Title
Machine learning for daily forecasts of Arctic sea-ice motion: an attribution assessment of model predictive skill

Authors
Hoffman, Lauren
Mazloff, Matthew R
Gille, Sarah T
et al.

Publication Date
2023-08-17

DOI
10.1175/aies-d-23-0004.1

Machine learning for daily forecasts of Arctic sea-ice motion: an attribution assessment of model predictive skill.

Lauren Hoffman,[a] Matthew R. Mazloff,[a] Sarah T. Gille,[a] Donata Giglio,[b] Cecilia M. Bitz,[c] Patrick Heimbach,[d] Kayli Matsuyoshi,[a]

[a] *Scripps Institution of Oceanography, University of California, San Diego, California, USA*
[b] *University of Colorado Boulder, Boulder, Colorado, USA* [c] *University of Washington, Seattle, Washington, USA* [d] *The University of Texas at Austin, Austin, Texas, USA*

Corresponding author: Lauren Hoffman, lahoffma@eng.ucsd.edu

ABSTRACT: Physics-based simulations of Arctic sea ice are highly complex, involving transport between different phases, length scales, and time scales. Resultantly, numerical simulations of sea-ice dynamics have a high computational cost and model uncertainty. We employ data-driven machine learning (ML) to make predictions of sea-ice motion. The ML models are built to predict present-day sea-ice velocity given present-day wind velocity and previous-day sea-ice concentration and velocity. Models are trained using reanalysis winds and satellite-derived sea-ice properties. We compare the predictions of three different models: persistence (PS), linear regression (LR), and convolutional neural network (CNN). We quantify the spatio-temporal variability of the correlation between observations and the statistical model predictions. Additionally, we analyze model performance in comparison to variability in properties related to ice motion (wind velocity, ice velocity, ice concentration, distance from coast, bathymetric depth) to understand the processes related to decreases in model performance. Results indicate that a CNN makes skillful predictions of daily sea-ice velocity with a correlation up to 0.81 between predicted and observed sea-ice velocity, while the LR and PS implementations exhibit correlations of 0.78 and 0.69, respectively. The correlation varies spatially and seasonally; lower values occur in shallow coastal regions and during times of minimum sea-ice extent. LR parameter analysis indicates that wind velocity plays the largest role in predicting sea-ice velocity on one-day time scales, particularly in the central Arctic. Regions where wind velocity has the largest LR parameter are regions where the CNN has higher predictive skill than the LR.

SIGNIFICANCE STATEMENT: We build and evaluate different machine learning (ML) models that make one-day predictions of Arctic sea-ice velocity using present-day wind velocity and previous-day ice concentration and ice velocity. We find that models that incorporate non-linear relationships between inputs (a neural network) capture important information (i.e. have a higher correlation between observations and predictions than linear and persistence models). This performance enhancement occurs primarily in deeper regions of the central Arctic where wind speed is the dominant predictor of ice motion. Understanding where these models benefit from increased complexity is important because future work will use ML to elucidate physically meaningful relationships within the data, looking at how the relationship between wind and ice velocity is changing as the ice melts.

1. Introduction

Sea-ice cover in the Arctic has been diminishing since the beginning of the satellite record. (Serreze et al. 2007; Stroeve et al. 2012; Stroeve and Notz 2018; Thoman et al. 2022). Negative trends in sea-ice concentration, thickness, and multiyear ice coverage (Carmack et al. 2015) have been reported throughout the Arctic, whereas the length of the melt season, drift speeds, and deformation rates are increasing (Stroeve and Notz 2018; Rampal et al. 2009; Onarheim et al. 2018). Climate model simulations indicate a substantial likelihood that the Arctic Ocean will become largely ice free during September by 2100 if warming exceeds 2ºC (Stroeve and Notz 2018; Notz and Stroeve 2018; Jahn 2018; Meredith et al. 2019). Transition to thinner and more fragile ice will have widespread environmental, geopolitical, and logistical impacts, including potential for new increased maritime activity (Bennett et al. 2020; Crawford et al. 2021; Cao et al. 2022), with which comes the need to know where sea-ice is and the need for skillful predictions of where it will be. In this study we contribute to addressing these issues by assessing the skill of machine learning models in making one-day predictions of sea-ice motion. We design these models to predict present-day ice motion based on previous-day observations, and show proof-of-concept for applications in operational forecasting that would allow information about the ice state to be obtained before satellite retrievals are processed. Additionally, we explore the extent to which these ML models will have enough skill to be used to represent the dynamical component of sea ice in a simulation framework that provides nowcasting of the state of Arctic sea ice.

Predictions of sea-ice motion have almost exclusively been attempted with numerical prediction models (Petrou and Tian 2019). While these state-of-the-art physics-based models for sea-ice prove useful, their inherent complexity comes with a high computational cost (Hunke et al. 2020). There are also several sources of uncertainty, including large sensitivity to initial conditions and physical assumptions (Blanchard-Wrigglesworth et al. 2015). In contrast to physics-based models, machine learning is emerging as a powerful tool for applications in the geosciences in cases where large volumes of data are available (Hsieh and Tang 1998; Toms et al. 2020). Machine learning predictions are driven by data and therefore do not depend on assumptions imposed on physical constraints. Although these constraints are crucial for some applications (e.g. where mass, heat, and momentum need to be conserved), in other applications they introduce additional uncertainty and complexity with little scientific benefit. While simple forms of machine learning (e.g. linear regression) have been commonly used in the geosciences, more advanced deep learning models (e.g. neural networks) have the potential to further elucidate physically meaningful relationships within data (McGovern et al. 2019; Toms et al. 2020). In this study, we assess the viability of using a neural network as a surrogate model to parameterize sea-ice motion in a numerical model setting on one-day time scales.

Machine learning models for sea-ice have been applied to improve estimates of ice properties from satellite remote sensing (Lee et al. 2016; Dumitru et al. 2019), to predict and understand sea-ice concentration on different time scales (Kim et al. 2020; Li et al. 2021; Andersson et al. 2021), and to make predictions of sea-ice motion (Petrou and Tian 2019; Zhai and Bitz 2021). ML models have been successful at improving predictions of sea-ice properties in comparison to state-of-the-art dynamical models. For example, the deep learning model IceNet outperformed the SEAS5 dynamical model from the European Centre for Medium-Range Weather Forecasts (ECMWF) for lead times greater than one month when making seasonal forecasts of summer ice (Andersson et al. 2021). Additionally, a CNN designed to make one-day predictions of ice motion showed higher correlations with satellite observations than CICE5, a leading physics-based model for sea ice (Zhai and Bitz 2021). The high performance of this CNN provides evidence that a CNN would be an effective surrogate model to replace the sea ice dynamical component of a numerical model for short-time-scale predictions. We build upon the work of Zhai and Bitz (2021) by further analyzing the nuances in the performance of a CNN in predicting ice motion, and by building the case for its use over a conventional linear regression approach.

We apply three different models, including persistence (PS), linear regression (LR), and convolutional neural network (CNN) to make predictions of sea-ice motion. In comparison to the other two models, a CNN has the benefits of incorporating spatial information and non-linear relationships between the inputs into its predictions. We build a CNN that has a similar architecture to that of Zhai and Bitz (2021) (differences are noted in the supplementary information, Table S1) and that is trained on the same input and output data. Our models show similar performances in making one-day predictions of sea-ice motion (Table S1). We expand on previous work by putting an emphasis on understanding the spatial and temporal variability in performance of the different models and how it is related to various properties of the ice. We divide the Arctic into four geographic regions (Fig. 1) based on the differences in skill between the CNN and LR models, and we analyze model performance within each.

2. Background

Sea-ice motion, as described by the momentum equation (Equation 1), is determined from a balance of the momentum tendency ($\frac{D}{Dt}(m\vec{u})$) with drag from the atmosphere ($\vec{\tau}_a$) and ocean ($\vec{\tau}_w$), the Coriolis force ($mf\hat{k}\times\vec{u}$), the ocean surface tilt ($mg\nabla H$), and the internal ice stress ($\nabla\cdot\sigma$) (Olason and Notz 2014; Feltham 2008). The term on the left represents the total derivative of mass, m times velocity, $\vec{u}$:

$$\frac{D}{Dt}(m\vec{u}) = \vec{\tau}_a + \vec{\tau}_w - mf\hat{k}\times\vec{u} - mg\nabla H - \nabla\cdot\sigma. \tag{1}$$

Changes in external forcing (i.e. winds, currents, radiation, etc.) influence the geometric and mechanical properties of the ice (thickness distribution, mass, strength, drag coefficients, etc.), which ultimately impact ice motion and deformation (Untersteiner et al. 2007). The American-Canadian Arctic Ice Dynamics Joint Experiment (AIDJEX) of 1970-1978 was one of the first major studies aimed at developing a comprehensive model of sea-ice motion under the influences of the ocean and atmosphere (Maykut et al. 1972; Untersteiner et al. 2007). Using data from the AIDJEX experiments, Thorndike and Colony (1982) introduced a relationship between sea-ice velocity and geostrophic wind that explained up to 70% of the variance in sea-ice velocity in the

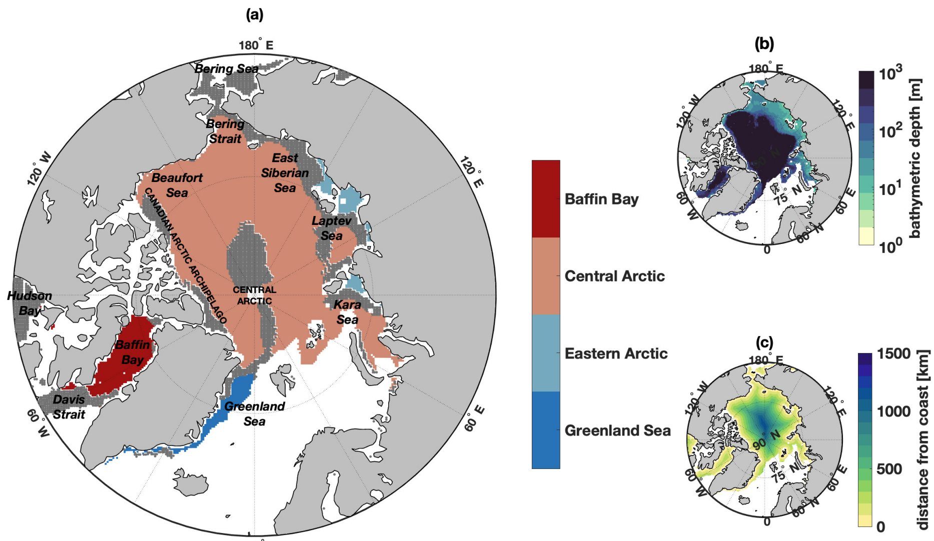

FIG. 1. Maps showing (a) spatial divisions (Greenland Sea, Eastern Arctic, Central Arctic, and Baffin Bay), (b) bathymetric depth [m] (note logarithmic scaling), and (c) the distance from coast [km]. Spatial divisions are based on overall performance of the CNN model and the difference between the performance of the CNN and LR models. The four divisions represent regions of: variable model performance and $corr_{LR} \gg corr_{CNN}$ (Greenland Sea, dark blue), low model performance and $corr_{LR} > corr_{CNN}$ (Eastern Arctic, light blue), high model performance and $corr_{LR} < corr_{CNN}$ (Central Arctic, light red), and variable model performance and $corr_{LR} \ll corr_{CNN}$ (Baffin Bay, dark red). Gray shading represents areas where the difference in correlation between the CNN and LR is not statistically significant or areas that are not included within this analysis. Data are not shown in regions where the ice concentration is zero or the satellite retrievals are absent for more than 20% of the year.

central Arctic. This relationship describes ice that is subject to high wind speeds on time scales of days to months. In this relationship, sea-ice velocity is related to geostrophic wind velocity through a speed reduction factor (the wind factor) and a turning angle, after removal of the long-term mean ice velocity field. In the absence of a steady ocean current, sea-ice moved about 8° to the right of the geostrophic wind at about 0.008 times the speed. This model is less successful for areas within 400 km of the coast, where stress gradients within the ice become more important due to the restriction of ice motion by geographical features (Thorndike and Colony 1982).

The internal stress gradient also depends on factors including the magnitude of the wind speed, ice concentration, and ice thickness. Ice with high values for thickness and concentration may have large stress gradients, which can result in a smaller dependence on wind. Conversely, ice with smaller stress gradients (low thickness and concentration) is found to have higher dependencies on wind (Hibler 1979). Decreases in correlation between wind and ice motion near the coast have often been attributed to ice stresses (Thorndike and Colony 1982; Kimura and Wakatsuchi 2000; Hibler 1979).

A relationship between ice motion and geostrophic wind was also examined by Kimura and Wakatsuchi (2000) and by Maeda et al. (2020), using sea-ice motion derived from satellite products and geostrophic wind derived from the sea-level pressure data from ERA Interim Reanalysis data produced by European Centre for Medium-Range Weather Forecasts (ECMWF) on 2.5° and 0.75° grids, respectively. In these studies, geostrophic wind was generally found to explain 70% of the variance in sea-ice velocity, with 60–90% of the variance explained in the central Arctic, and up to 40% in coastal regions (Fig. 3 in Maeda et al. (2020)). In addition to spatial variability, seasonal variations in the speed reduction factor and turning angle have been reported (Thorndike and Colony 1982; Kimura and Wakatsuchi 2000; Kwok et al. 2013; Maeda et al. 2020).

