Predicting Sea Ice Concentration with Uncertainty Quantification using Passive Microwave and Reanalysis Data: A Case Study in Baffin Bay

Ray Valencia, Xinwei Chen, Armina Soleymani, K. Andrea Scott

***Abstract*—In recent years, the adoption of deep learning (DL) techniques for predicting sea ice concentration (SIC) given both passive microwave (PM) data and reanalysis data has seen a growing interest. For use in downstream services, these sea ice concentration estimates should be accompanied by uncertainty estimates. To provide these estimates, we utilize a heteroscedastic Bayesian neural network (HBNN), which can estimate both model (epistemic) and data (aleatoric) uncertainty. We use both PM and atmospheric data as our input features, and demonstrate that both are needed for accurate SIC estimates. Results show that over an annual cycle, the months of melt onset, such as April, May, and June produce the highest uncertainties relative to other months, with total (epistemic + aleatoric) uncertainties of approximately 20%, while areas in the marginal ice zone, contributed highest total uncertainty of 25% spatially. When considering an average over the test year, the level of uncertainty due to the data (aleatoric) is consistent with other studies, at 10%-15%. The advantage of our approach is that the uncertainties are specific to the data instance, and both model and data uncertainty are estimated.**

I. Introduction

Sea ice concentration (SIC) is the measure of sea ice area relative to the total area in a particular region. The accurate estimation of SIC in the polar regions of the Arctic and Antarctic is necessary for studies concerning climate change [1], sustainability of local ecosystems [2], and human endeavors such as the shipment of goods, natural resource exploration, and fisheries [3]. Multiple remote sensing methodologies have been applied in SIC estimation, using approaches based on optical imagery, synthetic aperture radar (SAR) imagery [4], delay Doppler maps (DDMs) from Global Navigation Satellite System-Reflectometry (GNSS-R) [5], and passive microwave (PM) data [6], [7]. Optical imagery allows for a straightforward interpretation of sea ice concentration due to the strong contrast in albedo between open water and sea ice. However this interpretation is only feasible when cloud-free images are available. SAR is an active sensor that measures a backscattered signal from the Earth's surface generally in the low-frequency portion of the electromagnetic spectrum, and thus is not affected by atmospheric moisture or cloud cover. Data can be acquired from SAR sensors at a high spatial resolution (approximately 50-100 m). Imagery from SAR sensors can be difficult to interpret due to speckle noise and the sensitivity of the SAR signal to both imaging geometry and properties of the surface.

To the contrary, PM sensors measure microwave radiation emitted from the earth's surface and, at low frequencies, are also less affected by cloud cover. Instead of measuring SIC directly, PM sensors first measure the brightness temperatures (TB) and then calculate SIC via retrieval algorithms such as the bootstrap method [8] or the enhanced National Aeronautics and Space Administration (NASA) team (NT2) algorithm [9]. An excellent overview of PM SIC retrieval algorithms is given in [10], [11]. PM-TB based SIC estimation data calculated from these algorithms are freely available, allowing for great accessibility, but can be negatively affected by many factors that influence the SIC estimation accuracy. These include the presence of melt on the ice surface, thin ice, atmospheric moisture and wind-roughening of open water within the sensor footprint [12]–[14]. These can be further complicated due to seasonal/monthly changes of these factors [15]. To correct for atmospheric moisture and wind-roughening, the algorithms often use weather filters [16], [17] but these filters have been shown to not only remove weather effects but the ice itself [18]. An alternative approach is to correct the brightness temperatures before using them in the retrieval algorithm [19], [11], [15].

