to each region, pan-Arctic SLP data is always used, to allow for possible teleconnection effects. The regions are depicted in Fig. 1, shown with this dataset's sea ice concentration variability calculated over the entire control run. While a pre-industrial control simulation is not indicative of our current transient climate, our objective is to establish a baseline of performance for KAF in predicting sea ice anomalies by making use of a large training data set of a climate without a secular trend, so that useful historical analogs may be identified and predictive skill robustly assessed. The Arctic sea ice anomalies form this dataset exhibit inter-annual variability, but no drift.

Our target observable f for prediction is integrated anomalies in sea ice area (as opposed to sea ice extent which is sea ice area above 15% concentration). Sea ice anomalies in the test data period are calculated relative to the monthly climatology calculated from the training data set. While this should not be a concern in a pre-industrial control run with no secular trend, this may be of more importance in other scenarios. Persistence forecasts are initialized with the true anomaly (as opposed to the out-of-sample extension value), so all forecasts will have initial error metrics greater than damped persistence due to reconstruction error.

We consider combinations of SIC with SST, SIT, and SLP as predictor variables. While adding more variables, and thereby increasing the domain size and including more physics, should not result in the reduction of skill, in practice it may result in a loss of discriminating power of the kernel. A balance needs to be considered between the inclusion of variables that add more physics to the training data, and the ability of KAF to leverage this information in discerning useful historical analogs.

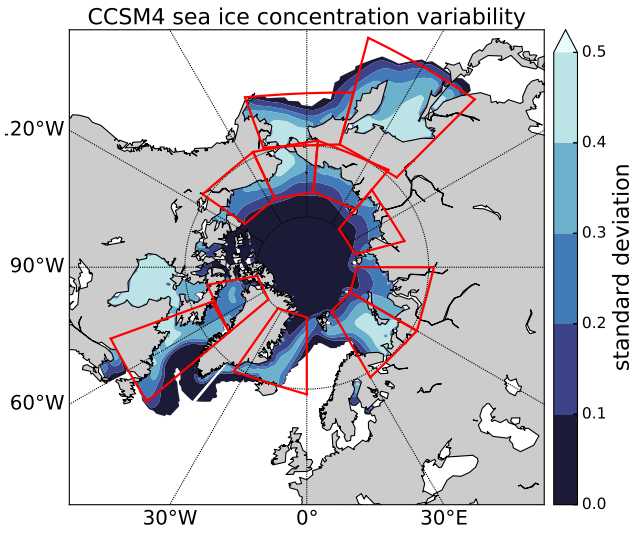

Fig. 1 Standard deviation of monthly sea ice concentration (SIC) from the CCSM4 control run, with regions considered in our forecasting are pan-Arctic (45N–90N), Beaufort Sea (125W–155W, 65N–75N), Chukchi Sea (155W–175E, 65N–75N), East Siberian Sea (175E–140E, 65N–75N), Laptev Sea (140E–105E, 70N–80N), Kara Sea (60E–90E, 65N–80N), Barents Sea (30E–60E, 65N–80N), Greenland Sea (35W–0E, 65N–80N), Baffin Bay (80W–50W, 70N–80N), Labrador Sea (70W–50W, 50–70N), Bering Sea (165E–160W, 55N–65N), and Sea of Okhotsk (135E–165E, 45N–65N).

Regional predictions use training data only from that region, which does not account for predictive information outside the region boundaries that may advect across region boundaries before it does so. However it does allow for better selection of historical analogs in that only local information is used in weighting analogs. In this tradeoff we have tested using pan-Arctic training data for predicting regional sea ice anomalies, and find better predictive skill when only regional data is used for training.

An embedding window of 12 months is used in constructing the kernels in Eq. (2); 6 and 24 month embedding windows were also tested for robustness, and while results were similar for a 6 month window, results with 24 months were marginally worse than 12 months. We use an ensemble size of 100 (number of non-zeros entries per row retained in P), which represents about 2% of the total sample size, but the results are not sensitive to ensemble size (see Comeau et al (2017)). Lastly, we use the shape parameter $\sigma_0 = 2$.

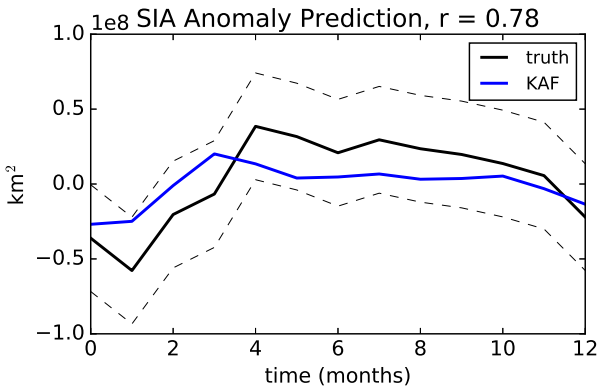

Fig. 2 Sample forecast trajectory of Arctic sea ice cover anomalies using SIC, SST, and SLP as predictors. Dashed lines represent 1 standard deviation of the sea ice cover anomaly over the test period.

4 Results

4.1 Pan-Arctic

We first focus on sea ice area anomalies, using SIC, SST, and SLP predictor data, and a 12 month embedding window. Figure 2 shows a sample forecast trajectory compared to the ground truth. While too much predictive value should not be inferred from single sample trajectories, it is common for forecasts to falter when near zero, as there is still difficulty in determining the sign of the forecast anomaly when the state is very near climatology, even with dynamic information encoded into the forecasting scheme. We show the degradation of the forecasts as lead times increase in Fig. 3, where forecasts are performed with 0, 3, 6, 9, and 12 month lead times. , the initial reconstruction matches the truth reasonably well, and forecasts at longer lead times become increasingly smoothed out towards climatology with increasing lead time.

To quantify how well the forecasts do on average, we consider the error metrics averaged over all points in the test period (400 years of monthly data, minus the length of the embedding window). In Fig. 4, we show pattern correlation conditioned on the target month for prediction and lead time, for KAF and damped persistence forecasts as a benchmark for comparison. Beyond initial reconstruction, KAF outperforms damped persistence in almost every regard, and is even above the 0.5 threshold for almost all of the first 6 months predicted range. Notable is KAF skill in predicting March and September anomalies, successful from 4–6 months out, whereas damped persistence loses skill in 1–2 months for these local extremes. Late spring is the most troublesome time to predict. We can see evidence of reemergence limbs in the damped persistence forecast, such as a fall-to-spring limb. Persistence suffers the worst in predicting August/September sea ice anomalies from 3–4 months

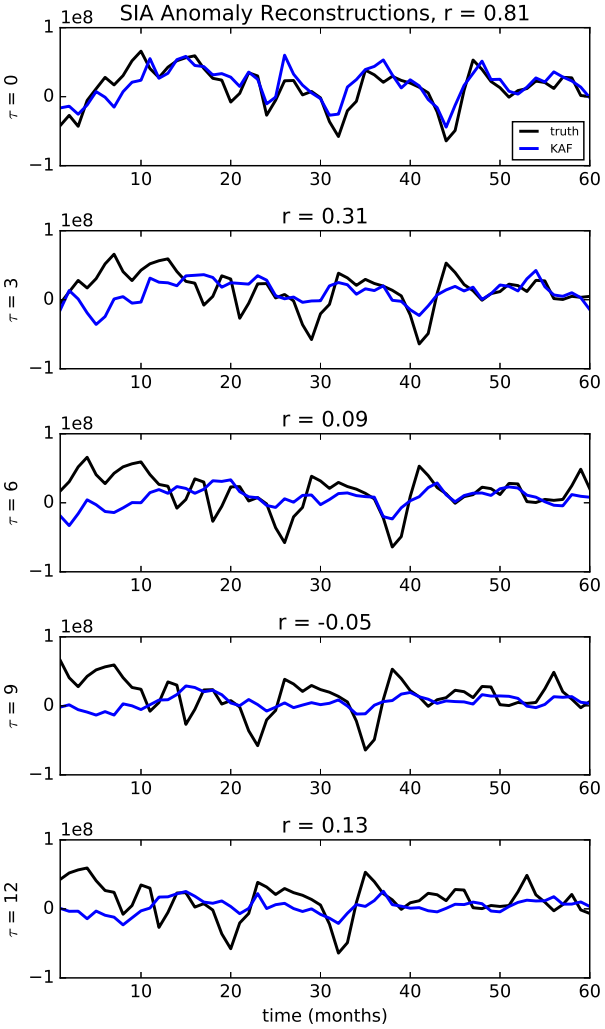

Fig. 3 Reconstructions at different time lags for Arctic sea ice area anomalies, taken as a random sample from the forecast period, initialized in January. The blue forecast at each point is made from a lead time indicated by the panel. Degradation of forecast fidelity is seen as the lead time increases.

out, and KAF has noticeably higher scores for predicting September pan-Arctic sea ice area anomaly, with skillful (PC $>$ 0.5) forecasts for lead times out to 9 months.