3. Data

In our analysis, models are trained to make one-day predictions of sea-ice velocity given present-day wind velocity, previous-day sea-ice concentration, and previous-day sea-ice velocity from various satellite and reanalysis sources, during 1989–2021. Using present-day wind as a predictor of present-day sea-ice velocity incorporates information that gives the model intrinsic skill. This approach is appropriate for the objective to make predictions on one-day time scales. We opt not to detrend to avoid contaminating the data with spurious removals. However, we do find that the model performance does not have any significant changes when run on data with the seasonal cycle removed (not shown). Processed data and methods for obtaining and processing raw data are made available by (Hoffman et al. 2023).

The ice velocity and concentration data are available from 25 October 1978 to 31 December 2021. However, evaluation of the uncertainty metrics for the Polar Pathfinder ice motion product shows a change in the error fields starting in the summer of 1987 (Figure S1) due to a difference in the sampling period when switching from using Scanning Multichannel Microwave Radiometer (SMMR, 48hr sampling period) to Special Sensor Microwave/Imagers (SSM/Is, 24hr sampling period) for brightness temperature (Tschudi et al. 2020). Additionally, ice concentration data from the Nimbus-7 passive microwave are only available every other day until 1987, and there is a gap in availability of the sea ice concentration data from 03 December 1987 to 12 January 1988. Thus, for consistency in the stability of the observation systems and the quantity of data used from each year, we use data from 1989-2021 to build our models. We use the satellite and reanalysis sources discussed below for consistency with Zhai and Bitz (2021). However, in comparison, we make a slight extension to the temporal subset of data over which the model is trained and tested.

a. Sea-Ice Velocity: Polar Pathfinder Version 4 Daily Sea Ice Motion vectors (PP)

The Polar Pathfinder product (PP; Tschudi et al. 2019) provides daily sea-ice motion vectors at a spatial resolution of 25 km in the Equal-Area Scalable Earth (EASE)-grid. The EASE-grid was defined by the NOAA/NASA Polar Pathfinder Program to support standardized spatial comparisons from gridded, satellite microwave data. In polar regions, the EASE-Grid takes the form of Lambert azimuthal equal-area projections that accurately represent area in all regions of the global sphere. (Brodzik et al. 2012). This data set is informed by optimal interpolation of a combination of observations from passive microwave inputs, buoys, and NCEP/NCAR reanalysis winds. The PP dataset relies on wind because during the summer, passive microwave and buoy sources become unreliable for melting ice (Tschudi et al. 2020). For wind-derived ice motions, ice is assumed to move at $\sim$1% of the wind speed and in the direction of the wind, based on the estimate from Thorndike and Colony (1982). An estimated uncertainty map is also provided, which we use for comparison when evaluating our models. We were unable to obtain a dataset that is independent from the PP product to validate the use of the PP for this case. We did find high correlation between the PP and the Ice-Tethered Profiler data (not shown), but these observations were used to create the PP product. Wang et al. (2022) found the PP to have low accuracy in speed, but high accuracy in angle in comparison to eleven other satellite products when evaluated against measurements from buoys from the International Arctic Buoy Program (IABP) and the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC).

b. Sea-Ice Concentration: Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data

The passive microwave sea-ice concentration product (Cavalieri et al. 1996) is generated from brightness temperature data derived from various sensors (SMMR, DMSP and SSM/I-SSMIS). This product provides daily measurements of sea-ice concentrations (fraction of ocean area covered by sea ice in each grid cell) in a 25×25 km polar stereographic projection. Here we re-grid to the 25-km EASE-grid for consistency with other ML model inputs. An intercomparison study of 10 satellite passive microwave sea-ice concentration data sets by Kern et al. (2019) found that while the Nimbus-7 product used in this work showed the largest difference between other products, all 10 products compared reasonably well to ship-based observations. Additionally, the Nimbus-7 product used in this study showed less than a 7% deviation from all other products from November–June, and less than a 15% deviation from July–October when comparing the monthly mean values of sea-ice concentration among the 10 products from June 2002 to September 2011. The product used in this study was also found to have a negative bias in sea-ice concentration throughout the Arctic in comparison to the ensemble mean of the 10 products (Fig. 8 from Kern et al. (2019)). While this negative bias was particularly large in the peripheral seas, it was close to zero (i.e. $<$ 6%) in the region of study of this work.

c. Wind Velocity: Japanese 55-year Reanalysis derived for ocean-ice models (JRA55-do)

The Japanese Meteorological Agency 55-year atmospheric reanalysis based surface dataset for driving ocean-sea ice models (JRA55-do) is used to prescribe wind velocity (Tsujino et al. 2018). Based on the JRA55 (Kobayashi et al. 2015), the JRA55-do is derived for use in ocean simulations, with surface fields adjusted relative to satellite climatological winds (SSM/I and QuikSCAT) using a spatially varying wind factor for wind speed and EOF analysis for wind direction (Tsujino et al. 2018). The JRA55-do better matches satellite wind fields in coastal areas than do other reanalysis products (Taboda et al. 2019). The JRA55-do provides 3-hourly estimates of total wind velocity at 10 m with a horizontal resolution of ~55 km. Here we calculate daily average wind vectors and re-grid to the 25-km EASE-grid.

d. Bathymetric Depth: International Bathymetric Chart of the Arctic Ocean (IBCAO)

We use bathymetric depth from IBCAO (Jakobsson et al. 2020) for comparisons of model performance after training. We make use of the Version 4.2 product without elevation data for the Greenland Ice Sheet on a 400 m × 400 m grid cell spacing, re-gridded to the 25-km EASE-grid.

4. Methods

a. Model Inputs

We employ a suite of machine learning and statistical models (PS, LR, and CNN) to predict present-day sea-ice velocity components ($u_{i,t}$ & $v_{i,t}$) using the input parameters:

- present-day zonal & meridional wind velocity ($u_{a,t}$ & $v_{a,t}$),
- previous-day zonal & meridional sea-ice velocity ($u_{i,t-1}$ & $v_{i,t-1}$), and
- previous-day sea-ice concentration (c_{t-1}).

Inputs are chosen based on results from Zhai and Bitz (2021), who showed that the above combination of parameters produced skillful output when used to predict sea-ice motion with a CNN. Sea-ice velocity might be expected to be dependent also on sea-ice thickness, in addition to our selected input fields (Hibler 1979; Thorndike and Colony 1982). However, feature exploration studies of CNN models applied to Community Earth System Model version 2 (CESM2) output by Zhai and Bitz (2021) found that the inclusion of sea-ice thickness as an input parameter does not greatly impact the overall skill and correlation of CNN predictions. Fortunately the thickness is not an important input, as satellite observations of sea-ice thickness prior to 2019 have a high uncertainty, are discontinuous in time and unavailable during the summer. Therefore this parameter is omitted from our analyses. We note that efforts are being made to extend the CryoSat-2 sea-ice thickness record back in time using machine learning techniques (Landy et al. 2022). However, these data are available bi-weekly and thus do not meet the requirements of this study for daily data.

Inputs are taken from satellite and reanalysis sources listed in section 3. All variables are normalized to zero mean and one standard deviation before being input into the models, based on the global statistics of the entire record used here from 1989-2021. Data are broken up into train, validation and test data sets with an 88%–6%–6% split (e.g. train with years 1989–2017, validate with years 2018–2019 and test with years 2020–2021). The train, validate, and test years are shuffled ten times to produce data for ten different ensemble runs for each ML model. We refer to an "ensemble run" as a run that is trained on a different temporal subset of data. We calculate performance statistics (discussed in section 4.c) for each ensemble run and average over the ten runs for final results. A CNN requires inputs to be of consistent size, with consistent spatial and temporal coverage, and without non-numerical (e.g. not-a-number or 'NaN') values. Thus, while it may make sense to remove data in regions where sea-ice motion data are not available (i.e. sea-ice concentration is zero or there is land) before training, due to the practical constraints of applying a CNN sea-ice velocity components are set to zero during training. A time-variable mask is used to remove these sea ice free points during model evaluation. Additionally, while uncertainty metrics are available for the Polar Pathfinder sea-ice motion product, we do not mask out any points during training due to the constraints of CNN models listed above. We note that taking uncertainty into account during training of PS and LR models is possible, but to maintain consistency between models we leave that for future work.

b. Model Setup

We compare prediction outputs from three different models: PS, LR, and CNN.

1) Persistence and Linear Regression Models

PS predicts the present-day sea-ice velocity to be the same as the previous day at each grid point (Equation 2):

$$u^*_{i,t} = u^*_{i,t-1}. \tag{2}$$

This offers a baseline measure of the variability of the system and of the minimum skill that any alternative models should attain. Here the vector u^*_i is a complex number, where the real and imaginary parts are the zonal and meridional components of the sea-ice velocity vector.

LR regresses each of the five input parameters (section 4.a) onto the sea-ice velocity components (Equation 3):

$$u^*_{i,t} = Au^*_{a,t} + Bu^*_{i,t-1} + Cc^*_{i,t-1} + D \tag{3}$$

Given inputs and outputs, LR solves for parameters A to D. In equation (3), A to D are complex constants, and the vectors u_i^*, u_a^* and c_i^* are complex numbers, where the imaginary part of c_i^* is set to zero. LR is carried out in two different manners: one is performed globally (LR-g), and uses each time snapshot as an independent sample for fitting, providing one equation for the entire modeled region in the Arctic; the other is performed grid-wise (LR), leading to a different regression equation for each grid point. For both LR configurations we employ ridge regression with a ridge parameter of $\lambda = 10^{-2}$ to limit the magnitude of the regression coefficients and prevent them from being unrealistically large (Marquardt and Snee 1975). The value of the ridge parameter is chosen based on the iterative approach in Marquardt and Snee (1975) where we make step changes from small to large values of λ and pick the value of λ for which the LR coefficients stabilize (i.e. are not infinitely large). We also note that data are not removed from the training set when $c_i = 0$, which may dampen the wind dependence in LR because the model is trained that $u_i = 0$ when $u_a \neq 0$ in these locations. As discussed in section 4.a, these data are not masked during training because the CNN requires numerical values (i.e. not 'NaN').

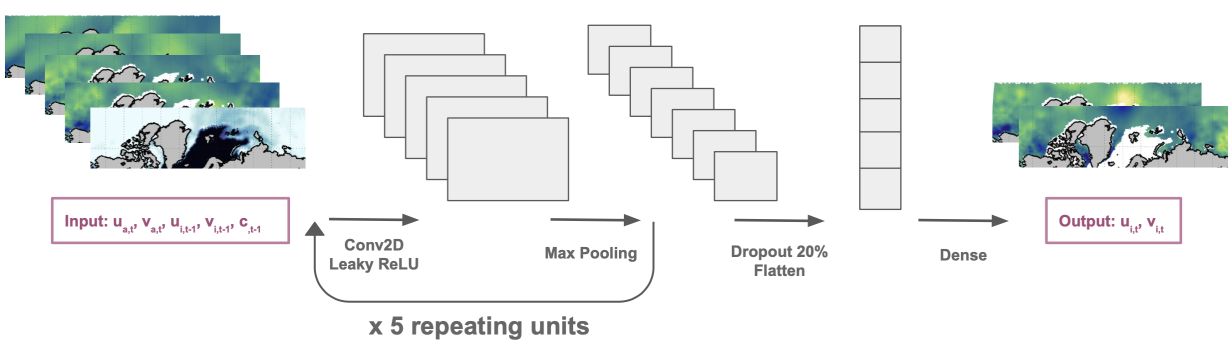

Fig. 2. Schematic of the convolutional neural network (CNN) used in this study for predicting present-day sea-ice velocity components, $u_{i,t}$ & $v_{i,t}$ (outputs), from present-day wind velocity, $u_{a,t}$ & $v_{a,t}$, previous-day sea-ice velocity, $u_{i,t-1}$ & $v_{i,t-1}$, and previous-day sea-ice concentration, $c_{i,t-1}$ (inputs). This CNN has five repeating units of a 2D convolution with a ReLU activation and max pooling, followed by a 20% droput layer, flattening and a dense layer.

2) Convolutional Neural Network (CNN) Architecture

A CNN is a type of ML model typically applied to visual images, where a computer is fed numerous (hundreds to millions) different images and learns from their patterns in order to make a prediction (O'Shea and Nash 2015). We use data sets that are image-like in that they have a specified value at various grid-points on a map (for images this would be colors at various pixel locations). Incorporation of spatial information when making predictions is one of the benefits of CNN over LR or PS models, in addition to the ability of a CNN to capture non-linearities in the relationships between the input predictors and the outputs. Our CNN (Fig. 2) is set up with five repetitions of the block unit: 2D convolution, ReLU (Rectified Linear Unit), and 2D max-pooling. This is followed by a 20% drop-out layer, a flattening to a one-dimensional vector, and finally a regression onto a 1D vector (dense layer) representative of the output predictions. This output is then concatenated into two maps of present-day zonal and meridional sea-ice velocity.