Among methods in SIC estimation from remote sensing imagery, there lies computational methods based on the ability of an algorithm to adaptively learn from data. This field of computational methods is known as machine learning (ML). A specific subset of ML, known as deep learning (DL), has garnered interest in the last few years. The concept of deep refers to the model architecture being partially composed of a series of layers where each layer has a set of weights that need to be learned. With recent advancements in hardware capabilities and increasing data volumes, developments in DL have flourished, demonstrating their ability to solve difficult problems and better learn spatio-temoporal patterns in data as opposed to other ML approaches [20]. For this reason, the use of deep neural networks (DNNs) in SIC estimation has gained popularity. For example, deep convolutional neural networks (CNNs) have been demonstrated to produce signif-

This work was supported in part by Environment and Climate Change Canada (ECCC), as well as the University of Waterloo's Engineering Excellence Fellowship. (*Corresponding author: Xinwei Chen*)

Ray Valencia, Xinwei Chen, Armina Soleymani, and K. Andrea Scott are with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: ra2valen@uwaterloo.ca; xinwei.chen@uwaterloo.ca;armina.soleymani@uwaterloo.ca; ka3scott@uwaterloo.ca).

icantly improved sea ice concentration estimates from SAR data during both melt and freeze-up, as compared to passive microwave data [4], [21]. While these studies used ice charts as training labels, the approach has also been investigated using PM data as training labels and using the CNN as change of scale operator [22], [23]. Among studies using ML or DNNs to estimate SIC from PM data, some have shown including atmospheric variables in the input features can lead to improved SIC estimates [24]–[26].

In the domain of remote sensing and climate sciences, uncertainty quantification (UQ) of data products is needed to use the products to support decision-making and climate policy [27]. For sea and lake ice remote sensing, this endeavor has only been recently explored. Asadi *et al.* [28] proposed a methodology utilizing multilayer perceptron's (MLP's) to quantify uncertainty in detection of ice and water in SAR imagery. Results showed the uncertainty was increased in regions where atmospheric features would be expected to lead to confusion between ice and water, supporting interpretability of the results. Additionally, UQ from a convolutional neural network was incorporated for lake ice mapping and sea ice segmentation using SAR images in Saberi *et al.* [29] and Chen *et al.* [30], respectively.

Previous studies on uncertainty quantification for sea ice concentration have shown uncertainty to originate from characteristics such as sensor noise, sensitivity of TB to surface properties, or spatial resolution limitations. This uncertainty has been quantified in previous methods to estimate SIC [19], [31], but related uncertainties in SIC retrieval within a DL framework have not been explored. Thus this study explores the use of a deep learning model to estimate SIC and quantify the uncertainty in these estimates, by leveraging both TB data and geophysical variables such as air temperature and wind speed. The deep learning model is implemented using an MLP architecture (or DNN), with methods for UQ [32], [33]. The performance and results of the model are evaluated via spatial investigation and statistical analysis. We discuss the seasonal variation of uncertainty and consider the effect of using differing SIC labels in training. The main contributions of this study are,

1) A proof of concept for a data driven approach using a DL model to estimate SIC and its uncertainty. The method allows seasonal and spatial variability of the uncertainties to be explored as a function of the input data, season and region.
2) It is the first study of sea ice concentration uncertainty from PM data that considers separate uncertainties for both the model and the data [34] and is able to generate uncertainty maps specific to the data instance. The level of uncertainty due to data is consistent with that found in earlier studies [19].