4.2 Regional Arctic

While predicting pan-Arctic sea ice area minimums and maximums has been of great interest, as more areas of the Arctic become accessible, an increased effort has been made in regional scale predictions. Snapshots of regional ice anomalies (calculated against regional climatologies) in Fig. 5 demonstrate different behavior around the Arctic basin. The out-of-sample extension values are plotted with the truth, and again should be thought of as the lead time 0 forecast. The central Arctic basins (Beaufort, Chukchi, East Siberian, &

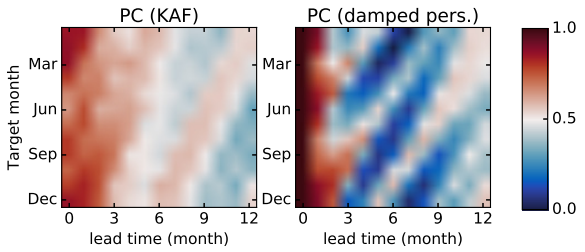

Fig. 4 Pattern correlation as a function of target month of prediction and lead time for pan-Arctic sea ice area anomaly forecasts using KAF (left) with SIC, SST, and SLP as predictors, and damped persistence (right). Considering a PC of 0.5 as a threshold for predictive skill, red indicates predictive skill, and blue indicates lack of skill. Beyond the initial reconstruction (lead time 0), KAF outperforms damped persistence in virtually every other regard. Early summer suffers the lowest predictive skill for short lead times, but is still predictable 4–5 months out. For predicting September and March sea ice area anomalies, note that damped persistence loses skill at short lead times, whereas KAF forecasts show predictive skill 4–6 months out for these months. Note a fall to spring reemergence limb ('growth-to-melt') in both the KAF and damped persistence plots.

Laptev, moving clockwise) are largely perennially ice covered, with less variability than other Arctic regions, and we will see that the abundance of time spent near climatology makes predicting anomalies difficult. Continuing clockwise to the Barents Sea, we begin to see the strong influence of the North Atlantic in regulating sea ice cover. More persistent anomalies are seen in the Barents and Greenland seas, which leads to greater predictability (Figs. 7 and 8). Moving across to the North Pacific basins, we have the Bering Sea and Sea of Okhotsk, regions in the marginal ice zone that may spend a couple months of the year completely free of ice. These regions in particular have been found to exhibit strong reemergence phenomena, in both SIC and SST fields (Alexander et al, 1999; Bushuk et al, 2014, 2015; Bushuk and Giannakis, 2015, 2017).

The aggregate error metrics, averaged over all months for each region in Figs. 6 (RMSE) and 7 (PC) show that KAF consistently outperforms damped persistence, (or at least fares no worse) once an initial reconstruction error is overcome, typically after only one month. In pattern correlation, disregarding PC scores below the 0.5 threshold may cut into some apparent gains of KAF over damped persistence, but it is worth noting the decay rate of KAF PC is slower than damped persistence, sometimes dramatically so (e.g. Bering and Labrador). The persistent nature of the North Atlantic adjacent basins seen in Fig. 5 manifests itself as slower than average decay of damped persistence. Also note the rise in damped persistence PC at the 12 month mark, for the North Pacific basins (Bering, Okhotsk) but also seen in pan-Arctic, suggesting reemergence phenomena.

Conditioning forecasts on the target month of prediction allows us to parse out seasonal impacts on predictability. The combined spatial and temporal effects of predictability highlight particularly skillful months and regions to predict, as seen in Fig. 8. Regional predictions suffer many lags and initialization months when there is no predictive skill, however times of success are clearly seen. In general there is a pattern that times of very low variability, either due to seasons when the region is completely ice-covered (e.g. central Arctic basins in the winter) or completely ice-free (e.g. Bering and Okhotsk in the summer), translate to poor predictive skill by KAF as measured by PC (since this score is normalized by variance). Late summer and fall are in general the most predictable to varying degrees, a pattern that largely appears in each region with the exception of the North Pacific Bering and Okhotsk regions, where predictive skill is low during these times. Notable is that the September anomaly is the most predictable, but only from a lead time of a couple months. The North Atlantic adjacent regions exhibit larger extent of predictive skill, and some form of reemergence seems to be aiding the forecasts. To demonstrate the gain in predictive skill of KAF over damped persistence, rather than plot damped persistence PC by target month as in Fig. 4, we instead plot the difference in pattern correlation scores, KAF over damped persistence, in Fig. 9. We zero out any value where both PC scores are below the threshold of 0.5, which we consider as not relevant to predictive skill. Most notable improvement over damped persistence is in the North Atlantic adjacent regions, though many regions also demonstrate improvement in predicting spring anomalies.

The areas of high PC in Fig. 8 by region and season are indicative of the different types of reemergence. In the central Arctic basins (e.g. Beaufort, Chukchi, E. Siberian, Laptev, and Kara), we see regions of high predictability are during the melt seasons, when growth-to-melt reemergence may aid in skill. Similarly, in the marginal ice zones of Bering Sea and Sea of Okhotsk, the regions of high predictability are in the growth seasons, where melt-to-growth reemergence is present.

4.3 Role of predictor variables

So far, the experiments we have shown have used SIC, SST, and (pan-Arctic) SLP as predictor variables, from which kernel evaluations to determine similarity are based (in Takens embedding space). To address the predictive power of using these components, in Fig. 10 we show the effect of combinations of SIC with each of SST, SLP, and SIC separately as predictors for pan-Arctic, as well as a representative perennial (Beaufort) and marginal (Bering) ice zone. In general, we find that KAF extracts much of its predictive power through SIC alone, with modest gains, or at times