We implement the CNN in python using the Tensorflow/Keras library (Abadi et al. 2015). Convolutional and ReLU layers are carried out with (1,1) strides and (3,3) filter sizes, whereas the max pooling strides and filter sizes are (2,2). For each of the respective repeating block units, there are 7, 14, 28, 56, and 112 filters. The training runs for 50 epochs with a batch size of 365 days. Optimization is carried out with an Adam optimizer and a normalized root mean squared error as the loss function (second term in Equation 5 discussed below). Similarly to the LR, we employ ridge regression with a ridge parameter of $\lambda = 10^{-2}$. Further descriptions of the architectural components of a CNN (i.e. layers, strides, filters, ReLU, max pooling, etc.) can be found in O'Shea and Nash (2015). Filter sizes are chosen based on the conventional VGGNet architecture (Szegedy et al. 2015). We do not carry out hyperparameter tuning for this study in order to maintain consistency with the architecture of Zhai and Bitz (2021), with the only differences being in the sizes and number of the filters due to differences in the sizes of the starting input maps.

c. Model Evaluation

As in Zhai and Bitz (2021), the model performances are evaluated and compared based on the correlation (Corr) and skill, given by:

$$corr_{x,y} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2}\sqrt{\sum_i^n (y_i - \bar{y})^2}}, \tag{4}$$

$$skill_{x,y} = 1 - \frac{\sqrt{\overline{(x_i - y_i)^2}}}{\sqrt{\overline{(x_i - \bar{x})^2}}}, \tag{5}$$

where x represents observations, and y represents predicted values of a sample size n. The correlation (Equation 4) is defined as the covariance between prediction and observation scaled by their standard deviations. The skill (Equation 5) is a representation of the fraction of the observed standard deviation explained by the model predictions, where the second term is the root mean squared error normalized by the standard deviation of the observations (Thomson and Emery 2014). The correlation ranges from -1 to 1, with 1 indicating a perfect positive relationship, -1 indicating a perfect negative relationship, and zero representing orthogonality. The skill can range from negative infinity to 1, with 1 representing a perfect match between model predictions and observations. The correlation is a measure of how well the phase variability in the data is explained by the model, whereas the skill is a measure of the absolute error in the model predictions.

These metrics are calculated using the test data set (varying years, as discussed in section 4.a) of which the models have no prior knowledge. Two different masks are made and both applied to the data during model evaluation: one is time-variable and evaluates model outputs only at times and in locations where sea-ice concentration is greater than zero; the other is constant with time and masks out all areas where sea-ice concentration is zero more than 20% of the time from 1992–2017. Metrics are calculated overall (section 5.a.1), at each grid point to provide spatial evaluation (section 5.a.2), over each month for temporal evaluation (section 5.a.3), and for different percentile ranges of various sea-ice properties (wind speed, ice speed, and ice concentration) to understand the role these play on the model performance (section 5.a.4). For temporal evaluations we calculate the monthly mean for each of the ten ensemble runs. Overall monthly means are then represented by the mean of the ten ensemble runs, and monthly errors are calculated as the standard error of the mean of the ten ensemble runs (as discussed in section 4.a). Temporal evaluations are carried out for for different regions within the Arctic. The divisions (Fig. 1a) are made based on spatial distributions of model performance metrics ($corr_{CNN}$ and $corr_{CNN} - corr_{LR}$ in Fig. 3c & f), representing regions of: (i) variable model performance and LR greatly outperforming CNN (i.e. variable $corr_{CNN}$ and $corr_{LR} \gg corr_{CNN}$; Greenland Sea, dark blue), (ii) low model performance and LR slightly outperforming CNN (i.e. low $corr_{CNN}$ and $corr_{LR} > corr_{CNN}$; Eastern Arctic, light blue), (iii) high model performance and CNN slightly outperforming LR (ie. high $corr_{CNN}$ and $corr_{LR} < corr_{CNN}$; Central Arctic, light red), and (iv) variable model performance and CNN greatly outperforming LR (i.e. variable $corr_{CNN}$ and $corr_{LR} \ll corr_{CNN}$; Baffin Bay, dark red).

d. Model Comparison

We also investigate the correlation and skill differences between the LR and CNN models, which requires an understanding of where the differences are significant. Significance tests on the differences are approximated with a cross-validated t test (Dietterich 1998; Tang et al. 2000). The cross-validated t test proceeds as follows: (i) for each of the ten ensemble runs, the correlation and skill for the LR and CNN are calculated for each grid point or percentile range for a given variable and transformed by Fisher's z transform (Equation 14.5.6 in Press et al. 1986) to remove skewness in the distribution; (ii) the difference between the transformed correlation and skill for the two models is calculated and averaged over the ten ensemble runs; (iii) a two-tailed t test is performed to detect whether the mean difference between the two models is significantly different from zero at the 95% confidence level. The cross-validated t test uses the degrees of freedom to calculate significance. For spatial comparisons (section 5.a.2) we estimate degrees of freedom using the temporal decorrelation scale to estimate the number of independent time series of sea-ice motion in the Arctic. This temporal decorrelation scale is taken as the e-folding scale of a Gaussian fit to the autocorrelation of the sea-ice speed calculated at different time-lags (Equations 10 and 11 in Sumata et al. 2018).

e. Analysis of Inputs

We analyze the spatial and temporal variability of different parameters related to ice motion (wind speed, u_a; ice speed, u_i; and ice concentration, c_i) to assess how the model performance compares to the model inputs. Spatial analyses look at maps of the average and standard deviation of each parameter over time from 1989–2021. This type of analysis is useful for comparing these properties to maps of the model performance metrics in order to understand different regimes within the Arctic. We also look at the seasonality of each of these properties. Similarly to the input analysis in section 4.c, monthly errors are calculated as the standard error of the mean of the ten ensemble runs, and temporal evaluations are carried out for different divisions that are chosen based on the model performances.

5. Results

a. Model Performance

1) Overall

We evaluate the overall performance of the different models by calculating the correlation and skill over all gridpoints and times (Table 1). The CNN has the highest correlation and skill, followed closely by the grid-wise linear regression (LR). The grid-wise LR largely outperforms the global LR (LR-g) that covers the entire Arctic, which is not much better than the simple PS model. These results confirm the advantage of using a model that captures non-linearity (CNN) and the heterogeneity of Arctic sea ice motion statistics (both CNN and LR). The better performance of the CNN, LR, and LR-g models in comparison to the PS confirms that sea-ice motion depends on wind and sea-ice concentration on daily time scales. Table 1 shows the pattern that an increase in model complexity leads to an increase in performance. Additionally, because correlation is a measure of how well the model is able to capture the phasing, while skill measures the model's ability to capture phasing and magnitude, the high correlation but lower skill suggests the models do well capturing the phasing but incur error in capturing the magnitude.

Model	Correlation	Skill
Persistence (PS)	0.69 ± 0.02	0.21 ± 0.02
Linear Regression, global (LR-g)	0.72 ± 0.01	0.30 ± 0.01
Linear Regression, gridwise (LR)	0.78 ± 0.02	0.37 ± 0.02
Convolutional Neural Network (CNN)	0.81 ± 0.01	0.42 ± 0.02

Table 1. Overall correlation and skill between observations and predictions of sea-ice velocity for four different models.

2) Spatial

Spatial variations in the correlation (Fig. 3) and skill (not shown) are similar for the PS, LR, and CNN models. Models perform well for predictions in the central Arctic, with decreasing performance in coastal locations. Low values of correlation (Fig. 3a–c) are visible in the Bering Strait, Bering Sea, Hudson Bay, East Siberian Sea, Laptev Sea, Kara Sea, and off the coast of

Greenland. Particularly poor model performances are found near the islands in the Eastern Arctic. The best model performance is seen north of Fram Strait and in the Beaufort Sea.

Typically, $corr_{CNN} > corr_{LR} > corr_{PS}$, similar to the results from section 5.a.1. The spatial differences in correlation between the models are shown in Fig. 3d–f. Regions in red indicate areas where the first model in the difference metric outperforms the second (i.e. $corr_{CNN} > corr_{PS}$ in Fig. 3d, $corr_{LR} > corr_{PS}$ in Fig. 3e, and $corr_{CNN} > corr_{LR}$ in Fig. 3f), whereas blue regions indicate the opposite (i.e. $corr_{CNN} < corr_{LR}$ in Fig. 3f). Gray regions show where the difference in correlation between the two models is not statistically significant. The CNN and LR models outperform the PS over the entire Arctic (Fig. 3d & e), with the exception of the western side of Baffin Bay where the PS outperforms the LR (blue). Overwhelmingly, the CNN outperforms the LR (red in Fig. 3f). Interestingly, the LR has a higher correlation (blue) in coastal regions where both models have decreased performance (i.e. near the islands in the Eastern Arctic and off the coast of Greenland).

The spatial patterns in model performance compared to the distance from the coast are confirmed in Fig. 4. Correlations for the CNN and LR models tend to be lower in coastal regions (Fig. 4a–b). This is also true for skill (not shown). For both models, locations that are greater than 400 km from the coast consistently have correlation greater than 0.7 (and skill greater than 0.3, not shown). The finding that the CNN outperforms the LR model for most cases is confirmed in Fig. 4c, where most of the data lie in the positive region (i.e. above the black line). Conversely, locations where the LR outperforms the CNN only occur within 400 km of the coast.

We also show that models have decreased performance in shallower regions (Fig. 4d–e). Overall, model performance increases with increasing seafloor depth. The relationship is logarithmic: performance increases rapidly with increasing depth for depths shallower than 1000 m, while the trend levels out for depths greater than 1000 m. Models exhibit correlations less than 0.7 and 0.5 (CNN and LR, respectively) only for locations with depths less than 1000 m. The CNN outperforms the LR for most cases (Fig. 4f). Most regions where the LR outperforms the CNN (below the black line) occur at depths shallower than 500 m, although there are some instances of higher correlation of the LR for greater depths.

We also analyze the spatial variability of the various properties related to sea-ice motion (wind speed, u_a; sea-ice speed, u_i; and sea-ice concentration, c_i). The mean and standard deviation of the properties listed above are mapped in Fig. 3g–l. Patterns in mean ice speeds tend to coincide with the spatial patterns in wind speed (Fig. 3 g & h), consistent with the known dependence of ice motion on wind speed (Thorndike and Colony 1982). Both ice and wind speed are relatively low in the coastal and island regions of the East Siberian Sea, the Canadian Arctic Archipelago, and off the northern and western coasts of Greenland. The highest mean wind speeds occur in the Davis Strait, off the eastern coast of Greenland, and in the Bering Strait; high mean ice speeds also occur in these regions, in addition to the Beaufort Sea. The region of low mean ice speeds to the north of the Canadian Arctic Archipelago coincides with high mean ice concentrations (Fig. 3h & i). Conversely, the region of low mean ice speeds in the East Siberian Sea coincides with lower mean ice concentrations.

Regions that show high variability (large standard deviations) in ice speed coincide with high mean ice speeds (i.e. in the Beaufort Sea, Baffin Bay, Davis Strait, and Greenland Sea), while regions with low variability coincide with lower mean ice speeds (to the north of the Canadian Arctic Archipelago and in the East Siberian sea) (Fig. 3 h & k). Variability in wind speed is found to be relatively consistent throughout the Arctic, with the exception of high variability off the eastern coast of Greenland (Fig. 3j). Regions with large variability in ice concentration typically correspond to regions with lower mean ice concentrations (i.e. in the East Siberian Sea, Baffin Bay, the Kara Sea, and the Bering Strait). These are the regions where the largest amount of seasonal ice melt typically occurs (not shown), which contributes to the large variability and lower mean ice concentrations.

3) Temporal

For the region containing the entire Arctic, the CNN typically has the highest correlation, followed by the LR and then the PS model (Fig. 5a). During June–September the difference in correlation between the CNN and LR models is not statistically significant. Temporal structure is visible in the correlation for all of the models. The LR model performance (Fig. 5a) has a larger range of seasonal variability than the other two models. Maximum correlation and skill for the PS and CNN models occurs during October–December, while the LR has a correlation maximum in June–August. All three models experience a minimum performance in April.

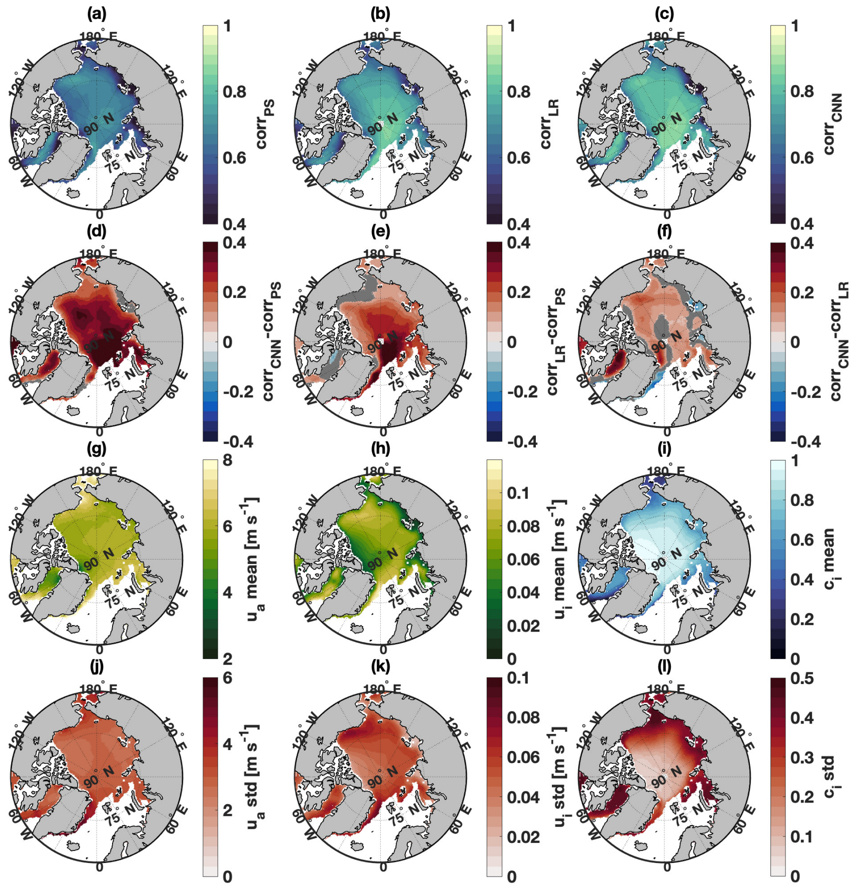

FIG. 3. (a–c) Mapped correlation for predictions of sea-ice velocity made by the (a) PS, (b) LR, and (c) CNN models. (d–f) The difference in correlation between models: (d) $corr_{CNN}$-$corr_{PS}$, (e) $corr_{LR}$-$corr_{PS}$, and (f) $corr_{CNN}$-$corr_{LR}$. The gray regions in d–f represent locations where the difference in correlation between the two models is not statistically significant. (g–i) Mean and (j–l) standard deviation in time of various properties related to sea-ice motion from satellite and reanalysis products(wind speed, u_a (g & j); sea-ice speed, u_i (h & k); and sea-ice concentration, c_i (i & l).