II. Background

Sea ice concentration can be estimated from TB data by exploiting the difference between horizontally and vertically polarized TBs, which is greater for water than for ice. Linear interpolation between tie points that express typical values of polarization difference or its normalized equivalent (polarization ratio) can be used, though in practice other frequencies need to be considered to differentiate between different ice types. These methods are complicated by sensitivity of TB to wind-roughening (across 6.9 GHz -89 GHz) and atmospheric moisture, which is normally quantified through integrated water vapor (WV) and integrated cloud liquid water(LW) [11], [35]. Sensitivity to CW increases monotnoically with frequency, whereas sensitivity to WV increases non-monotonically with frequency, peaking at the water vapor absorption line of 22.235 GHz [36]. These sensitivities impact the upwelling and downwelling contributions of the brightness temperature. Sensitivity to wind speed (WS) is polarization-dependent and is due to wind-roughening of the surface, which leads to surface scattering and impacts surface emission. Each of these sensitivities can result in TB signatures over weather-affected open water that are similar to those of intermediate ice concentrations leading to spurious sea ice concentration retrievals [18], [37]. Open water is more affected by atmospheric contamination than consolidated sea ice [11] because the surface contribution to the overall brightness temperature is lower over open water than sea ice due to its lower emissivity. In practice, to mitigate atmospheric contamination and reduce spurious sea-ice concentration retrievals, different weather filters are used in SIC retrieval algorithms. To correct for WV, the gradient ratios of 19 and 22 GHz frequency are used: GR(1922V) $\leq$ 0.04, and to correct for LW and WS, the gradient ratios of 19 and 37 GHz frequency are used: GR(1937V) $\leq$ 0.045 [37], [38]. An alternative approach is to use a radiative transfer model (RTM) to correct the TB data for atmospheric effects prior to their use in the interpolation method [11], [18]. This correction consists of subtracting from the observed TB the difference between one retrieved from the RTM considering atmospheric effects, and one retrieved from the RTM without atmospheric effects. These atmospheric effects are included via data (WS, CLW and WV) typically from weather prediction models or reanalyses.

Due to these atmospheric sensitivities, the problem of estimating SIC from PM data can be described as a non linear regression problem. In contrast to earlier studies, a multilayer perceptron (MLP) model is used here for SIC estimation. MLP models are a type of NN model and are a popular choice for non linear regression problems due to their flexibility [39]–[42]. MLP models are a type of feedforward NN characterized by several layers of neurons, which at minimum consist of an input layer, output layer, and a hidden layer in between the input and output. When these NNs have multiple hidden layers, they are considered subset of DNNs.

Although DNNs produce valuable results, they are often taken with no consideration for the uncertainty of their estimates. As DNNs are increasingly used in decision making, the requirement to provide uncertainty estimates has seen a rise of interest in various domains including health sciences [43], computer vision [33], automated vehicles [44], remote sensing [34], [45], [28] and others [46]. Additionally, if we can

measure the uncertainty in our models, it brings the possibility of reducing uncertainty, producing higher confidence in the model estimates.

Uncertainty in DL can be categorized into two types, epistemic and aleatoric uncertainty. Epistemic uncertainty is described as the uncertainty due to the model [33]. To reduce epistemic uncertainty one can either produce a model that encapsulates the problem better, or add more data. Methods to effectively measure epistemic uncertainty in NNs include Monte Carlo (MC) Dropout [47], Bayesian NNs [32], and Deep Ensembles [48].

In contrast, aleatoric uncertainty is described as the uncertainty that originates from the intrinsic randomness of the input or data [33]. This typically cannot be reduced in the same manner as the epistemic uncertainty. The aleatoric uncertainty can be further categorized into heteroscedastic and homoscedastic uncertainty. In particular, homoscedastic uncertainty is the uncertainty when the noise is assumed to be identical for all points in the data, whereas heteroscedastic uncertainty is the uncertainty when the noise is assumed to be variable across the data.

III. Study Area and Data

The area of interest for this study consists of Baffin Bay, Davis Strait, and most of Nares Strait towards the the Lincoln Sea (Figure 1). This area is primarily a seasonal ice zone. It is chosen because it is a region experiencing significant declines in SIC [49], and increases in shipping activity [50], and contains part of the Tallurutiup Imanga National Marine conservation area, which is an important habitat for marine mammals and seabirds [49], [51].

The data utilizes the full years of 2014 and 2015 and is comprised of a total of six geophysical variables as our features, which are 10-meter wind speed (WS) (m/s), integrated total column water vapour (WV) (mm), integrated total column liquid water (LW) (mm), 2-meter air temperature (AT) (K), vertical brightness temperature (TBV) (K), and horizontal brightness temperature (TBH) (K). These are used as input during training, validation, and testing of the model. The 2014 data is used solely in training/validation, with an 80/20 split, while all of 2015 is used in testing (inference).