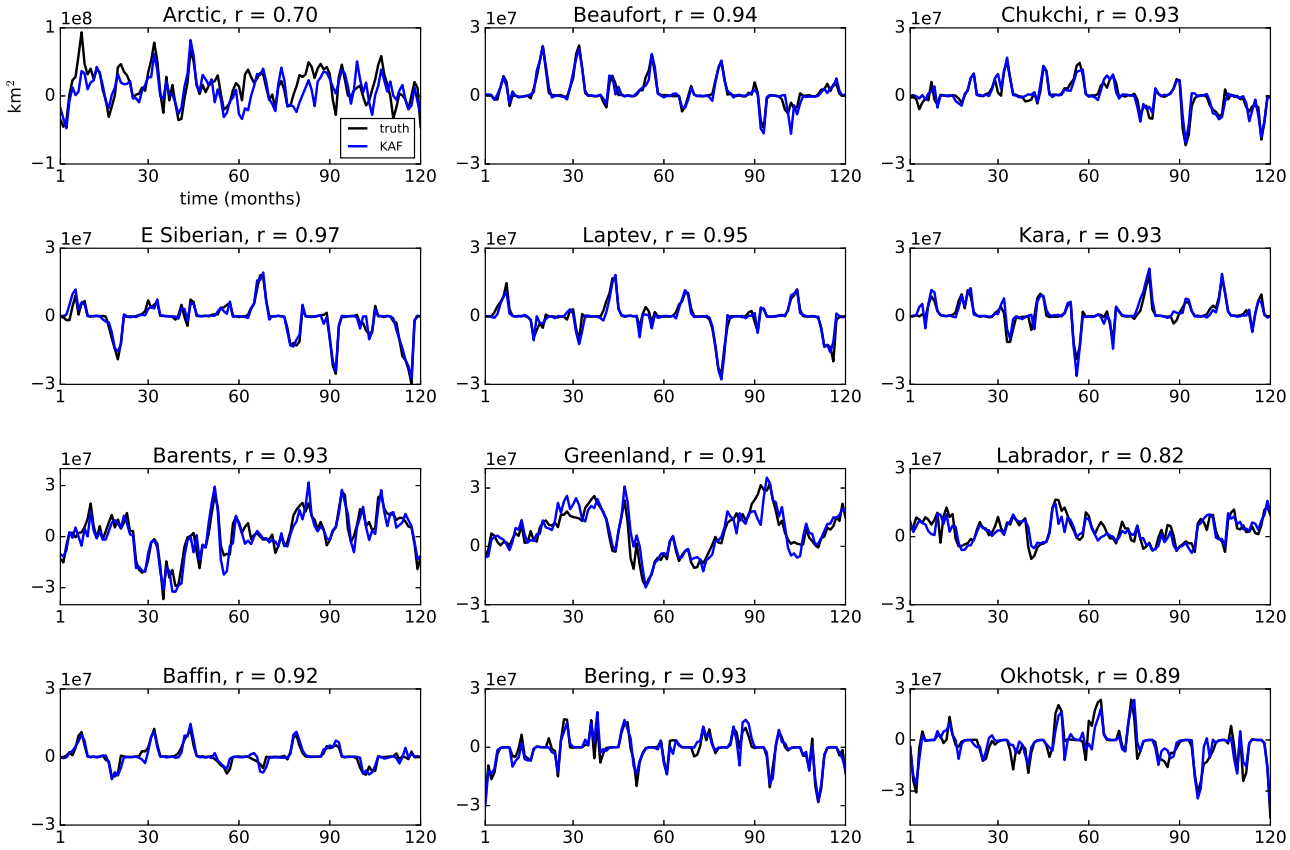

Fig. 5 Initial reconstructions (lead time 0) of regional sea ice area anomalies compared to truth, taken as a random sample from the forecast period, initialized in January. Note the regions that have more persistent anomalies (Barents and Greenland), are North Atlantic adjacent, which has been found in other studies to be regions of relatively high predictability.

loses, when including an additional predictor. While there is marginal difference in adding SST or SLP to SIC, including SIT is actually detrimental to predictions in the pan-Arctic sea ice area anomaly setting. This may seem surprising, given other studies' emphasis on the importance of sea ice thickness measurements, however in the context of kernel evaluation, increasing the dimension of our state vector may yield less discernible informative historical analogs. A similar degradation of performance when including SIT data in the kernel was observed in the study of Bushuk and Giannakis (2017) on SIT-SIC reemergence mechanisms. This behavior was attributed to the slower characteristic timescale of SIT data, resulting in this variable dominating the phase velocity-dependent kernel in Eq. (1). This could in part be mitigated by allowing for a longer embedding window for SIT; Fig. 11 shows PC scores for SIC anomalies using SIC, SST, and SIT predictors, where increasing the embedding window for SIT from $q = 12$ to $q = 48$ months yields some increase in predictive skill on what appear to be reemergence limbs, with a corresponding decrease in skill at short lead times. In such interannual embedding windows the NLSA kernel is known to capture low-frequency modes of variability and the associated modulations at the annual cycle through its leading eigenfunctions; e.g. the North Pacific Gyre Oscillation in high latitudes (Bushuk et al, 2014) and ENSO in the tropics (Slawinska and Giannakis, 2017; Giannakis and Slawinska, 2017). These eigenfunctions naturally capture reemergence phenomena, and are expected to play an important role in the behavior of the kernel in forecasting, which is consistent with the behavior seen in Fig. 11.

Turning to an atmospheric predictor variable, while the inclusion of pan-Arctic SLP does not hamper our prediction skill, it offers only marginal improvement. This is most likely due to the fact that the quantities used are monthly averaged, and perhaps too temporally coarse to reflect the chaotic atmospheric influence on sea ice cover on shorter time scales, or that SLP itself is not predictable on month long time scales.

4.4 Regional volume anomalies

We also consider the problem of forecasting sea ice volume anomalies, which in general show more persistence than sea ice cover anomalies. This increased persistence in volume

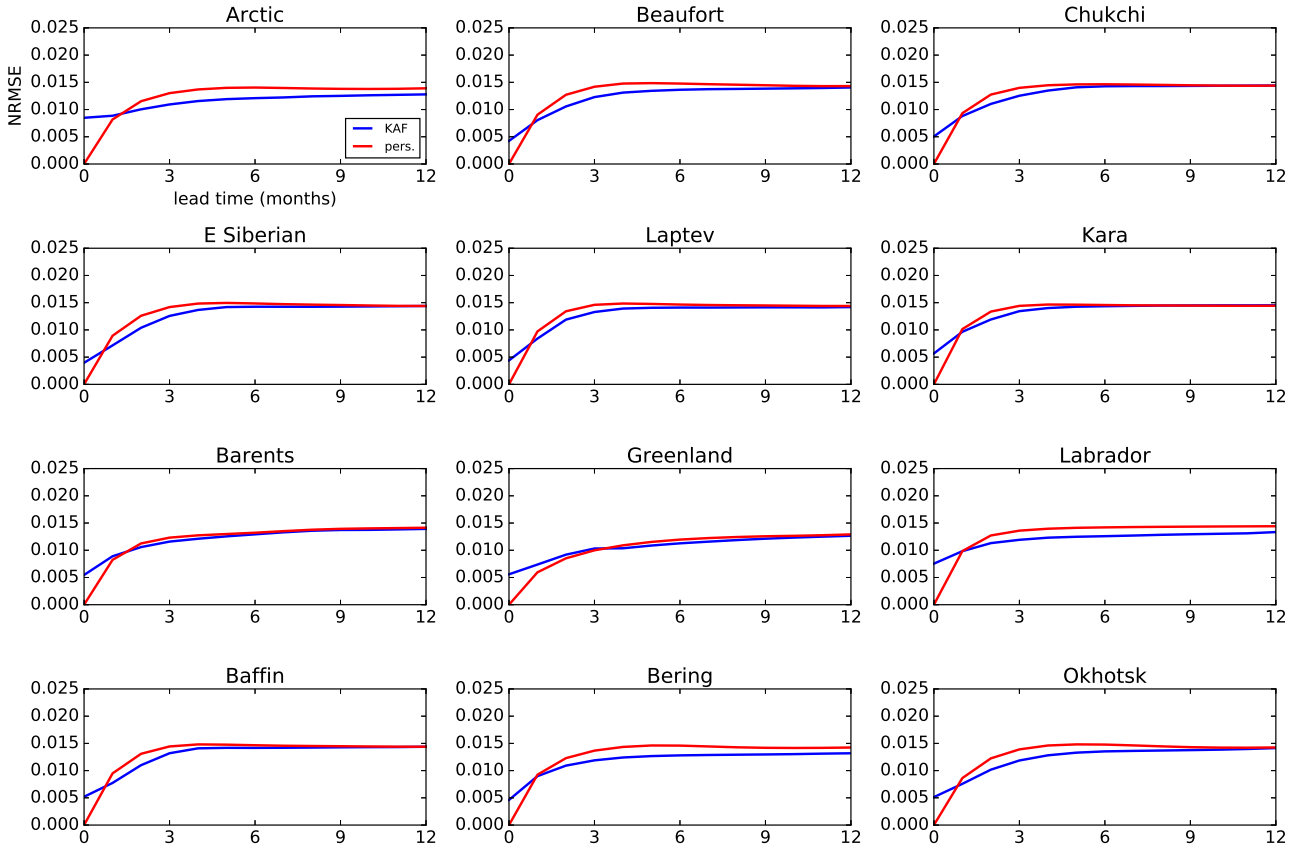

Fig. 6 RMSE for regional sea ice area anomaly predictions, averaged over all months, for KAF and damped persistence as a benchmark. KAF suffers reconstruction errors at lead time 0, and then outperforms damped persistence usually after one month.

over concentration anomalies is due in part because thinner ice (which drives concentration anomalies) is more sensitive to being advected by winds to areas that are more or less prone to melting. Figure 12 shows regional forecast PC for predicting sea ice volume anomalies, having only observed SIC, SST, and SLP. In this example we are predicting an unobserved variable, and in some areas (late fall, early winter) KAF is remarkably skillful in reconstructing the unobserved quantity, with skill from lead times of a couple months out. However, these do not compare favorably against a damped persistence forecast using the ground truth (not shown), due to inherent persistence of volume anomalies (compared to area anomalies which persist on a shorter timescale), though this would also not be a fair comparison given the KAF forecasts are not observing the full observable. We note that in comparing to Fig. 8 we see similar patterns of predictive skill by region, suggesting that again the different types of reemergence are aiding in skill.