The temporal evaluations are divided into regions (Fig. 1a) based on the spatial variability of their performance, as discussed in section 4.c. The impacts of this spatial division on model performance are shown in Fig. 5b–d, while Fig. 5e–g represent differences in the correlation between the different models. Here the black lines represent metrics calculated with all of the data included, and the different shades of red and blue represent the respective spatial regions from Fig. 1a. Diamonds in Fig. 5e–g indicate months where the difference between the two models is statistically significant. The correlation for the region within the Central Arctic division (light red)

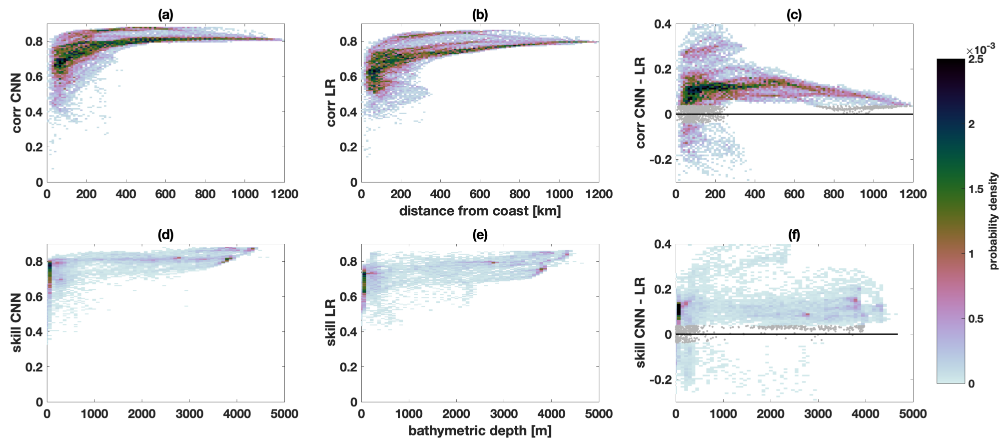

FIG. 4. PDFs for model performances compared to their distance from the coast (a–c) and bathymetric depth (d–f), with (a & d) for the CNN and (b & e) for the LR. (c & f) The difference ($corr_{CNN}$-$corr_{LR}$) between the correlation of the two models. Gray shading (c & f) represents correlation differences between the two models that are not statistically significant. Results for skill (not shown) are similar to correlation.

does not deviate much from that of the entire Arctic (black) because the Central Arctic region is large and covers most of the region containing the entire Arctic. However, there are significant changes in monthly values of correlation for all other divisions (Greenland Sea, Eastern Arctic, and Baffin Bay divisions). For all three models, the Eastern Arctic (light blue) division exhibits a similar seasonal cycle to the entire Arctic (i.e. minimum correlation in March–April), but has a consistently lower monthly correlation in comparison to the other divisions for all models, except during the months of July–October.

The Greenland Sea (dark blue) and Baffin Bay (dark red) divisions exhibit a relatively high correlation from October–May that decreases toward a minimum in August or September (Fig. 5b–d). The Greenland Sea division (dark blue) has a higher correlation than the other divisions from October–April for all three models. The Greenland Sea division shows a lower correlation than the region containing the entire Arctic from the months of June–September, reaching a minimum in August for all three models that is significantly lower than correlations for the entire Arctic (i.e the CNN has a minimum of 0.54 for the Greenland Sea division in comparison to 0.80 for the overall Arctic). The Baffin Bay division (dark red) exhibits the largest deviations in correlation from the overall Arctic for all models, showing up as a large decrease during the months of May–November. The Baffin Bay division has higher correlations in December–April, and the lowest

August–September minimum out of all of the divisions for all models (i.e. the August correlation of the CNN within the Baffin Bay division is 0.28 in comparison to 0.80 for the entire Arctic). The performance minima that occur in August–September for the Greenland Sea and Baffin Bay regions are much lower than the April minima for the region containing the entire Arctic. This pattern of decreased model performance during months of minimum sea-ice extent (Greenland Sea and Baffin Bay divisions) suggests a link between model performance and sea-ice concentration, which will be further evaluated in section 5.c.

The differences in correlation between the models for the different divisions are shown in Fig. 5e–g. The LR and CNN typically outperform the PS for all divisions (i.e. diamonds indicating statistically significant difference in model performance are above zero in Fig. 5e & f). The LR outperforms the CNN in all months for the Eastern Arctic division. However, statistically significant differences from zero are only present December–May. The CNN outperforms the LR during the months of September–May for the Central Arctic, and September–June for the Baffin Bay division. However, the difference between the correlation of the CNN and LR is not statistically significant during the months of June–October for the Central Arctic or July and September–November for the Baffin Bay division. These differences in model correlation will be further analyzed in section 5.c.

We also compare the temporal variability in performance to that of the various properties related to sea-ice motion (wind speed, u_a; sea-ice speed, u_i; and sea-ice concentration, c_i). The ensemble mean monthly averages of various properties related to sea-ice motion are shown in Fig. 5h–j. Analysis is further broken down into the four divisions within the Arctic, which are chosen based on values of the model correlations (Fig. 1).

For all regions the seasonal cycles for ice speed and wind speed (Fig. 5h & i) generally line up, with minima typically occurring during the summer months and maxima in the winter. The seasonal pattern of minimum wind speeds occurring from June–July, and maximum speeds anywhere from October–February is consistent throughout all regions, except for the Eastern Arctic division where minima are found in December–March, and maxima occur in September–October. The Greenland Sea division has greater seasonal variability in wind and ice speeds than the other divisions, with comparatively high maximum speeds in November–April. Seasonal patterns in ice speed show minima in June–July for the Central Arctic, June–August for the Greenland Sea, and May–

October for the Baffin Bay division. The Eastern Arctic division shows the opposite seasonal trend, instead exhibiting minimum ice speed from December–May. Sea-ice concentration also follows a seasonal cycle within each division, typically reaching a maximum in March and a minimum in September (Fig. 5j). The Baffin Bay division exhibits the lowest and longest duration minimum ice concentration (i.e. $c_i < 0.5$ from July–October). From December–May the Greenland Sea division has a lower ice concentration than the other divisions, which are all similar during this time.

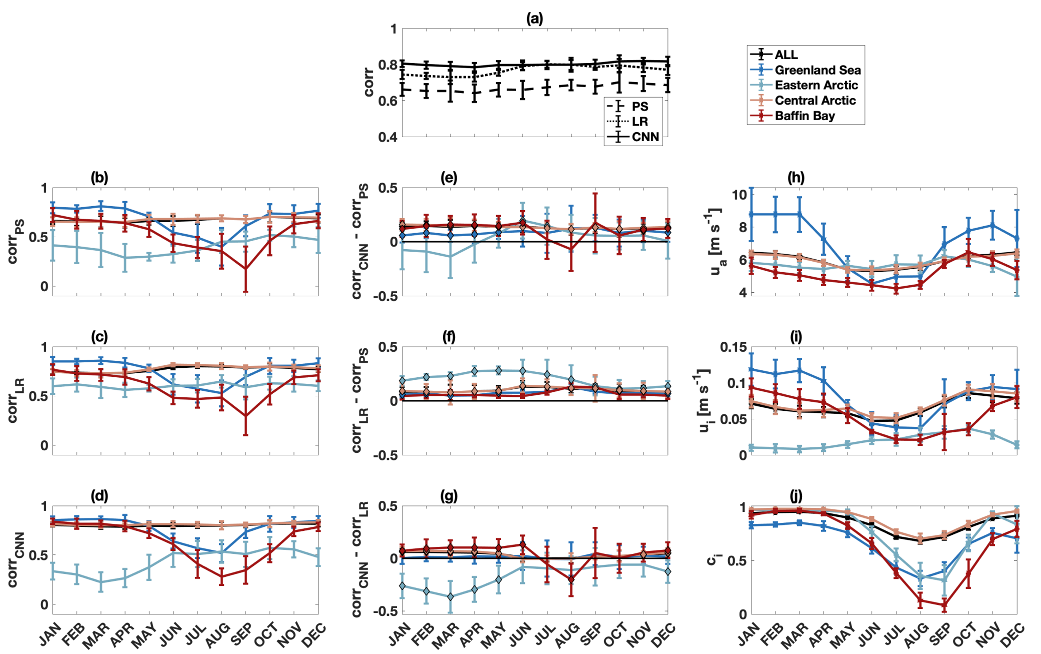

FIG. 5. (a–d) Ensemble mean monthly correlation for the prediction of sea-ice velocity by three different models: (a) all models, (b) persistence, (c) linear regression, and (d) CNN. (e–g) The difference between the correlation of the (e) CNN and PS, (f) LR and PS, and (g) CNN and LR models. (h–j) Ensemble mean monthly values of various properties related to sea-ice motion (wind speed, u_a (h); sea-ice speed, u_i (i); and sea-ice concentration, c_i (j). Metrics are calculated for five different regions: containing the entire area of the Arctic (black), and within the spatial divisions indicated in Fig. 1a (shades of red and blue). Error bars represent ensemble mean standard deviations. Diamonds in Fig. 5e–g indicate months where the difference between the two models represented is statistically significant.

4) Model Performance for Percentiles of Inputs

The model performance is compared to properties related to sea-ice motion (wind speed, u_a; sea-ice speed, u_i; and sea-ice concentration, c_i) to probe the variability in model correlation in space and time. Figure 6 shows the correlation metrics calculated from subsets of test data for all models (PS in dark blue, LR in teal, and CNN in green). Subdivisions are based on percentile ranges (5% intervals) of the various properties. The performance metrics (correlation (a–c) and the difference in correlation between the various models (d–f)) are plotted against the average of each percentile range (i.e. 0–5%, 5–10%, etc.) for each property. Skill metrics (not shown) have similar patterns to the correlation. We find that the correlation increases with increasing wind speed, sea-ice speed, and sea-ice concentration for all models (Fig. 6a–c). These relationships have statistically significant r^2 values when fit to a second-order polynomial with a least squares regression.

The CNN and LR consistently outperform the PS model, as these two difference metrics ($corr_{CNN}$-$corr_{PS}$ & $corr_{LR}$-$corr_{PS}$) are positive for all u_a, u_i, and c_i (blue and teal lines in Fig. 6 d–f). The CNN has a higher correlation than the LR (green lines in Fig. 6 d–f), except for the case where $c_i < 0.5$ (Fig. 6f). The metrics for the difference between the CNN and the other two models (i.e. $corr_{CNN}$-$corr_{PS}$ & $corr_{CNN}$-$corr_{LR}$) have statistically significant relationships with wind speed, ice speed, and ice concentration: the difference between the two models decreases for increases in wind and ice speed (Fig. 6d & e), and increases with increases in ice concentration (Fig. 6f). The difference metric $corr_{LR}$-$corr_{PS}$ shows a similar relationship to u_i, but not u_a or c_i. Additionally, the difference between the CNN and the LR is less dependent on u_i than the other two difference metrics (i.e. the slope of the green line is less than the slopes of the teal and blue lines in Fig. 6e). This can be attributed to the correlation of the PS model being much lower than that of the CNN or LR when ice speeds are close to 0 m s $^{-1}$. The results in Fig. 6d–f are robust whether we use all data or remove non-significant points.

b. Linear Regression Parameters: Relationship Between Sea-Ice Motion and Input Parameters

Analysis of the linear regression parameters provides insight on the locations where each of the inputs is important for predicting sea-ice motion. The parameters from the full LR (A-C in Equation 3) described in section 1 are mapped in Fig. 7. Here Fig. 7a–c represents the magnitude

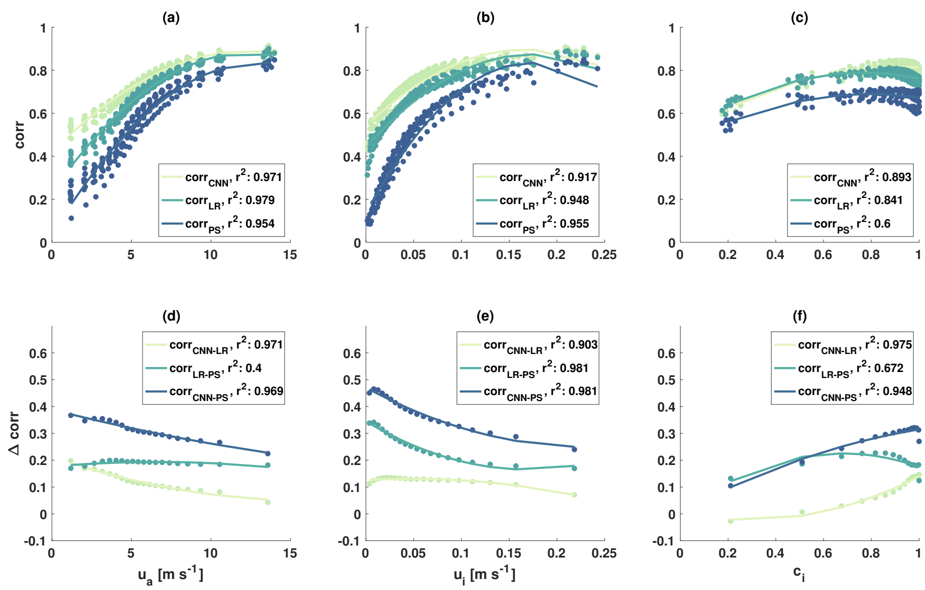

FIG. 6. Correlation of the CNN (a–c) and the difference between CNN and LR correlation (d–f) as a function of various properties related to sea-ice motion (wind speed, u_a (a & d); sea-ice speed, u_i (b & e); and sea-ice concentration, c_i (c & f). The correlation is calculated with subsets of test data based on percentiles (5 percent intervals) of the various parameters. The x-axis represents the mean value of the data in each 5% interval of each parameter. Correlation differences (d–f) that are not statistically significant are not shown.

of the regression coefficients for normalized wind speed, sea-ice speed, and sea-ice concentration on the sea-ice velocity (i.e. $\sqrt{\Re^2 + \Im^2}$ of A to C, where $\Re$ and $\Im$ represent the real and imaginary components of these coefficients). These values range between 0 and 1 in the figure because they are normalized to the maximum overall coefficient. Larger values indicate that sea-ice velocity has a larger linear dependence on a particular parameter.