The variables WS, WV, and LW are commonly used to apply corrections to TBs in SIC retrieval methods [11]. AT was chosen because of its relationship on atmospheric emission through CW and WV [36] as well as surface emission and melt. These atmospheric variables (WS, WV, LW, and AT) are acquired from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5th Generation (ERA5) data at the Copernicus climate data store [52]. The WS is calculated by taking the magnitude of the 10-meter meridional-component and 10-m zonal-component. All variables in the ERA5 reanalysis data have a nominal spatial resolution of 31 km, and are provided hourly over a 24 hour period. For the present application, a single average over the full 24 hours of a given day is used.

Brightness temperatures at both horizontal and vertical polarization from the Advanced Microwave Scanning Radiometer 2 (AMSR2) are used in this study, with a nominal gridded resolution of 12.5 km [53]. For choice of frequency band to use as input to the MLP, we look to utilize one with a smaller instrument field of view, corresponding to higher spatial resolution of the data. AMSR2 has channels corresponding to frequencies of 18.7, 36.5 and 89 GHz, with the higher frequency 89 GHz channels having a finer spatial resolution than the 18.7 and 36.5 GHz channels. However in addition to scattering from rain and cloud ice particles, the emission at 89 GHz is very sensitive to atmospheric water vapor and cloud liquid water. Thus the frequencies of 18.7 and 36.5 GHz were chosen due to their lower sensitivity [35], consistent with earlier studies [54]. Data from lower frequencies channels are also available across historical passive microwave sensors, enabling the method to be more easily extended to climate data records.

Values of SIC are required as labels during training, and thus for the first dataset used, the SIC values are calculated by the NT2 algorithm using AMSR2 TBs. These SIC data are used solely as training labels during training and validation of the model, and are not used as an input feature. In the NT2 algorithm [55], SICs are estimated using gradient ratios (GRs), and polarization ratios (PRs) calculated using 19 H/V, 37 V, and 89 H/V GHz channels. In the Arctic, three types of ice are treated separately (multi-year, first-year, and thin ice). For each ice type, GRs and PRs are calculated for 12 representative atmospheric profiles using a radiative transfer model to first calculate the brightness temperatures. A lookup table is built from these GRs and PRs. The ice concentration is calculated for a given pixel by choosing the concentration value that matches the actual and estimated PR and GR. This algorithm uses 89 GHz frequency to differentiate areas with low ice concentrations from those with refreezing-related surface effects (snow layering and glazing). It also allows for a specific new ice tie point to be considered, depending on GR37V19V, based on the emissivity difference at the frequencies for new ice.

Finally, a nearest neighbor interpolation scheme is used to upsample the ERA5 reanalysis data to respective brightness temperature sets. Specifically, the ERA5 data is upsampled to match the 12.5 km nominal spatial resolution of the AMSR2 brighness temperatures. As each feature has different ranges of values, each feature is normalized to values between [0,1] prior to input to the MLP.

IV. Methodology

The methodology of this paper is comprised of parts to predict sea ice concentration, as well as the measurement of epistemic and aleatoric uncertainty. As noted, the problem of predicting SIC can be framed as solving a non-linear regression problem, from which an MLP architecture is appropriate and chosen. Next, by utilizing Bayes by Backprop [32], we can