When we include SIT as predictor data with an increased embedding window of $q = 48$ months, we expectedly see a substantial increase in skill. In Fig. 13, we show the difference in PC score of KAF over damped persistence (similar to Fig 9, but for sea ice volume anomalies). Damped persistence outperforms KAF at short lead times (0–3 months), largely due to the initial reconstruction error in KAF. For longer lead time (3–12 months), KAF retains predictive skill with PC scores that far exceed those of damped persistence in many regional forecasts (the difference in PC scores is zeroed out if both scores are below 0.5, as in Fig 9). This gain in predictive skill in regional forecasts does not translate to pan-Arctic forecasts, where the difference between damped persistence and KAF is quite small, but follows the pattern that damped persistence scores higher at short lead times (0–6 months), and KAF scores higher at longer lead times (6–12 months).

5 Discussion & Conclusions

In this paper, we used KAF (Zhao and Giannakis, 2016; Comeau et al, 2017; Alexander et al, 2017), a nonparametric method using weighted ensembles of analogs, to predict Arctic sea ice area anomalies, then volume anomalies, for both pan-Arctic and regional scales, examining the effects of including SIC, SST, SLP, and SIT as predictors for our method. We find in general that for predicting sea ice

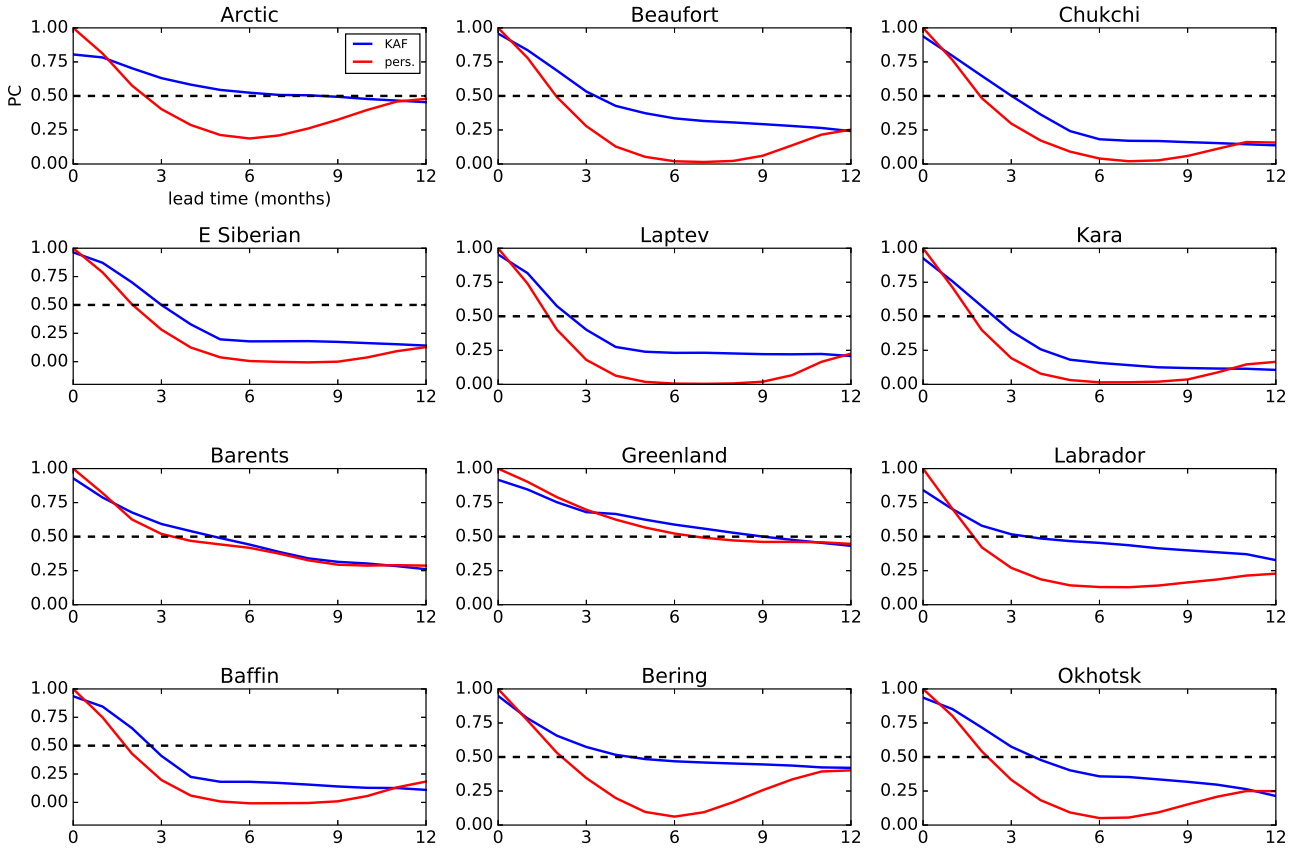

Fig. 7 Pattern correlation scores for regional sea ice area anomaly prediction by region, averaged over all months. Reconstruction errors are less noticeable in this metric (apart from pan-Arctic), and KAF exceeds damped persistence in every region. Note the North Atlantic regions are the most persistent, as demonstrated by slow decay of PC. There, regional improvements over damped persistence are marginal, though in pan-Arctic, the improvement is several months.

concentration anomalies, KAF outperforms the damped persistence forecast, or at minimum does not perform worse (with the exception of the inherent lag 0 reconstruction error) and the outperformance lead times range between 1 and 9 months. Moving to regional scale basins and conditioning on the target month of prediction, we see clear regional-seasonal domains when KAF succeeds, as well as those when it fails (along with damped persistence).

The North Atlantic seems to have a strong impact on sea ice area anomalies, as the adjacent regions (Barents and Greenland Seas) exhibit the strongest persistent anomalies, and have the highest predictability. The marginal ice zone basins in the North Pacific (Bering Sea, Sea of Okhotsk) show similar behavior, but to a lesser degree. The basins in the central Arctic (Beaufort, Chukchi, East Siberian, and Laptev) have less variability to their sea ice cover, and thus being close to climatology for much of the time makes predicting excursions from the climatology difficult. Late summer and early fall are in general the best times to predict with KAF, with skillful forecasts at lead times 3–6 months. Late winter and spring is in general the time period of least predictability, except for marginal ice zones and pan-Arctic. The areas of high predictability by region and season are consistent with reemergence phenomena, with central Arctic basins benefiting from growth to melt reemergence, and marginal ice zones benefiting from melt to growth reemergence.

We find most of the predictive information for sea ice concentration is in SIC alone, with each of SST, SLP and SIT providing marginal improvements, although in some cases the inclusion of SIT actually hampers predictive skill. While we have success in reconstructing sea ice volume anomalies without using SIT as a predictor at the regional level, we see drastically improved performance with the inclusion of SIT in predicting pan-Arctic volume anomalies, particularly at the regional level, where forecasts remain skillful at 12 month lead times.

Ultimately, the goal is to move to an operational prediction based on observational data, for which this is a first step. By using model data, we are able to make use of a long control run that has sampled the climate's natural variability and perform statistically robust estimates of skill. Limitations on