Results show that wind speed has the largest importance in predicting sea-ice velocity within the Central Arctic (Fig. 7a). Near the coast, the LR coefficient for previous-day sea-ice velocity is elevated (Fig. 7c) complementary to the high values in the interior for wind speed (Fig. 7a). Fig. 7d–e represents the rotation angles of the wind and sea-ice velocity to the predicted next-day sea-ice velocity. The wind angle has an average of 24.9° ± 11.3° throughout the Arctic, which is fully consistent with Nansen's observations aboard the Fram of angles between 20 and 40° (Ekman 1905), and falls within one standard deviation of previous research (Thorndike and Colony 1982; Serreze et al. 1989; Maeda et al. 2020) who found wind angles of -5 to 18°, 0 to 19°, and -10 to

30° (depending on season; winter to summer), respectively. The spatially averaged angle between present and previous-day sea-ice speed is -8.3° ± 6.4°, with spatial variations as seen in Fig. 7e. When looking at the data, the expected spatial mean of the angle difference between previous and present-day sea-ice velocity is 0.2° (not shown), which is within two standard deviations of the angle found from the LR parameters.

Wind velocity is found to have the maximum LR coefficient for predicting sea-ice velocity throughout the Central Arctic (dark blue in Fig. 7f). Locations near the coast are dominated by the sea-ice speed (pink regions). This is consistent with results from previous studies (Thorndike and Colony 1982; Kimura and Wakatsuchi 2000; Maeda et al. 2020) that conclude that the dependence of sea-ice velocity on wind velocity is not as strong in coastal locations where ice stresses become more important. Additionally, the low coefficient for wind velocity found in the Fram Strait off the east coast of Greenland, where the transpolar drift acts as a strong and persistent export pathway for Arctic sea-ice (Weiss 2013), has previously been attributed to strong surface ocean currents (Kimura and Wakatsuchi 2000).

The LR coefficient for wind speed is related to the spatial patterns in the mean c_i (Figs. 7a & 3i). We find low values for the LR parameter for wind speed in the Canadian Arctic Archipelago, a region where c_i is high and has little temporal variability (Fig. 3i &l), which is consistent with results from Kimura and Wakatsuchi (2000); Maeda et al. (2020). However, regions of low mean c_i often have smaller values for the LR wind coefficient (i.e. coastal regions in the eastern Arctic, Baffin Bay, and the Bering Strait). This contradicts results from Kimura and Wakatsuchi (2000); Maeda et al. (2020), where areas with high ice concentration exhibit a relatively small wind factor as a result of internal stresses becoming more important in regions where ice is thick and concentrated. However, we note that in contrast to Kimura and Wakatsuchi (2000), our model also includes u_i as a predictor, which increases in importance near the coast. Additionally, our analysis has one LR coefficient at each spatial location throughout all time from 1992–2017, which provides a description of the relationship between the wind factor and the average c_i at each location. In contrast, Maeda et al. (2020) have a different LR equation for each month, providing a better picture of the relationship between the wind factor and the instantaneous c_i, which is more likely to display impacts of ice stresses.

Values of the LR coefficients are related to the performance of the LR model and to the difference between the CNN and LR model performance. Figure 8 shows the relationship between the LR coefficients and the model correlation (Fig. 7a–c), and the difference between the correlation of the CNN and the LR (Fig. 7d–e), as calculated at each grid point. Larger LR coefficients for wind speed are associated with larger correlation of the LR model (Fig. 8a) in addition to an improved performance of the CNN over the LR (Fig. 8d). Conversely, a larger LR coefficient for sea-ice speed is associated with lower correlation (Fig. 8b) and does not show a statistically significant relationship with the difference metric, $corr_{CNN}$ - $corr_{LR}$ (Fig. 8e). A larger LR parameter for ice concentration is linked to higher model correlation (Fig. 8c) and tends toward the LR outperforming the CNN (Fig. 8f). The skill (not shown) exhibits the same patterns as the correlation.

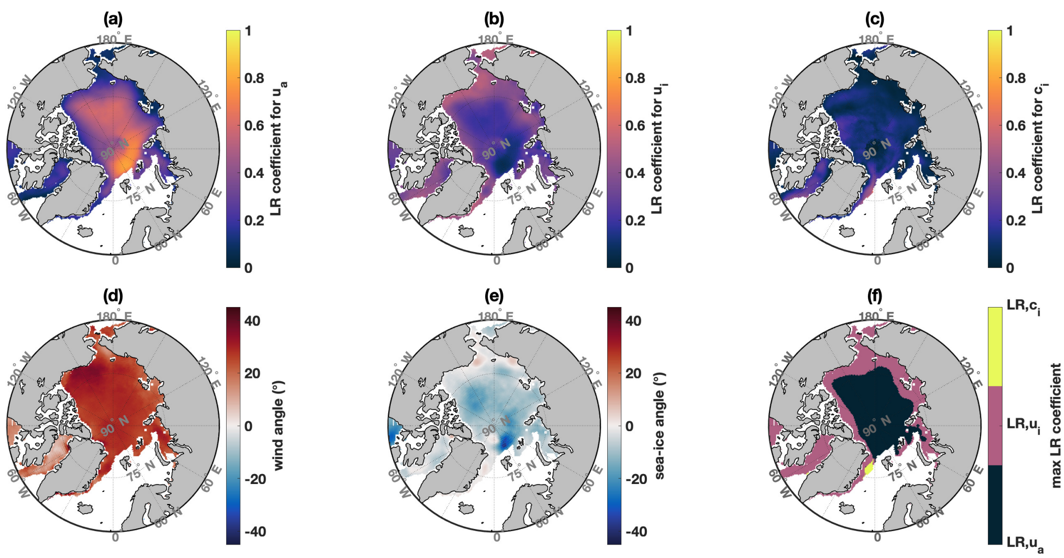

FIG. 7. (a–c) Magnitude of the normalized linear regression coefficient for the relationship between sea-ice velocity components and input parameters (a, wind speed, A; b, sea-ice speed, B; c, sea-ice concentration, C) normalized to the maximum of a–c. (d–e) Mean angle of (d) wind speed and (e) sea-ice speed to the predicted next-day sea-ice speed. (f) Maximum linear regression parameter (a–c) for predicting sea-ice velocity at each location. Wind and ice speed parameters are derived from calculating the magnitude of the parameters for the velocity components.

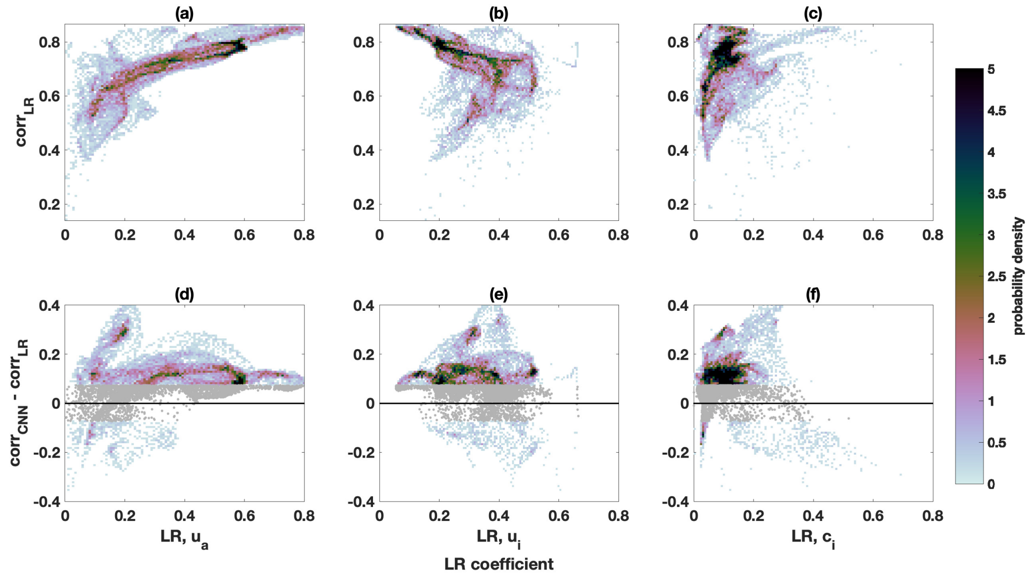

FIG. 8. PDFs for LR correlation (a–c) and the difference between the correlation of the CNN & LR correlation (d–f) compared to the LR coefficient magnitudes for (a & d) wind speed, LRu_a; (b & e) sea-ice speed, LRu_i; and (c & f) sea-ice concentration, LRc_i.

c. Attribution assessment of model predictive skill

We address our aims to understand (i) reductions in forecast skill and (ii) discrepancies in the performance of the different models by comparing the variability of these performance metrics (i.e. $corr_{CNN}$ and $corr_{CNN}$ - $corr_{LR}$) to variables related to ice motion (i.e. distance from coast, d_c; bathymetric depth, d; wind speed, u_a; ice speed, u_i; ice concentration, c_i; and the LR coefficients for wind speed, A, ice speed, B, and ice concentration, C). We focus on the difference between the CNN and the LR, because the CNN and LR both outperform the PS for almost all spatial locations.

In section 5 we find high model performance is linked to large distances from the coast, depths (Fig. 4 in section 5.a.2), wind speed, ice speed, ice concentration, (Fig. 6 in section 5.a.4), and values of the LR coefficients for wind speed & ice concentration (Fig. 8 in section 5.b). Additionally, the difference between the correlation of the CNN and LR models is typically smaller for high wind speed and ice speed, and larger for high sea-ice concentration (Fig. 6d–f in section 5.a.4), large distances from the coast, and large depths (Fig. 4 in section 5.a.2). We aim to confirm these findings by comparing the spatial and temporal variability in model correlation (Figs. 3a–f & 5a–g) to that of the various properties linked to ice motion (Figs. 3g–l & 5h–j), as well as to the spatial variability of the LR coefficients (Fig. 7a–c).

We analyze four spatial divisions (Fig. 1a) that are made based on overall model performance and the difference between the performance of the CNN and LR models. The Greenland Sea division (dark blue in Fig. 1a) covers the region to the east of Greenland where the model correlation is variable, but the LR largely outperforms the CNN. The Eastern Arctic division (light blue in Fig. 1a) represents the region of the eastern Arctic where the correlation is low and the LR outperforms the CNN. The Central Arctic division (light red in Fig. 1a) includes the central Arctic, the Beaufort Sea, and the regions to the north of the Canadian Arctic Archipelago. The Baffin Bay division (dark red in Fig. 1a) is the region where the model correlation is variable, but the CNN consistently outperforms the LR. The gray shading in Fig. 1a indicates regions that are not included in the following analysis. We discuss how the variability in the input parameters is linked to (i) model performance, (ii) the difference between the performance of the CNN and LR models, and (iii) the values for the LR coefficients in each division. We note the distinction between inter-divisional comparisons and analysis within each division, both of which are discussed below.

A summary of the inter-divisional comparisons is shown in Fig 9. Here the average values of the metrics and properties are shown for each division, and error bars represent the standard deviation. While the mean over any given division falls within one standard deviation of the mean for the other division for many properties, significance testing shows that for each property the differences between the mean value for each individual division and all other divisions are statistically significant (not shown). Analysis within each division is summarized in Fig. 10, which shows the ensemble-averaged correlation between each of the performance metrics and each of the properties related to ice motion within each division. The correlation between the maps of the performance metrics (Figs. 3a & f) and the average of the properties throughout time (Figs. 3g–i) are shown in Fig. 10a & b. The correlation between the daily time series of the performance metrics and the spatially averaged properties (similar to Figs. 5d & g vs. Figs. 5h–j, but using daily rather than monthly values) are shown in Fig. 10c & d. The properties are compared to the model correlation (circles, Figs. 10a & c) and the difference between the CNN and LR correlation, $corr_{CNN}$ - $corr_{LR}$ (triangles, Figs. 10b & d). The different divisions are represented by the different colors, as indicated in the legend. Values greater than zero are representative of cases where increases in the property are linked to increases in the model performance metric, while values less than zero indicate an inverse relationship between the property and performance metric.