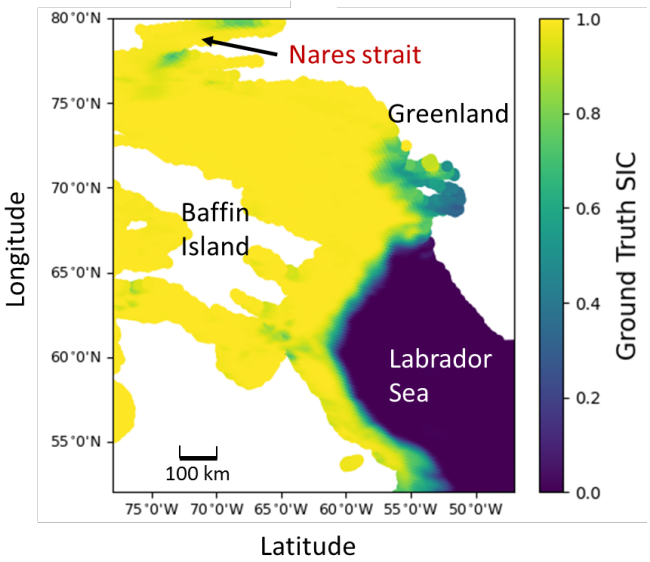

Fig. 1: Study area map showing the monthly average SIC for May 2015, where SIC is calculated by the NT2 algorithm, having a nominal spatial resolution of 12.5 km. White regions are land.

transform non-bayesian models like our MLP into bayesian models (i.e. BNN) such that the weights are represented as probabilistic distributions allowing us to measure the epistemic uncertainty in our model, but preserving the original architecture. Following this, aleatoric uncertainty is captured by transforming a BNN into a heteroscedastic BNN (HBNN) while using the heteroscedastic loss function [33]. We point to the original papers for full technical details but provide a brief overview of each methodology here.

A. Bayes by Backprop

The weights and biases of a NN, such as an MLP, are generally provided as deterministic values [39]. One way to quantify the uncertainty in the model is to instead use probabilistic distributions to represent the weights. For example, allowing each weight to assume a Gaussian distribution with a mean and variance that can be learned.

Bayes by backprop (BBB) provides a method to achieve this utilizing variational inference [32]. Given the weights $\mathbf{w}$ of our model, to attain distributions over $\mathbf{w}$ requires the learning of the posterior predictive distribution (PPD) $P(\mathbf{w}|\mathcal{D})$, where $\mathcal{D}$ is the data characterized by ($\mathbf{x}$, $\mathbf{y}$). In the context of this study, $\mathbf{x}$ are the input features of geophysical variables, and $\mathbf{y}$ SIC training labels. Due to the number of weights in a neural network, to calculate the PPD would need immense computational resources and is intractable. Thus, variational inference is used to reformulate the problem through minimization of the Kullback-Leibler (KL) divergence, requiring the use of an approximate distribution $q(\mathbf{w}|\theta)$ instead of the PPD, $P(\mathbf{w}|\mathcal{D})$. To find the optimal θ, the following tractable loss function as minimized in BBB is used [32],

$$L_{BBB} = \frac{1}{N_{MC}} \sum_{j=1}^{N_{MC}} [\log q(\mathbf{w}_j|\theta) - \log P(\mathbf{w}_j) - \log P(\mathcal{D}|\mathbf{w}_j)]. \tag{1}$$

Here N_{MC} is the number of sample weight vectors $\mathbf{w}_j$ drawn from the variational posterior using a Monte Carlo (MC) sampling scheme. In equation 1 the prior component is defined to be a scaled mixture of Gaussian's following [32] while the likelihood is the MSE loss. The variational posterior is parameterized by $\theta = (\boldsymbol{\mu}, \boldsymbol{\rho})$ where $\boldsymbol{\mu}$ is the mean vector of the distribution and $\boldsymbol{\rho}$ is a hyperparameter used to parameterize the standard deviation vector of the weights $\boldsymbol{\sigma_w}$, such that $\boldsymbol{\sigma_w}$ is always non-negative. Here, $\boldsymbol{\sigma_w}$ is defined pointwise via the softplus function as [32],

$$\boldsymbol{\sigma_w} = log(1 + exp(\boldsymbol{\rho})). \tag{2}$$

To achieve this parameterization, BBB utilizes the local reparameterization trick. This technique allows for $\boldsymbol{\mu}$ and $\boldsymbol{\rho}$ to be incorporated into the weights $\mathbf{w}$ in the model. The parameters $\boldsymbol{\mu}$ and $\boldsymbol{\rho}$ are then updated via backpropagation to find the optimal θ.