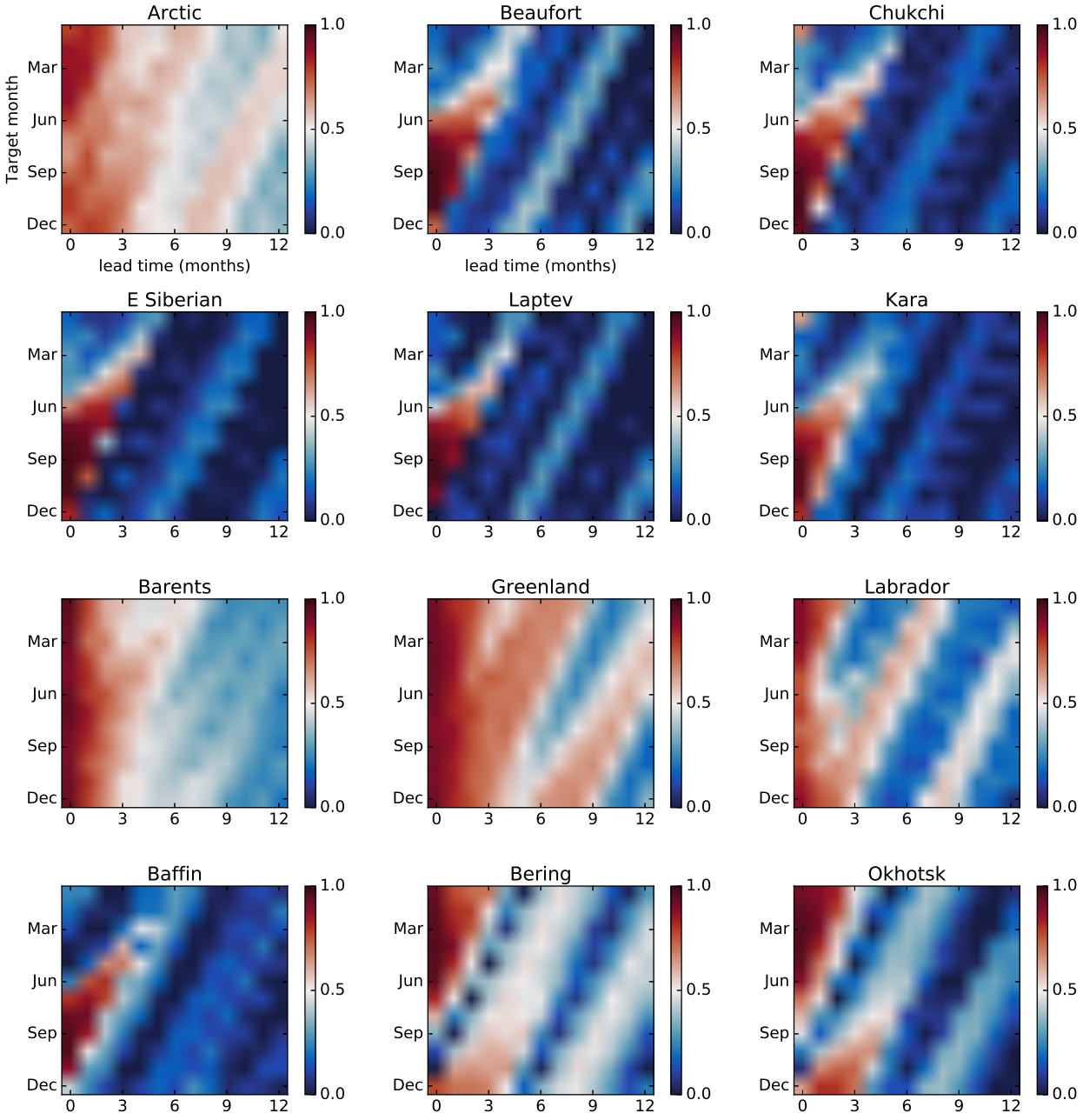

Fig. 8 Pattern correlation for KAF forecasts of sea ice concentration anomalies as a function of target month and lead time by region. Pan-Arctic values are reproduced from the first panel of Fig. 4. Notice pan-Arctic Sep seems lightest color for lag 0, but is over 0.5 for 3 – 4 months. In general across regions, predicting fall is typically the most skillful, with the exception of the North Pacific basins in marginal ice zones (Bering Sea and Sea of Okhotsk), where higher predictability is in the winter. The North Atlantic adjacent regions have a larger extent of predictive skill, with apparent limbs of high correlation that may be reemergence processes aiding KAF.

the quality (i.e. presence of model biases or observation errors) and the length of training data will impact the performance of KAF, as it would any other statistical method. Experiments with much shorter lengths of control training data (e.g. 40 years) show a sharp decrease in KAF predictive skill (Fig 14), underscoring the need for a rich enough set of training data where the system's full internal variability has been explored, even without the presence of a changing climate. Utilizing KAF to predict internal variability in conjunction with some method to account for the changing mean Arctic state would implicitly assume the internal climate variability itself is not changing, which also merits consideration.

Our future research plan is to use NLSA to extract an underlying 'trend' in the data as a way of non-parametrically

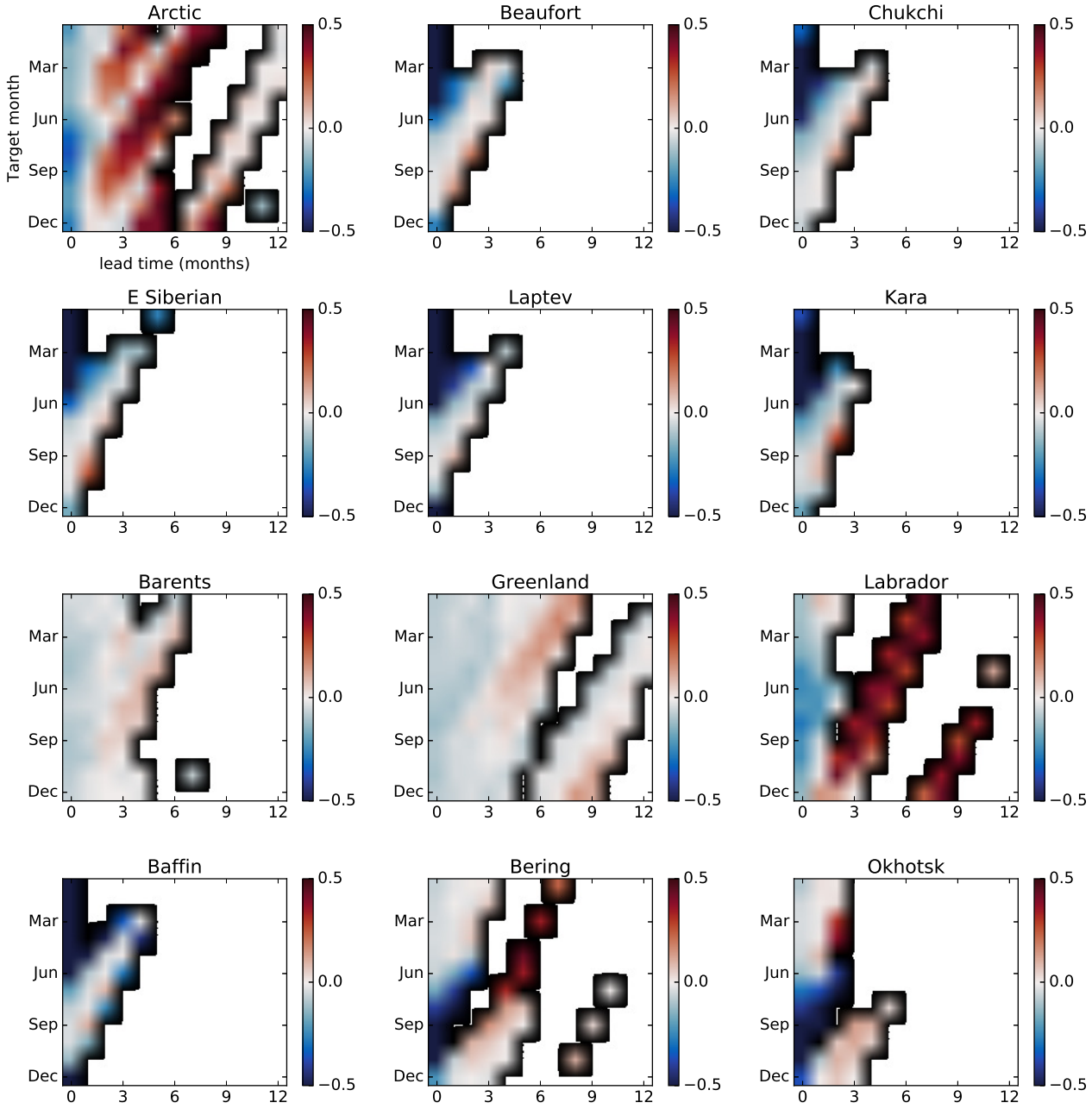

Fig. 9 Difference in pattern correlation scores of KAF over damped persistence to illustrate the gain in predictive skill, with zero in place of any value where both scores are below 0.5. Red indicates KAF outperforming damped persistence, and blue vice-versa. For pan-Arctic, a big improvement over damped persistence is observed late in the year, with a gap suggesting reemergence phenomena. Strong improvement is observed in predictability in North Atlantic adjacent regions, as well as in predicting (early) spring and autumn anomalies in many regions.

determining a trend (as opposed to fitting a linear or quadratic regression). This trend could then be extended to a forecast time using some form of extrapolation or out-of-sample extension technique, while the anomalies from this trend would be forecasted by the KAF method using long training datasets as in this study. Other research directions include using a blended damped persistence and analog forecasting approach to avoid the initial reconstruction errors at short time scales, as well as forecasts using kernels targeted at specific observables.