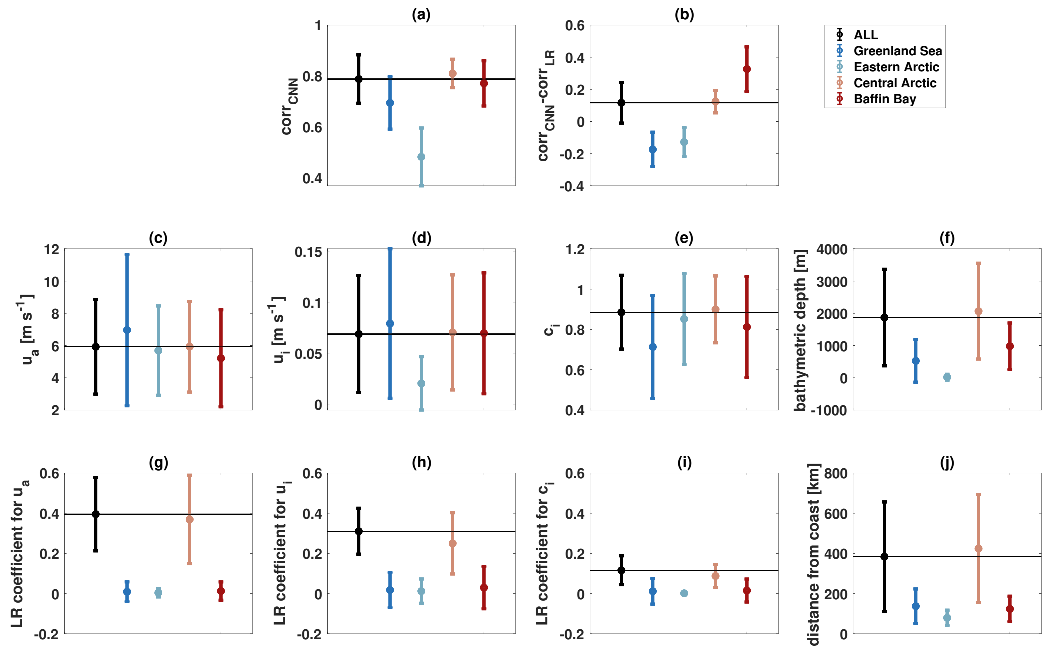

FIG. 9. Overall mean of the performance metrics, ((a) $corr_{CNN}$, and (b) $corr_{CNN}$ - $corr_{LR}$), and properties related to ice motion ((c) wind velocity, u_a ($m\ s^{-1}$); (d) ice velocity, u_i ($m\ s^{-1}$); (e) ice concentration, c_i; (f) bathymetric depth, d (m); (j) distance from coast, d_c (km); and the LR coefficients for (g) u_a, (h) u_i, and (i) c_i). Different colors represent the different spatial divisions, as indicated in the legend. Error bars represent the standard deviation within each division. The black line in panel each represents the mean value for the overall Arctic ('ALL' in the legend) for comparison.

1) Model predictive skill vs. properties related to ice motion

Inter-divisional comparisons suggest that low correlation of the CNN is typically linked to low depth, distance from coast, and ice speed, which is consistent with results from Fig. 6. For example, the Eastern Arctic division has the lowest $corr_{CNN}$, as well as the lowest mean of the properties listed above in comparison to the other divisions (Fig 9).

Visual inspection of spatial (Fig. 3) and temporal (Fig. 5) results also support this. For example, the low $corr_{CNN}$ found in the Eastern Arctic division (Fig. 3c) is coincident with low values for depth, distance from coast (Fig. 1b–c), wind speed, ice speed, and ice concentration (Figs. 3g–i).

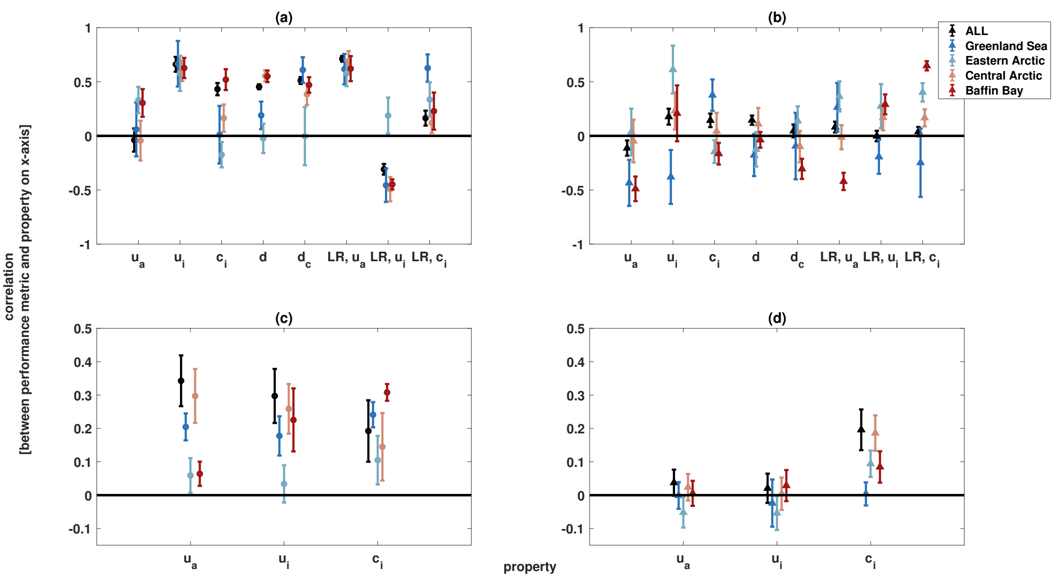

FIG. 10. Ensemble mean of the correlation between the model performance metrics (circles for $corr_{CNN}$ in a & c; triangles for $corr_{CNN}$ - $corr_{LR}$ in b & d) and the various properties related to ice motion within each of the spatial divisions (different shades of red and blue, as indicated in the legend). Correlations are calculated to understand how (a & b) spatially mapped performance metrics are related to spatial variability in time-averaged wind speed, u_a; ice speed, u_i; ice concentration, c_i; depth, d; distance from coast, d_c; LR parameter for wind, LRu_a; LR parameter for ice speed, LRu_i; and LR parameter for ice concentration, LRc_i; and (c & d) temporal variability in performance is linked to daily averages of u_a, u_i, and c_i within each division. Error bars represent the standard deviation of the ensemble runs within each division.

Temporally, the exceptionally low correlation in the Eastern Arctic division from November–May (Fig. 5b–d) is coincident with values of u_i for the Eastern Arctic division that are lower than all of the other divisions (Fig 5i). Additionally, the Central Arctic division exhibits a higher correlation than the other divisions, particularly during May–October, where the Central Arctic has higher u_a, u_i and c_i in comparison to the other divisions. Temporal analysis also shows that divisions that have a lower seasonal minimum c_i also exhibit a lower correlation relative to the other divisions, and in August–September the ordering for both c_i and $corr_{CNN}$ between divisions is: Baffin Bay < Eastern Arctic < Greenland Sea < Central Arctic.

Within each division, large $corr_{CNN}$ is typically related to high depth, distance from coast, wind speed, ice speed, and ice concentration, which is consistent with results from Fig. 6. This can be seen in Fig. 10a & c, where data points for all divisions are typically greater than zero (above the black line), which indicates that spatial (Fig. 10a) and temporal (Fig. 10c) variability of the properties listed on the x-axis are linked to variability in the correlation of the CNN. There are a few exceptions to this relationship when comparing spatial variability of performance metrics to the mean field of the properties: large wind speed is linked to low $corr_{CNN}$ within the Central Arctic and the overall Arctic; within the Eastern Arctic division large ice concentration, depth, and distance from coast are linked to low $corr_{CNN}$. Interestingly, many of these exceptions lie within the Eastern Arctic division, where overall depth, distance from coast, wind speed, ice speed, and ice concentration are significantly lower than other divisions. However, the values of these exceptions are within one standard deviation of zero, which indicates neither a positive or negative correlation between the model performance and the respective property. We note that the spatial comparisons (Fig. 10a) make use of the mean fields of u_a, u_i, and c_i, while temporal analyses (Fig. 10c) look at the daily time series that are averaged over the spatial domain of each division. We use spatial and temporal analyses here as a confirmation of results in Fig. 6, but do not expect perfect adherence due to the differences caused by averaging across space and time.

While Fig. 10a & c provides a quantitative analysis of the comparisons of spatial (Fig. 3c vs. g–i) and temporal (Fig. 5d vs. h–j) variability in the model correlation with respect to these properties, we can also see the link through visual inspection. For example, spatial patterns of high correlation within the Greenland Sea division (i.e. increasing from west to east; Fig. 3a–c) are coincident with high depth, distance from coast, ice speed, and wind speed, while low correlation is seen in locations with high ice concentration. Within the Eastern Arctic division, low correlation is largely linked to low depth and ice speed (Fig. 10a). High correlation within the Central Arctic division is generally coincident with high depth, distance from coast, ice speed, and ice concentration. Slightly lower correlations are seen in regions with lower values of u_a and u_i (western side), and lower c_i (eastern side and near the Bering Strait). Interestingly, the Beaufort Sea has high values of skill and correlation despite its proximity to land. However, the Beaufort Sea is relatively deep and has exceptionally high mean u_a and u_i in comparison to other coastal regions, properties that are linked to higher model performances (Figs. 4d & 6a–b). Lastly, high model correlation in the Baffin Bay division (Fig. 3) is aligned with large depth (Fig. 1b), u_a, u_i, and c_i (Fig. 3g–i). These spatial patterns of correlation within each of the divisions tend to be consistent with results from

Figs. 4d & 6a–b, the main exception being for the link between low correlation and high c_i within the Greenland Sea division, and the high correlation found close to the coast in the Beaufort Sea.

Temporally, the seasonal cycle for correlation follows that of u_a, u_i, and c_i, with minimum model correlations occurring during the months of minimum u_a, u_i, and c_i (August–September) for most models and divisions. The exceptions here are the Eastern and Central Arctic divisions where the correlation does not follow the seasonal cycle for c_i. This is likely a result of the low u_a and u_i in the Eastern and Central Arctic division during this time. Additionally, low seasonal variability in correlation within the Central Arctic division could be linked to the relatively small seasonal variations in u_a, u_i, and c_i in comparison to the other divisions.

2) Difference between predictive skill of the CNN and LR models vs. properties related to ice motion

Inter-divisional analysis suggests that low values for the difference metric, $corr_{CNN}$ - $corr_{LR}$ (the Greenland Sea and Eastern Arctic divisions in Fig 9b), are linked to low depth, distance from coast, and ice concentration (the Greenland Sea and Eastern Arctic division in Fig 9 e–f & j). Additionally, a low difference metric is linked to high u_a and u_i in the Greenland Sea division. Conversely, low u_i is linked to a low difference metric in the Eastern Arctic division (Fig 9d). The high difference metric in the Baffin Bay division is also linked to a lower mean u_a. As noted above, while the mean value of a particular division may fall within one standard deviation of that for other divisions, significance testing shows that the differences between means among divisions for a given property are statistically significant. For the case of c_i, these inter-divisional comparisons are consistent with results from Fig. 6d–f, where a high difference metric is linked to high c_i. Additionally, these results are consistent with the relationship between high $corr_{CNN}$ - $corr_{LR}$ and low wind and ice speeds found in Fig. 6d–f for the Greenland Sea (u_a & u_i) and Baffin Bay (u_a) divisions, but not the Eastern Arctic division (u_a & u_i).

Visual inspection of spatial (Fig. 3f vs. Figs. 1b–c & 3h–j) and temporal (Fig. 5g vs.Fig. 5h–j) results also supports this. Spatially, the low difference metric, $corr_{CNN}$ - $corr_{LR}$, in combination with relatively low depth, distance from coast, ice concentration, and exceptionally high wind and ice speeds in the Greenland Sea division compared to the rest of the Arctic is consistent with results in Fig. 6d–f. Additionally, temporal analysis shows the difference metric for the Greenland

Sea division remains lower than that for the entire Arctic (dark blue line is below black line), while u_a and u_i are higher in the Greenland Sea division than other divisions during the months of October–April. Similarly, for the Eastern Arctic division a relatively low depth, distance from coast, and ice concentration are linked to a low difference metric. However, contrary to patterns found in Fig. 6, the difference metric in the Eastern Arctic division is low, while u_a and u_i are also low in both spatial and temporal analyses. The difference metric for the Eastern Arctic division is lower than that for the Greenland Sea division from January–April, despite lower c_i and higher u_a and u_i in the Greenland Sea division, all of which are expected to contribute to a lower difference metric (Fig. 6). Spatially, the high difference metric in the Central Arctic division is linked to high c_i, low u_a, and high u_i relative to other divisions, which is consistent with results in Fig. 6, with the exception of the tendency of u_i. However, in temporal analysis of the Central Arctic division, the difference metric is particularly high compared to other divisions when u_i is lower in January–May, which is consistent with results in Fig. 6. The notably high difference metric in the Baffin Bay division compared to other divisions is linked to low u_a in both spatial and temporal (December–June in Fig. 5g & h) analyses.

Within each division, comparisons of $corr_{CNN}$ - $corr_{LR}$ with the properties related to ice motion are more nuanced, as data points in Fig 10b & d do not consistently lie above or below zero for a given property, particularly with spatial comparisons using the mean fields (Fig 10b). From results in Fig. 6, we would expect points in Fig 10b to be above zero for c_i and below zero for u_a and u_i (i.e. high $corr_{CNN}$ - $corr_{LR}$ is linked to high c_i, low u_a, and low u_i), which is only the case for some divisions. The region containing the entire Arctic (black) is consistent with this pattern for all variables on the $x-$axis, except for u_i. Additionally, these results are consistent with Fig. 6 for the following cases: the coincidence of high $corr_{CNN}$ - $corr_{LR}$ with low u_a, low u_i, and high c_i in the Greenland Sea division; high $corr_{CNN}$ - $corr_{LR}$ coincident with low u_a, but high c_i and depth in the Central Arctic region; the coincidence of high $corr_{CNN}$ - $corr_{LR}$ with low u_a in the Baffin Bay region. We find the following exceptions to the trends in Fig. 6: the coincidence of high $corr_{CNN}$ - $corr_{LR}$ and low d and d_c in the Greenland Sea division; high $corr_{CNN}$ - $corr_{LR}$ coincident with high u_i and low c_i in the Eastern Arctic division; and high $corr_{CNN}$ - $corr_{LR}$ coincident with high u_i, low c_i, low depth, and low distance from coast in the Baffin Bay division.