B. Heteroscedastic Loss

In BNNs, such as that constructed from BBB, the data dependent portion of the loss function, i.e. the likelihood cost, is measured by the MSE loss which corresponds to a Gaussian distribution with homoscedastic variance. To estimate the uncertainty, the variance is assumed to be non-constant, i.e. heteroscedastic [33], [56], by replacing the MSE loss in the likelihood with the heteroscedastic loss function.

$$L_{HNN} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \Big(\frac{||y_i - \hat{y}_i||^2}{\hat{\sigma}^2_{a(i)}} + log(\hat{\sigma}^2_{a(i)}) \Big), \tag{3}$$

where y_i and $\hat{y}_i$ represent elements of the vectors $\mathbf{y}$ and $\hat{\mathbf{y}}$ respectively, and $\hat{\sigma}_a$ is the aleatoric standard deviation of the model. This term is obtained from a network that has two output neurons, one value for the predicted mean $\hat{y}$, and the other value for $\hat{\sigma}_a$. This value of $\hat{\sigma}_a$ differs from $\hat{\sigma}_w$, since $\hat{\sigma}_w$ is the standard deviation that parameterizes the weight uncertainty. NNs that produce this output are sometimes referred to as Heteroscedastic NNs (HNN), and as we are transforming the BNN from Bayes by backprop to include this additional architecture, it is appropriate to distinguish it as a heteroscedastic Bayesian NN (HBNN), or specifically HBBB. As elaborated in [33], the epistemic and aleatoric uncertainty

for a given testing sample $\mathbf{x}$ can be decomposed from the predictive uncertainty Var($\mathbf{x}$) as

$$\text{Var}(\mathbf{x}) \approx \underbrace{\frac{1}{N_{MC}}\sum_{i=1}^{N_{MC}}\hat{\mathbf{y}}_i^2 - (\frac{1}{N_{MC}}\sum_{i=1}^{N_{MC}}\hat{\mathbf{y}}_i)^2}_{\text{epistemic uncertainty}} + \underbrace{\frac{1}{N_{MC}}\sum_{i=1}^{N_{MC}}\hat{\sigma}_{a(i)}^2}_{\text{aleatoric uncertainty}}. \quad (4)$$

V. Experimental Setup

For our method, we follow [24] and define an MLP to have 4 hidden layers between the input and output layers, with each hidden layer comprised of 10 neurons. Between each layer of the model, a rectified linear unit (ReLU) function is used and for choice of optimiser, we utilise Adam. Through trial and error, we found that a learning rate of 0.001 worked best combined with mini batches comprised of 1000 samples per batch. The data is split into train/validation/test sets and are described in the data section. Note that this MLP architecture is chosen for its efficiency and quality predictions of SIC [24]. A summary of the MLP hyperparameters can be found in Table I.

For BBB, hyperparameter tuning was also required. For N_{mc} it was found that values between 10-50 provided robust results, with any values greater than 50 only providing minimal improvements, but greater computational requirements. Thus an appropriate value of N_{mc} to set to best balance computational efficiency and garner good results was 30. Prior parameters (σ_{p1}, σ_{p2}, π) were chosen, based on Blundell *et al.*. [32]. Subsequently, the mean $\boldsymbol{\mu}$ of the variational posterior is initialized to be a vector of 0's for all weight values, such that the distribution is centered, while $\boldsymbol{\rho}$ is set to -2 [32]. A summary of the BBB hyperparameters can be found in Table II. To then measure aleatoric uncertainty, we replace the MSE loss of BBB, with the heteroscedastic loss, producing a HBNN, or HBBB specifically.

Lastly, we perform experiments as analyzed in Section VI, with models using varying combinations of input features and training labels (Table III). Here MLP denotes models trained only on an MLP, with no capabilities to measure epistemic nor aleatoric uncertainty. Models denoted as BBB are Bayes by backprop based, which measure epistemic uncertainty. Models with the added H, as in $HBBB$ utilize the heteroscedastic loss to produce a heteroscedastic neural network with different combinations of features. The full implementation of the model used in this study is programmed in Python, using the machine learning libraries scikit-learn, and PyTorch.