Acknowledgements The research of Andrew Majda and Dimitrios Giannakis is partially supported by ONR MURI grant 25-74200-F7112. Darin Comeau was supported as a postdoctoral fellow through this grant. Dimitrios Giannakis and Zhizhen Zhao are partially supported by NSF grant DMS-1521775. Dimitrios Giannakis also acknowledges support from ONR grant N00014-14-1-0150. We thank Mitch Bushuk

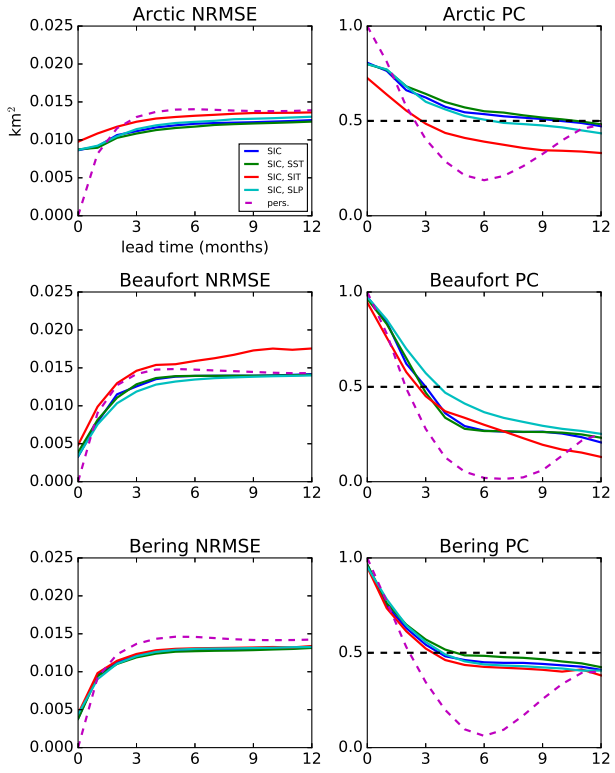

Fig. 10 Prediction results for Arctic sea ice area anomalies using different predictor variables for the pan-Arctic, a region in a mainly perennial ice zone (Beaufort Sea), and a region in a marginal ice zone (Bering Sea). Most of the skill is from SIC alone, although SIC and SST performs (marginally) the best. Interestingly, adding ice thickness information actually hinders pan-Arctic sea ice area anomaly prediction.

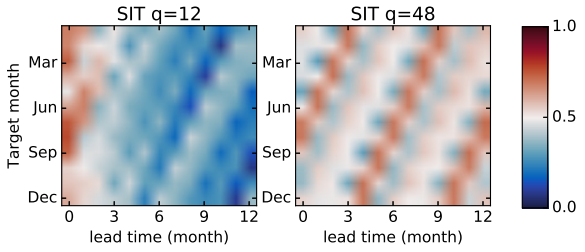

Fig. 11 Prediction results (PC scores) using SIC, SST, and SIT predictors for Arctic sea ice area anomalies with varying embedding windows for SIT. Increasing the embedding window from q=12 months (default) to q=48 months for SIT only yields an increase in predictive skill on reemergence limbs.

for helpful discussions We also thank two anonymous reviewers for their helpful comments in reviewing this manuscript.

References

Alessandrini S, Delle Monache L, Sperati S, Nissen J (2015) A novel application of an analog ensemble for short-term wind power forecasting. Renewable Energy 76:768–781

Alexander MA, Deser C, Timlin MS (1999) The reemergence of SST anomalies in the North Pacific Ocean. Journal of Climate 12(8):2419–2433

Alexander R, Zhao Z, Szkely E, Giannakis D (2017) Kernel analog forecasting of tropical intraseasonal oscillations. Journal of the Atmospheric Sciences 74(4):1321–1342

Atencia A, Zawadzki I (2015) A comparison of two techniques for generating nowcasting ensembles. part ii: Analogs selection and comparison of techniques. Monthly Weather Review 143(7):2890–2908

Blanchard-Wrigglesworth E, Bitz CM (2014) Characteristics of Arctic sea-ice thickness variability in GCMs. Journal of Climate 27(21):8244–8258

Blanchard-Wrigglesworth E, Armour KC, Bitz CM, DeWeaver E (2011a) Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations. Journal of Climate 24(1):231–250

Blanchard-Wrigglesworth E, Bitz C, Holland M (2011b) Influence of initial conditions and climate forcing on predicting Arctic sea ice. Geophysical Research Letters 38(18)

Blanchard-Wrigglesworth E, Cullather R, Wang W, Zhang J, Bitz C (2015) Model forecast skill and sensitivity to initial conditions in the seasonal Sea Ice Outlook. Geophysical Research Letters 42(19):8042–8048

Broomhead DS, King GP (1986) Extracting qualitative dynamics from experimental data. Physica D: Nonlinear Phenomena 20(2-3):217–236

Bushuk M, Giannakis D (2015) Sea-ice reemergence in a model hierarchy. Geophysical Research Letters 42(13):5337–5345

Bushuk M, Giannakis D (2017) The seasonality and interannual variability of Arctic sea ice reemergence. Journal of Climate 30(12):4657–4676, DOI 10.1175/JCLI-D-16-0549.1

Bushuk M, Giannakis D, Majda AJ (2014) Reemergence mechanisms for North Pacific sea ice revealed through nonlinear Laplacian spectral analysis. Journal of Climate 27(16):6265–6287

Bushuk M, Giannakis D, Majda AJ (2015) Arctic sea ice reemergence: The role of large-scale oceanic and atmospheric variability. Journal of Climate 28(14):5477–5509

Bushuk M, Msadek R, Winton M, Vecchi GA, Gudgel R, Rosati A, Yang X (2017) Summer enhancement of Arctic sea ice volume anomalies in the september-ice zone. Journal of Climate 30(7):2341–2362

Chevallier M, Salas-Mélia D (2012) The role of sea ice thickness distribution in the Arctic sea ice potential predictability: A diagnostic approach with a coupled gcm. Journal of Climate 25(8):3025–3038

Chevallier M, Salas y Mélia D, Voldoire A, Déqué M, Garric G (2013) Seasonal forecasts of the pan-Arctic sea ice ex-