Comparisons between temporal variability of the difference metric and the various properties are more straightforward, and tend to show results that are consistent with what is found in Fig. 6, where a high difference metric is linked to low u_a, low u_i, and high c_i. This is true (i.e. data for u_a and u_i exist below the black line, and points for c_i are above), except for in the case of the region containing the entire Arctic, the Central Arctic division, and the Baffin Bay division for both u_a and u_i, as well as the Greenland Sea division for c_i. Additionally, while the ensemble mean value of the correlation between u_i and the difference metric is negative for the Greenland Sea division, it lies within one standard deviation of zero.

Looking at the time series (Fig. 5g–j) it is clear that the low difference metric in the Eastern Arctic division from December–May is linked to low u_i and high c_i, which is the opposite of what is expected from Fig. 6. Within the Central Arctic division low $corr_{CNN}$ - $corr_{LR}$ is linked to low c_i in June–October, while a slightly higher $corr_{CNN}$ - $corr_{LR}$ from December–May is linked to high u_a and low u_i. Within the Baffin Bay division low $corr_{CNN}$ - $corr_{LR}$ is linked to low c_i (Figs. 5g & j & 10d): during months of low c_i, the difference metric is not statistically different from zero (May–November, except August), while for all other months the opposite is true, and $corr_{LR} < corr_{CNN}$. Additionally, high $corr_{CNN}$ - $corr_{LR}$ during January–April is coincident with a low u_a. Temporal results from Fig. 5g–j tend to be consistent with results from Fig. 6, with the following exceptions: coincidence of low $corr_{LR} < corr_{CNN}$ with low u_i and high c_i from December–May within the Eastern Arctic division; coincidence of high $corr_{LR} < corr_{CNN}$ and high u_a from December–May in the Central Arctic division.

3) Impact of LR parameters on model performance metrics

We find that the performance metrics ($corr_{CNN}$ and $corr_{CNN}$ - $corr_{LR}$) are related to the values of the LR coefficients for the different input parameters (Fig. 8 in section 5.b). These results come from comparing the LR coefficient at each location (Fig. 7a–c) with the mapped values for the performance metrics (Fig. 3c & f). We use divisional analyses to confirm the maximum LR coefficient in each division (Fig. 9g–i vs. Fig. 7f), as well as the relationship between the performance metrics and the LR coefficients within each division (Fig. 10a & b vs. Fig. 8). We also aim to understand whether the variable with the highest LR coefficient has the strongest relationship to model performance. 3.

Inter-divisional comparisons (Fig. 9g–i) show that the mean LR coefficient for u_a is higher than all other coefficients in the Central Arctic division and the region covering the entire Arctic. For all other divisions the mean of the LR coefficients are within one standard deviation of each other and the maximum coefficient within each division is not conclusive. The mean LR coefficient within the overall Arctic and the Central Arctic division (Fig. 9g–i) is consistent with what is seen spatially (Fig. 7f). We find that variability in model performance is not necessarily linked most strongly to the property that exhibits the dominant LR coefficient within each division (i.e. a high LR coefficient for u_a does not necessarily mean that the correlation between either performance metric and u_a will be stronger than that between the performance metric and u_i or c_i). In other words, the high value of the LR coefficient for u_a in comparison to that for u_i or c_i for the Central Arctic division in Fig.9g is not linked to the correlation between model performance and u_a being higher than that for u_i or c_i in Fig 10.

In Fig. 8, high model correlation is found in locations with large LR coefficient for u_a and c_i, but a low LR coefficient for u_i. Analysis of the LR coefficient within each division (Fig. 10a & b) confirms this and shows that high $corr_{CNN}$ is related to high a high LR coefficient for u_a and c_i within all four divisions. The relationship between high $corr_{CNN}$ and a low LR coefficient for u_i is also seen for all divisions except the Eastern Arctic division (light blue in Fig. 10a). While the general trend in Fig. 8 suggests high correlation to be linked to a low LR coefficient for u_i, it is clear that when $corr_{CNN} < 0.6$ (which is the case for the Eastern Arctic division, where the mean $corr_{CNN}$ is 0.5 ± 0.02), a high LR coefficient for u_i is linked to higher $corr_{CNN}$.

The relationship between a high difference metric and a high LR coefficient for u_a seen in Fig. 8 is confirmed within all divisions, except for the Baffin Bay division (Fig. 10b), however the Central Arctic division is within one standard deviation of zero. The relationship between high $corr_{CNN}$ - $corr_{LR}$ and a low LR coefficient for c_i seen in Fig. 8c is only found within the Greenland Sea division. While the general pattern in Fig. 8c suggests a link between high $corr_{CNN}$ - $corr_{LR}$ and a low LR coefficient for c_i, this is largely true where the LR coefficient for c_i is high (> 0.6), which is the case for the Greenland Sea division (0.69 ± 0.34). When the LR coefficient for $c_i < 0.6$ the opposite is true, and high $corr_{CNN}$ - $corr_{LR}$ is linked to a high LR coefficient for c_i, which is the case for the Greenland Sea, Eastern Arctic, and Baffin Bay divisions. Thus, Fig. 10b confirms results from Fig. 8.

6. Conclusions

a. A CNN can make skillful predictions of sea-ice motion on one-day time scales.

As sea-ice in the Arctic declines and opens new pathways for maritime transportation, the skill of sea-ice motion predictions becomes increasingly important (Bennett et al. 2020; Cao et al. 2022). This work uses machine learning models to make one-day predictions of sea-ice motion for operational forecasting. We show that a CNN can make skillful predictions of sea-ice velocity and outperforms other statistical models in most instances. In comparison to the other models, the CNN has the benefit of incorporating non-linearities between inputs and spatial information when making predictions. We also show that a grid-wise linear regression (LR) model performs almost as well as a CNN in most instances, and comes with the benefit of decreased complexity in comparison to neural networks. Both the CNN and LR models outperform the baseline PS model. Additionally, we find that the CNN shows improved performance in comparison to the models of Maeda et al. (2020); Kimura and Wakatsuchi (2000) discussed in section 2: the correlation of the CNN is as low as 0.4 in the Eastern Arctic, and 0.7 in the Canadian Arctic Archipelago (Fig. 3c), where Maeda et al. (2020) find correlation between ice motion and geostrophic wind as low as 0 and 0.4 in the same regions. Lastly, while comparing the model performance to that of a dynamical model was outside the scope of this study, our model was an extension of that presented by Zhai and Bitz (2021) (differences between the two models are identified in Table S1), which was found to have higher correlations for sea-ice velocity with satellite observations than the CICE5 dynamical model for sea ice.

b. Model predictive skill and discrepancies between model performances are linked to various properties related to sea-ice motion.

Model performances vary spatially and seasonally, and are linked to variability in properties related to sea-ice motion. Although there are exceptions that come with having different combinations of these properties, in general, better model performance is linked to:

- increased bathymetric depth and distance from the coast
- larger mean values of u_a, u_i, and c_i
- larger LR coefficients for u_a and c_i; smaller LR coefficient for u_i

The CNN outperforms the LR in most cases. We have shown that the following are related to increases in the performance of the CNN over the LR:

- larger distance from coast and greater bathymetric depth
- smaller mean u_a and u_i, and larger mean c_i
- larger LR coefficient for u_a, and smaller LR coefficients for c_i

Interestingly, the LR model tends to outperform the CNN model in some coastal regions where non-linear effects might be expected to play a large role. However, the locations where this happens exhibit shallow depths, and when coastal waters are deep (i.e. the Beaufort Sea) the CNN outperforms the LR. We note that sharp discontinuities between ocean and land pixels may reduce the quality of the CNN predictions due to the way the CNN incorporates spatial filters and non-local information in its predictions Sonnewald et al. (2021). This may also impact our result that the LR outperforms the CNN at shallower depths because depth increases with increasing distance from the coast. To address this, future analyses we will apply a non-local LR at each grid-point for a more direct comparison between LR and CNN models. However, even with non-localities built in, the LR doesn't apply spatial filters in the same way that the CNN does, so we may not be able to reproduce the same decreases in performance inherent to the CNN in coastal regions.

The LR typically outperforms the CNN in regions where wind speed is not the dominant LR coefficient: ice velocity is the dominant LR coefficient in the coastal regions of the eastern Arctic, and sea-ice concentration dominates the LR predictions in the coastal region to the east of Greenland. Conversely, wind speed is found to be the dominant LR coefficient wherever the CNN outperforms the LR. This suggests that the relationship between wind velocity and ice velocity includes non-linearities that are captured by the CNN (and not the LR), leading to an improved performance.

We find that larger LR coefficients for a given parameter are not necessarily linked to larger parameter values (e.g. in the Greenland Sea division, ice concentration is the dominant predictor in regions where wind and ice speed are exceptionally high). However, we find that the LR coefficient for wind speed tends to be lower in regions with low mean c_i. This contradicts previous findings, where areas with high c_i are known to exhibit larger internal ice stresses, which leads to a reduction in the dependence of ice motion on wind (Kimura and Wakatsuchi 2000; Maeda et al.

2020). We note that this particular conclusion does not take into account instantaneous effects, as it is a comparison between a mean c_i over time and a LR coefficient that is descriptive of ice motion over the duration of the study. Future work could decrease the time period over which LR is run to obtain equations that are more descriptive of instantaneous effects such as that of ice stresses due to high c_i. Lastly, we find that variability in model performance is not necessarily linked to the dominant LR coefficient within each region.

c. Wind velocity plays the largest role in predicting ice velocity.

We find that the spatial average of the wind factor over the Arctic is 0.72% (Fig. S2). The wind factor is higher for regions in the Central Arctic in comparison to coastal regions, confirming historical results (Thorndike and Colony 1982; Serreze et al. 1989; Kimura and Wakatsuchi 2000; Maeda et al. 2020). We also show an average turning angle to the wind of $24.9° \pm 11.3°$, which is consistent with the cited historical results. Analysis of LR parameters shows that of all of the input predictors, wind velocity has the largest importance in predicting sea-ice velocity. This relationship is particularly strong in the central Arctic, and is reduced in coastal regions. Furthermore, an increased dependence of the models on wind speed is related to increased model performance for the CNN, which provides further evidence as to why the models are not as skillful at predicting ice speed in coastal regions (i.e. ice speed is not as dependent on the training information in these regions). Future work will build off of these results and look at using outputs from machine learning models to understand how the relationship between wind and ice velocity is changing in time as the ice melts.

Acknowledgments. LH, MRM, and PH were supported by ONR (grant N00014-20-1-2772). MRM was supported by NSF (award OPP-1936222). STG was supported by NSF (award OPP-1936222) and by U.S. Department of Energy (DOE) (Award DE-SC002007). DG was supported by NSF Award 1928305. CMB was supported by NASA (award 80NSSC21K0745). Figures in this report were prepared using MATLAB, Matplotlib: A 2DGraphics Environment Hunter (2007). Colormaps were obtained using the cmocean package (Thyng et al. 2016) and the CubeHelix Colormap Generator (Stephen23 2023). We thank our reviewers for their helpful feedback.

Data availability statement. We acknowledge all sources of publicly available data that were used in this study. The JRA55-do dataset can be accessed at `https://climate.mri-jma.go.jp/pub/ocean/JRA55-do/`. Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors, Version 4 are made available by the National Snow and Ice Data Center (NSIDC) and can be accessed at `https://nsidc.org/data/nsidc-0116/versions/4`. Sea Ice Concentrations from Nimbus-7 Passive Microwave Data, Version 1 are made available by the NSIDC and can be accessed at `https://doi.org/10.5067/8GQ8LZQVL0VL`. The International Bathymetric Chart of the Arctic Ocean (IBCAO) are available at `https://www.gebco.net/data_and_products/gridded_bathymetry_data/arctic_ocean/`. All of the data and files used for processing for this paper can be accessed at `https://doi.org/10.6075/J0X06774`.

References

Abadi, M., and Coauthors, 2015: TensorFlow: Large-scale machine learning on heterogeneous systems. URL https://www.tensorflow.org/, software available from tensorflow.org, Version 2.8.0.

Andersson, T. R., and Coauthors, 2021: Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nature Communications*, **12**, 5124, https://doi.org/10.1038/s41467-021-25257-4.

Bennett, M. M., S. R. Stephenson, K. Yang, M. T. Bravo, and B. De Jonghe, 2020: The opening of the transpolar sea route: Logistical, geopolitical, environmental, and socioeconomic impacts. *Marine Policy*, **73**, 104 178, https://doi.org/10.1016/j.marpol.2020.104178.

Blanchard-Wrigglesworth, E., R. Cullather, W. Wang, J. Zhang, and C. Bitz, 2015: Model forecast skill and sensitivity to initial conditions in the seasonal sea ice outlook. *Geophysical Research Letters*, **42 (19)**, 8042–8048, https://doi.org/10.1002/2015GL065860.

Brodzik, M. J., B. Billingsley, T. Haran, B. Raup, and M. H. Savoie, 2012: Ease-grid 2.0: Incremental but significant improvements for earth-gridded data sets. *ISPRS International Journal of Geo-Information*, **1 (1)**, 32–45, URL https://www.mdpi.com/2220-9964/1/1/32.