TABLE I: Hyperparameter Summary for MLP Architecture

Input Dimensions	8
Hidden Layers	4
Neurons in each layer	10
Output Dimension	2
Learning Rate	0.001
Batch Size	1000
Activation Function	ReLU
Optimizer	Adam

TABLE II: Hyperparameter Summary for BNN

Initial Posterior Mean (μ)	0.0
Initial Posterior Rho (ρ)	-2
Prior sigma 1 (σ_{p1})	1.0 [32]
Prior sigma 2 (σ_{p2})	0.01 [32]
Prior pi (π)	0.5 [32]
Monte Carlo Samples (N_{mc})	30

VI. Results & Analysis

The results for this study are subdivided into parts and are described as follows. We first explore the $HBBB$ as it performs over the ice season for the test year of 2015 with respect to the average uncertainty and average root mean square error (RMSE) per month, where the RMSE is a measure of the differences between the predicted SIC and SIC labels from the test year. We then compare $HBBB$ inference capabilities to other chosen methodologies for May of 2015, as this month was found to have the highest uncertainty and second high RMSE compared to other months. Spatial evaluation on the $HBBB$ model as also performed by producing maps of average monthly predicted SIC, RMSE, and uncertainties during the melting season (i.e., April, May, June, and July of 2015). The model is also tested an independent region, (i.e., the Western Arctic, which was not in the training data), and obtains consistent estimation accuracy. This region includes the western part of the Canadian Arctic Archipelago and a portion of the Beaufort Sea, correpsonding to the Western Arctic ice charting region used by the Canadian Ice Service. This region contains both first-year and multiyear ice. In addition, we analyze combinations of input features using the model of $HBBB$. In the last result subection, estimates derived from the NT2 algorithm and the HBBB model are compared with SIC retrieved from the Landsat sensor to further validate the effectiveness of estimation accuracy and uncertainties.

A. Monthly Observations

When observing the RMSE as produced by the $HBBB$ model in Fig. 2, a pattern is apparent that shows a gradual increase of RMSE starting from January, peaking in May, declining to the bottom in October, and slowly increasing again in November and December. The lowest RMSE values correspond to the months of October and September, while the highest are found in the months of April and May. Compared to the RMSE, a similar pattern can be observed for the total uncertainty (where total uncertainty is equal to the sum of epistemic and aleatoric uncertainty), but with smaller differences in magnitude. Solely observing the epistemic uncertainty, it is apparent that the highest epistemic uncertainty (model uncertainty) falls within the months of March, April, and May. The aleatoric uncertainty (data uncertainty) peaks in April, and shows comparable peaks in May and December. It is

TABLE III: Models and Data Summary

Model Name	Uncertainty Measure	SIC Label	TB sensor	Input Features
MLP	N/A	NT2	AMSR2	TBH, TBV, (both 18.7 and 36.5 GHz) AT, WS, WV, LW
BBB	Epistemic	NT2	AMSR2	TBH, TBV, (both 18.7 and 36.5 GHz) AT, WS, WV, LW
$HBBB$	Epistemic + Aleatoric	NT2	AMSR2	TBH, TBV, (both 18.7 and 36.5 GHz) AT, WS, WV, LW
$HBBB_{only-tb}$	Epistemic + Aleatoric	NT2	AMSR2	TBH, TBV (both 18.7 and 36.5 GHz)
$HBBB_{no-tb}$	Epistemic + Aleatoric	NT2	N/A	AT, WS, WV, LW

noteworthy that the highest values for RMSE and uncertainty are in April and May, respectively, which coincides with melt onset in this region [57], and the lowest values coincide with times of low SIC, even in the absence of weather filters or TB corrections, as used in other studies [11], [15]–[17], [19]. We also observe that the aleatoric (data) uncertainty estimated by the model shows similarity amongst months in the same seasonal time frames (i.e. melt vs. freeze up periods), data for these groups of months may have comparable properties and noise.