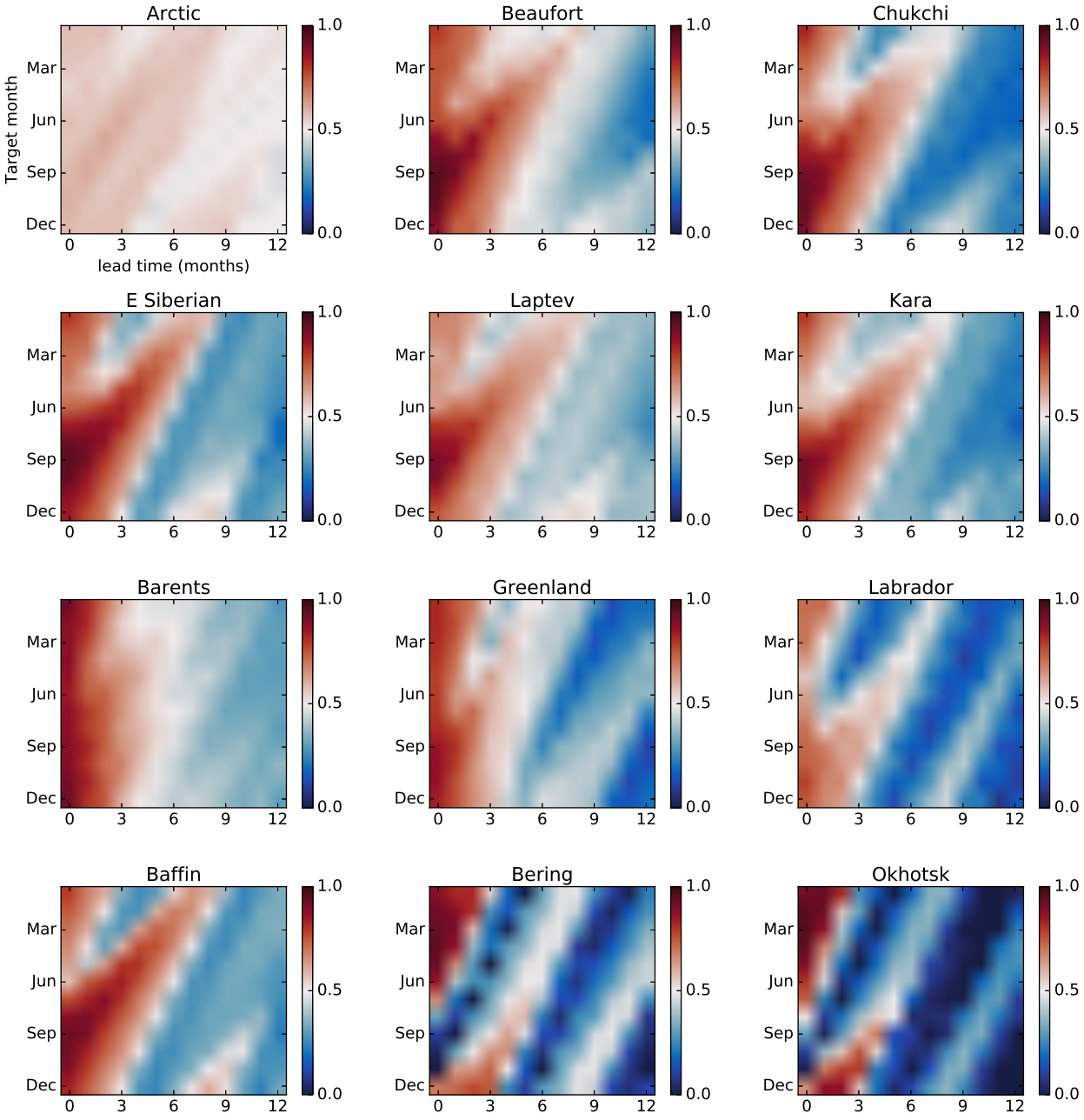

Fig. 12 Forecasts for Arctic sea ice volume anomalies, predicted using only SIC, SST, and SLP data. By not using SIT as a predictor variable, we are predicting an unobserved quantity, and damped persistence forecasts based on the true anomaly outperform KAF on all spatial and temporal scales (not shown).

tent using a GCM-based seasonal prediction system. Journal of Climate 26(16):6092–6104

Comeau D, Zhao Z, Giannakis D, Majda AJ (2017) Data-driven prediction strategies for low-frequency patterns of North Pacific climate variability. Climate Dynamics 48(5):1855–1872

Day J, Tietsche S, Hawkins E (2014) Pan-Arctic and regional sea ice predictability: Initialization month dependence. Journal of Climate 27(12):4371–4390

Deyle ER, Sugihara G (2011) Generalized theorems for nonlinear state space reconstruction. PLoS One 6(3):e18,295

Van den Dool H (1994) Searching for analogues, how long must we wait? Tellus A 46(3):314–324

Drosdowsky W (1994) Analog (nonlinear) forecasts of the Southern Oscillation index time series. Weather and Forecasting 9(1):78–84

Fernández A, Rabin N, Fishelov D, Dorronsoro JR (2013) Auto-adaptative laplacian pyramids for high-dimensional

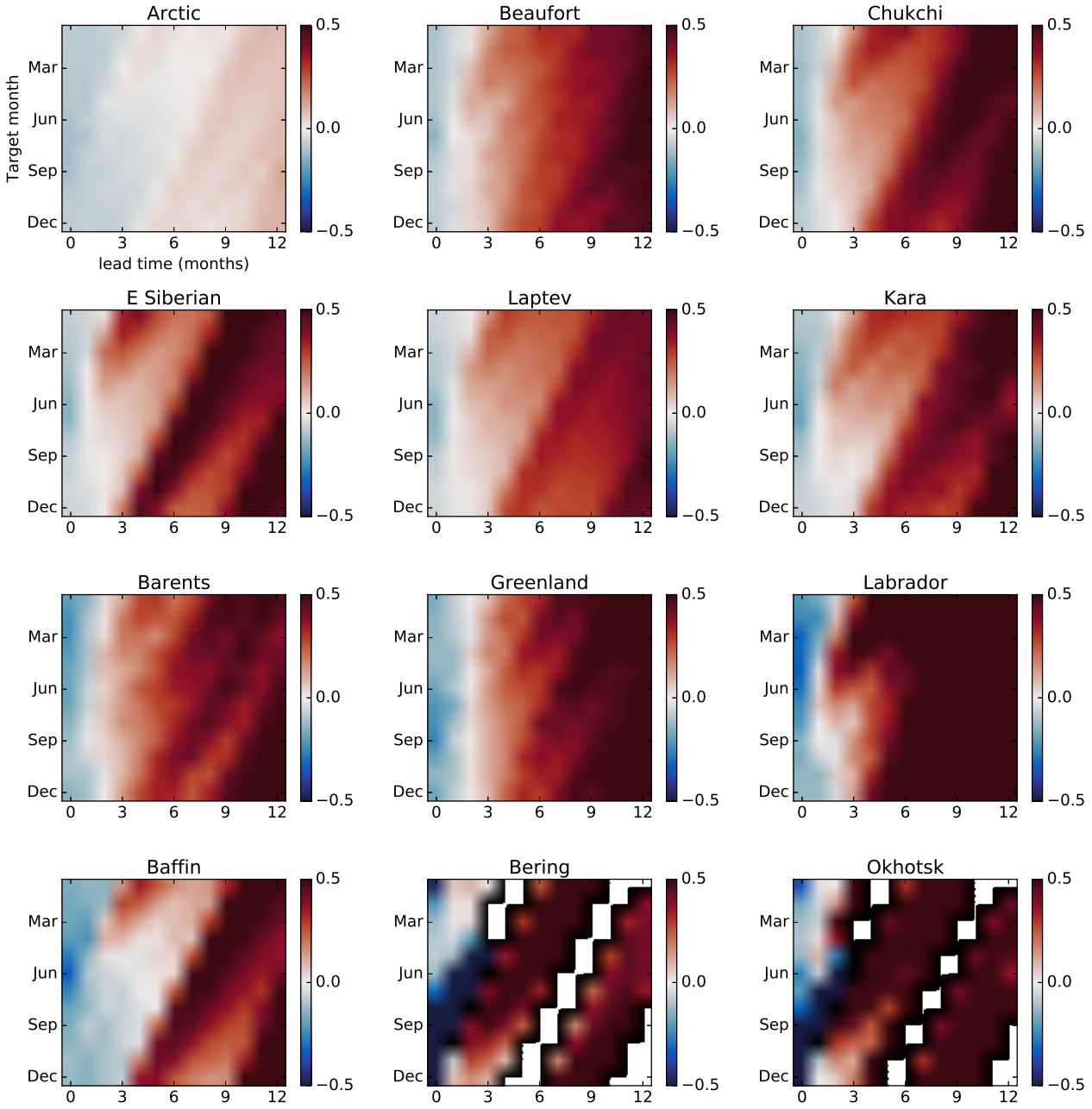

Fig. 13 Difference in PC score of KAF over damped persistence for predicting sea ice volume anomalies, with predictor variables SIC, SST, and SIT. SIC and SST have the default $q = 12$ month embedding window, whereas SIT has an extended embedding window of $q = 48$ months. At short lead times damped persistence outperforms KAF (blue colors), however beyond 3 month lead times KAF strongly outperforms damped persistence in many regional forecasts, with a minor improvement in pan-Arctic anomalies.

data analysis. arXiv preprint arXiv:13116594

Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR, Lawrence DM, Neale RB, Rasch PJ, Vertenstein M, et al (2011) The community climate system model version 4. Journal of Climate 24(19):4973–4991

Germe A, Chevallier M, y Mélia DS, Sanchez-Gomez E, Cassou C (2014) Interannual predictability of arctic sea ice in a global climate model: regional contrasts and temporal evolution. Climate dynamics 43(9-10):2519–2538

Giannakis D (2015) Dynamics-adapted cone kernels. SIAM Journal on Applied Dynamical Systems 14(2):556–608

Giannakis D, Majda AJ (2012a) Comparing low-frequency and intermittent variability in comprehensive climate models through nonlinear Laplacian spectral analysis. Geophysical Research Letters 39(10)

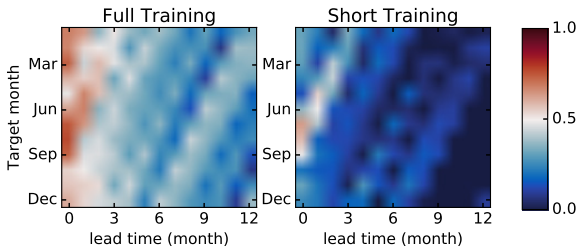

Fig. 14 Prediction results (PC scores) using SIC, SST, and SIT predictors for Arctic sea ice area anomalies comparing different length training records; full training of 400 years, and short training of only 40 years (similar length to the satellite record). Predictive skill is significantly hampered when the training record is not long enough to contain a suitable set of analogs.