Cao, Y., and Coauthors, 2022: Trans-arctic shipping routes expanding faster than the model predictions. *Global Environmental Change*, **73**, 102 488, https://doi.org/10.1016/j.gloenvcha.2022.102488.

Carmack, E., and Coauthors, 2015: Toward quantifying the increasing role of oceanic heat in sea ice loss in the new Arctic. *Bulletin of the American Meteorological Society*, **96 (12)**, 2079–2105, https://doi.org/10.1175/BAMS-D-13-00177.1.

Cavalieri, D., C. Parkinson, P. Gloersen, and H. Zwally, 1996: Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1. NASA National Snow and Ice Data Center Distributed Active Archive Center, URL https://nsidc.org/data/NSIDC-0051/versions/1, access Data: March 2021.

Crawford, A., J. Stroeve, A. Smith, and A. Jahn, 2021: Arctic open-water periods are projected to lengthen dramatically by 2100. *Communications Earth & Environment*, **2 (1)**, 109, https://doi.org/10.1038/s43247-021-00183-x.

Dietterich, T., 1998: Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comptation*, **10**, 1895–1923, https://doi.org/10.1162/089976698300017197.

Dumitru, C. O., V. Andrei, G. Schwarz, and M. Datcu, 2019: Machine learning for sea ice monitoring from satellites. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **XLII-2/W16**, 83–89, https://doi.org/10.5194/isprs-archives-XLII-2-W16-83-2019.

Ekman, W. V., 1905: On the influence of the earth's rotation on ocean-currents. *Arkiv for Matematik, Astronomi Och Fysik*, **2 (11)**, 1–52.

Feltham, D. L., 2008: Sea ice rheology. *Annual Review of Fluid Mechanics*, **40 (1)**, 91–112, https://doi.org/10.1146/annurev.fluid.40.111406.102151.

Hibler, W., 1979: A dynamic thermodynamic sea ice model. *Journal of Physical Oceanography*, **9 (4)**, 815–846, https://doi.org/10.1175/1520-0485(1979)0090815:ADTSIM2.0.CO;2.

Hoffman, L., M. R. Mazloff, S. T. Gille, D. Giglio, C. M. Bitz, P. Heimbach, and K. Matsuyoshi, 2023: Data from: Machine learning for daily forecasts of arctic sea-ice motion: an attribution assessment of model predictive skill. UC San Diego Library Digital Collections., publish Data: June 2023, https://doi.org/10.6075/J0X06774.

Hsieh, W. W., and B. Tang, 1998: Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society*, **79 (9)**, 1855 – 1870, https://doi.org/10.1175/1520-0477(1998)079⟨1855:ANNMTP⟩2.0.CO;2.

Hunke, E., and Coauthors, 2020: Should sea-ice modeling tools designed for climate research be used for short-term forecasting? *Current Climate Change Reports*, **6**, 121–136, https://doi.org/10.1007/s40641-020-00162-y.

Hunter, J. D., 2007: Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, **9 (3)**, 90–95, https://doi.org/10.1109/MCSE.2007.55.

Jahn, A., 2018: Reduced probability of ice-free summers for 1.5ºC compared to 2ºC warming. *Nature Climate Change*, **8**, 409–413, https://doi.org/10.1038/s41558-018-0127-8.

Jakobsson, M., L. Mayer, C. Bringensparr, and et al., 2020: The international bathymetric chart of the arctic ocean version 4.0. *Sci Data*, **7 (176)**, https://doi.org/10.1038/s41597-020-0520-9.

Kern, S., T. Lavergne, D. Notz, L. T. Pedersen, R. T. Tonboe, R. Saldo, and A. M. Sørensen, 2019: Satellite passive microwave sea-ice concentration data set intercomparison: closed ice and ship-based observations. *The Cryosphere*, **13 (12)**, 3261–3307, https://doi.org/10.5194/tc-13-3261-2019.

Kim, Y. J., H.-C. Kim, D. Han, S. Lee, and J. Im, 2020: Prediction of monthly arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks. *The Cryosphere*, **14 (3)**, 1083–1104, https://doi.org/10.5194/tc-14-1083-2020.

Kimura, N., and M. Wakatsuchi, 2000: Relationship between sea-ice motion and geostrophic wind in the northern hemisphere. *Geophysical Research Letters*, **27 (22)**, 3735–3738, https://doi.org/10.1029/2000GL011495.

Kobayashi, S., and Coauthors, 2015: The jra-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, **93 (1)**, 5–48, https://doi.org/10.2151/jmsj.2015-001.

Kwok, R., G. Spreen, and S. Pang, 2013: Arctic sea ice circulation and drift speed: Decadal trends and ocean currents. *Journal of Geophysical Research: Oceans*, **118 (5)**, 2408–2425, https://doi.org/https://doi.org/10.1002/jgrc.20191.

Landy, J., G. Dawson, and M. Tsamados, 2022: A year-round satellite sea-ice thickness record from cryosat-2. *Nature*, **609**, 517–522, https://doi.org/10.1038/s41586-022-05058-5.

Lee, S., J. Im, J. Kim, M. Kim, M. Shin, H.-c. Kim, and L. J. Quackenbush, 2016: Arctic sea ice thickness estimation from cryosat-2 satellite data using machine learning-based lead detection. *Remote Sensing*, **8 (9)**, https://doi.org/10.3390/rs8090698.

Li, M., R. Zhang, and K. Liu, 2021: Machine learning incorporated with causal analysis for short-term prediction of sea-ice. *Frontiers in Marine Science*, **8 (649378)**, https://doi.org/10.3389/fmars.2021.649378.

Maeda, K., N. Kimura, and H. Yamaguchi, 2020: Temporal and spatial change in the relationship between sea-ice motion and wind in the arctic. *Polar Research*, **39**, https://doi.org/10.33265/polar.v39.3370.

Marquardt, D. W., and R. D. Snee, 1975: Ridge regression in practice. *The American Statistician*, **29 (1)**, 3–20, https://doi.org/10.1080/00031305.1975.10479105.

Maykut, G., A. Thorndike, and N. Untersteiner, 1972: Aidjex scientific plan. *AIDJEX Bulletin*, **15**, 1–67.

McGovern, A., R. Lagerquist, D. J. Gagne, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, **100 (11)**, 2175 – 2199, https://doi.org/10.1175/BAMS-D-18-0195.1.

Meredith, M., and Coauthors, 2019: Chapter 3: Polar regions. in: IPCC special report on the ocean and cryosphere in a changing climate.

Notz, D., and J. Stroeve, 2018: The trajectory towards a seasonally ice-free arctic ocean. *Current Climate Change Reports*, **4**, 407–416, https://doi.org/10.1007/s40641-018-0113-2.

Olason, E., and D. Notz, 2014: Drivers of variability in Arctic sea-ice drift speed. *Journal of Geophysical Research: Oceans*, **119 (9)**, 5755–5775, https://doi.org/10.1002/2014JC009897.

Onarheim, I. H., T. Eldevik, L. H. Smedsrud, and J. Stroeve, 2018: Seasonal and regional manifestation of arctic sea ice loss. *Journal of Climate*, **13 (12)**, 4917–4932, https://doi.org/10.1175/JCLI-D-17-0427.1.

O'Shea, K., and R. Nash, 2015: An introduction to convolutional neural networks. arXiv, https://doi.org/10.48550/ARXIV.1511.08458.

Petrou, Z. I., and Y. Tian, 2019: Prediction of sea ice motion with convolutional long short-term memory networks. *IEEE Transactions on Geoscience and Remote Sensing*, **57 (9)**, https://doi.org/10.1109/TGRS.2019.2909057.

Press, W., B. Flannery, S. Teukolsky, and W. Vetterling, 1986: *Numerical Recipes*. Cambridge University Press, 636-639 pp.

Rampal, P., J. Weiss, and D. Marsan, 2009: Positive trend in the mean speed and deformation rate of arctic sea ice, 1979-2007. *Journal of Geophysical Research*, **114 (C5)**, C05 013, https://doi.org/10.1029/2008JC005066.

Serreze, M. C., R. G. Barry, and A. S. McLaren, 1989: Seasonal variations in sea ice motion and effects on sea ice concentration in the canada basin. *Journal of Geophysical Research: Oceans*, **94 (C8)**, 10 955–10 970, https://doi.org/10.1029/JC094iC08p10955.

Serreze, M. C., M. M. Holland, and J. Stroeve, 2007: Perspectives on the arctic's shrinking ice cover. *Science*, **315 (5815)**, 1533–1536, https://doi.org/10.1126/science.1139426.

Sonnewald, M., R. Lguensat, D. C. Jones, P. D. Deuben, J. Brajard, and V. Balaji, 2021: Bridging observations, theory and numerical simulation of the ocean using machine learning. *Environmental Research Letters*, **16 (073008)**, https://doi.org/10.1088/1748-9326/ac0eb0.

Stephen23, 2023: Cubehelix colormap generator: Beautiful and versatile! GitHub, URL https://github.com/DrosteEffect/CubeHelix/releases/tag/2.3.2, access Date: January 2023.

Stroeve, J., and D. Notz, 2018: Changing state of arctic sea ice across all seasons. *Environmental Research Letters*, **13 (10)**, 103 001, https://doi.org/10.1088/1748-9326/aade56.

Stroeve, J., M. C. Serreze, M. M. Holland, J. E. Kay, J. Malanik, and A. P. Barrett, 2012: The Arctic's rapidly shrinking sea ice cover: A research synthesis. *Climate Change*, **110**, 1005–1027, https://doi.org/10.1007/s10584-011-0101-1.

Sumata, H., and Coauthors, 2018: Decorrelation scales for arctic ocean hydrography – part i: Amerasian basin. *Ocean Science*, **14 (1)**, 161–185, https://doi.org/10.5194/os-14-161-2018.

Szegedy, C., and Coauthors, 2015: Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9, https://doi.org/10.1109/CVPR.2015.7298594.

Taboda, F. F., C. A. Stock, S. M. Griffies, J. Dunne, J. G. John, J. R. Small, and H. Tsujino, 2019: Surface winds from atmospheric reanalysis lead to contrasting oceanic forcing and coastal upwelling patterns. *Ocean Modelling*, **133**, 79–111, https://doi.org/10.1016/j.ocemod.2018.11.003.

Tang, B., W. W. Hsieh, A. H. Monahan, and F. T. Tangang, 2000: Skill comparisons between neural networks and canonical correlation analysis in predicting the equatorial pacific sea surface temperatures. *Journal of Climate*, **13 (1)**, 287–293, https://doi.org/10.1175/1520-0442(2000)013⟨0287:SCBNNA⟩2.0.CO;2.

Thoman, R. L., and Coauthors, 2022: The arctic. *Bulletin of the American Meteorological Society*, **103 (8)**, S257 – S306, https://doi.org/https://doi.org/10.1175/BAMS-D-22-0082.1, URL https://journals.ametsoc.org/view/journals/bams/103/8/BAMS-D-22-0082.1.xml.

Thomson, R. E., and W. J. Emery, 2014: *Statistical Methods and Error Handling*. Data Analysis Methods in Physical Oceanography (Third Edition), Elsevier, 219-311 pp., https://doi.org/10.1016/B978-0-12-387782-6.00003-X.

Thorndike, A., and R. Colony, 1982: Sea ice motion in response to geostrophic winds. *Journal of Geophysical Research: Oceans*, **87 (C8)**, 5845–5852, https://doi.org/10.1029/JC087iC08p05845.

Thyng, K. M., C. A. Greene, R. D. Hetland, H. M. Zimmerle, and S. F. DiMarco, 2016: True colors of oceanography: Guidelines for effective and accurate colormap selection. *Oceanography*, **29 (3)**, https://doi.org/10.5670/oceanog.2016.66.

Toms, B. A., E. A. Barnes, and I. Ebert-Uphott, 2020: Physically interpretable neural networks for the geosciences: Applications to earth system variability. *Journal of Advances in Modeling Earth Systems*, **12 (9)**, e2019MS002 002, https://doi.org/10.1029/2019MS002002.

Tschudi, M., W. Meier, J. Stewart, C. Fowler, and J. Maslanik, 2019: Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors, Version 4. NASA National Snow and Ice Data Center Distributed Active Archive Center, URL https://nsidc.org/data/NSIDC-0116/versions/4, access Data: March 2021.

Tschudi, M. A., W. N. Meier, and J. S. Steward, 2020: An enhancement to sea ice motion and age products at the National Snow and Ice Data Center (NSIDC). *The Cryosphere*, **14 (5)**, 1519–1536, https://doi.org/10.5194/tc-14-1519-2020.

Tsujino, H., and Coauthors, 2018: JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean Modelling*, **130**, 79–139, https://doi.org/10.1016/j.ocemod.2018.07.002.

Untersteiner, N., A. Thorndike, D. Rothrock, and K. Hunkins, 2007: AIDJEX Revisited: A look back at the U.S.-Canadian Arctic Ice Dynamics Joint Experiment 1970-78. *Arctic*, **60 (3)**, 327–336, URL http://www.jstor.org/stable/40512908.

Wang, X., R. Chen, C. Li, Z. Chen, F. Hui, and X. Cheng, 2022: An intercomparison of satellite derived Arctic sea ice motion products. *Remote Sensing*, **14 (5)**, https://doi.org/10.3390/rs14051261.

Weiss, J., 2013: *Drift, Deformation, and Fracture of Sea Ice: A Perspective Across Scales*. Springer Dordrecht, https://doi.org/10.1007/978-94-007-6202-2.

Zhai, J., and C. M. Bitz, 2021: A machine learning model of Arctic sea ice motions. https://doi.org/10.48550/ARXIV.2108.10925.