Giannakis D, Majda AJ (2012b) Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. Proceedings of the National Academy of Sciences 109(7):2222–2227

Giannakis D, Majda AJ (2013) Nonlinear Laplacian spectral analysis: capturing intermittent and low-frequency spatiotemporal patterns in high-dimensional data. Statistical Analysis and Data Mining 6(3):180–194

Giannakis D, Majda AJ (2014) Data-driven methods for dynamical systems: Quantifying predictability and extracting spatiotemporal patterns. Mathematical and Computational Modeling: With Applications in Engineering and the Natural and Social Sciences p 288

Giannakis D, Slawinska J (2017) Indo-pacific variability on seasonal to multidecadal timescales. part ii: Multiscale atmosphere-ocean linkages. Journal of Climate (2017)

Guemas V, Blanchard-Wrigglesworth E, Chevallier M, Day JJ, Déqué M, Doblas-Reyes FJ, Fučkar NS, Germe A, Hawkins E, Keeley S, et al (2014) A review on arctic sea-ice predictability and prediction on seasonal to decadal time-scales. Quarterly Journal of the Royal Meteorological Society

Holland MM, Stroeve J (2011) Changing seasonal sea ice predictor relationships in a changing Arctic climate. Geophysical Research Letters 38(18)

Holland MM, Bitz CM, Tremblay L, Bailey DA, et al (2008) The role of natural versus forced change in future rapid summer Arctic ice loss. Arctic sea ice decline: observations, projections, mechanisms, and implications pp 133–150

Holland MM, Bailey DA, Vavrus S (2011) Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model, version 3. Climate Dynamics 36(7-8):1239–1253

Holland MM, Bailey DA, Briegleb BP, Light B, Hunke E (2012) Improved sea ice shortwave radiation physics in CCSM4: the impact of melt ponds and aerosols on Arctic sea ice*. Journal of Climate 25(5):1413–1430

Hunke E, Lipscomb W (2008) CICE: The Los Alamos sea ice model, documentation and software, version 4.0, Los Alamos National Laboratory tech. rep. Tech. rep., LA-CC-06-012

Koenigk T, Mikolajewicz U (2009) Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model. Climate dynamics 32(6):783–798

Lindsay R, Zhang J, Schweiger A, Steele M (2008) Seasonal predictions of ice extent in the Arctic Ocean. Journal of Geophysical Research: Oceans 113(C2)

Lorenz EN (1969) Atmospheric predictability as revealed by naturally occurring analogues. Journal of the Atmospheric sciences 26(4):636–646

Neale RB, Richter JH, Conley AJ, Park S, Lauritzen PH, Gettelman A, Williamson DL, et al (2010) Description of the ncar community atmosphere model (cam 4.0). NCAR Tech Note NCAR/TN-485+ STR

Ogi M, Yamazaki K, Wallace JM (2010) Influence of winter and summer surface wind anomalies on summer Arctic sea ice extent. Geophysical Research Letters 37(7)

Packard NH, Crutchfield JP, Farmer JD, Shaw RS (1980) Geometry from a time series. Physical Review Letters 45(9):712

Rabin N, Coifman RR (2012) Heterogeneous datasets representation and learning using diffusion maps and Laplacian pyramids. In: SDM, SIAM, pp 189–199

Rigor IG, Wallace JM (2004) Variations in the age of arctic sea-ice and summer sea-ice extent. Geophysical Research Letters 31(9)

Rigor IG, Wallace JM, Colony RL (2002) Response of sea ice to the arctic oscillation. Journal of Climate 15(18):2648–2663

Robinson JC (2005) A topological delay embedding theorem for infinite-dimensional dynamical systems. Nonlinearity 18(5):2135

Sauer T, Yorke JA, Casdagli M (1991) Embedology. Journal of statistical Physics 65(3-4):579–616

Schröder D, Feltham DL, Flocco D, Tsamados M (2014) September Arctic sea-ice minimum predicted by spring melt-pond fraction. Nature Climate Change 4(5):353–357

Sigmond M, Fyfe J, Flato G, Kharin V, Merryfield W (2013) Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system. Geophysical Research Letters 40(3):529–534

Sigmond M, Reader MC, Flato GM, Merryfield WJ, Tivy A (2016) Skillful seasonal forecasts of arctic sea ice retreat and advance dates in a dynamical forecast system. Geophysical Research Letters 43(24):12,457–12,465

Slawinska J, Giannakis D (2017) Indo-pacific variability on seasonal to multidecadal timescales. part i: Intrinsic sst modes in models and observations. Journal of Climate

30(14):5265–5294, DOI 10.1175/JCLI-D-16-0176.1

Smith LC, Stephenson SR (2013) New trans-Arctic shipping routes navigable by midcentury. Proceedings of the National Academy of Sciences 110(13):E1191–E1195

Smith R, Jones P, Briegleb B, Bryan F, Danabasoglu G, Dennis J, Dukowicz J, Eden C, Fox-Kemper B, Gent P, et al (2010) The Parallel Ocean Program (POP) reference manual: Ocean component of the Community Climate System Model (CCSM). Los Alamos National Laboratory, LAUR-10-01853

Stroeve J, Hamilton LC, Bitz CM, Blanchard-Wrigglesworth E (2014) Predicting September sea ice: Ensemble skill of the SEARCH sea ice outlook 2008–2013. Geophysical Research Letters 41(7):2411–2418

Stroeve JC, Kattsov V, Barrett A, Serreze M, Pavlova T, Holland M, Meier WN (2012) Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations. Geophysical Research Letters 39(16)

Takens F (1981) Detecting strange attractors in turbulence. Springer

Tietsche S, Notz D, Jungclaus JH, Marotzke J (2013) Predictability of large interannual Arctic sea-ice anomalies. Climate dynamics 41(9-10):2511–2526

Tietsche S, Day J, Guemas V, Hurlin W, Keeley S, Matei D, Msadek R, Collins M, Hawkins E (2014) Seasonal to interannual Arctic sea ice predictability in current global climate models. Geophysical Research Letters 41(3):1035–1043

Tietsche S, Hawkins E, Day JJ (2016) Atmospheric and oceanic contributions to irreducible forecast uncertainty of arctic surface climate. Journal of Climate 29(1):331–346

Tilling RL (2016) Near-real-time arctic sea ice thickness and volume from cryosat-2. The Cryosphere 10(5):2003

Wang L, Yuan X, Ting M, Li C (2015) Predicting summer Arctic sea ice concentration intra-seasonal variability using a vector auto-regressive model. Journal of Climate (2015)

Wang W, Chen M, Kumar A (2013) Seasonal prediction of Arctic sea ice extent from a coupled dynamical forecast system. Monthly Weather Review 141(4):1375–1394

Xavier PK, Goswami BN (2007) An analog method for real-time forecasting of summer monsoon subseasonal variability. Monthly Weather Review 135(12):4149–4160

Yeager SG, Karspeck AR, Danabasoglu G (2015) Predicted slowdown in the rate of atlantic sea ice loss. Geophysical Research Letters 42(24)

Zhang J, Rothrock D (2003) Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. Monthly Weather Review 131(5):845–861

Zhang J, Steele M, Lindsay R, Schweiger A, Morison J (2008) Ensemble 1-year predictions of Arctic sea ice for the spring and summer of 2008. Geophysical Research Letters 35(8)

Zhao Z, Giannakis D (2016) Analog forecasting with dynamics-adapted kernels. Nonlinearity 29(9):2888