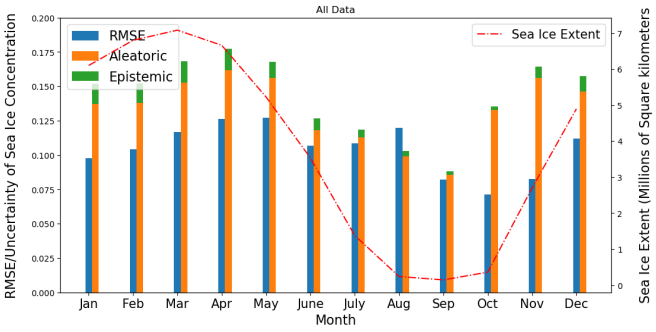

Fig. 2: Root mean squared error (RMSE), epistemic uncertainty, aleatoric uncertainty using the $HBBB$ model. The aleatoric and epistemic uncertainty are shown as a stacked bar, representing total uncertainty. The sea ice extent (SIE) values averaged per month over the study region for the testing of year of 2015 are also shown for reference to the seasonal cycle. Note the RMSEs and uncertainties are highest during May and April, respectively, which corresponds to melt onset in this region [57].

B. Performance comparison between methods

Across the methods used to estimate SIC, per Table IV (where best results are bolded), there is a gradual decrease in the RMSE with increasing model complexity. The RMSE is highest for the MLP model, and lowest for the $HBBB$ model. This provides confidence that each methodology improves the accuracy of SIC. The epistemic uncertainty is nearly 0.066 for the BBB or 6.6% in May, while $HBBB$ with all the features as input reduces this uncertainty to 1.2%.

C. Spatial evaluation of SIC and uncertainties

The spatial distribution of predicted SIC from the $HBBB$ model for April, May, June, and July 2015, shown in the second row Fig. 3, is in good agreement with the expected distribution (the first row of Fig. 3), with areas of consolidated ice in the northern portion of Baffin Bay, and open water farther south. It also captures the predicted SIC of the sea ice in Nares Strait near 80°N.

As shown in the RMSE maps in the third row of 3, where the $HBBB$ model predictions differ from the expected SIC is in areas near the ice edge, such as in the southern portion of Baffin Bay near Davis Strait, which can be defined to be areas where SIC is between 15% and 85% [58]. The aleatoric uncertainty of these areas (shown in the fifth row of 3) is relatively high, corresponding to values in the range of 3-4%. As for the epistemic uncertainty, regions dominated by open water with very low SIC tend to have the lowest values. The distribution patterns of uncertainty can be illustrated further when plotting the epistemic and aleatoric uncertainties for various SIC bins, as seen in Fig. 5. Both types of uncertainty are lowest in SIC bins of 0 to 0.3, increasing in mean uncertainty as the SIC increases. Aleatoric uncertainty values reach the highest in the SIC bins of 0.7-0.8, and decrease for the SIC bin of 0.9-1.0. SIC values of 0.9-1.0 have epistemic uncertainty and aleatoric uncertainty values higher than those for SIC bins between 0.0 and 0.3, possibly reflecting the presence of leads in the ice cover, or changing surface emissivity due to increased snow wetness and melt.

From spatial patterns of the aleatoric uncertainty in the $HBBB$ algorithm observed in Fig. 3, the uncertainties are highest along the MIZ and ice edge. Lastly, both aleatoric and epistemic uncertainties have slightly elevated values along the coastlines. When observing feature maps of TB (Fig. 6(a), (b), (c), and (d)), these elevated uncertainties may be due to land contamination from the TB data for pixels that overlap with the land-ocean boundaries. In addition, to further validate the generalization ability of the model, the model is also tested on another region, i.e., Western Arctic area. Estimated SIC, RMSE and uncertainties averaged over June, July, and August in 2015 are presented in Fig. 4 below. The RMSE values obtained from testing results on data collected from June, July, and August are 0.124, 0.100, and 0.154, respectively. The relatively low estimation error proves that the model is able to perform well in another region.

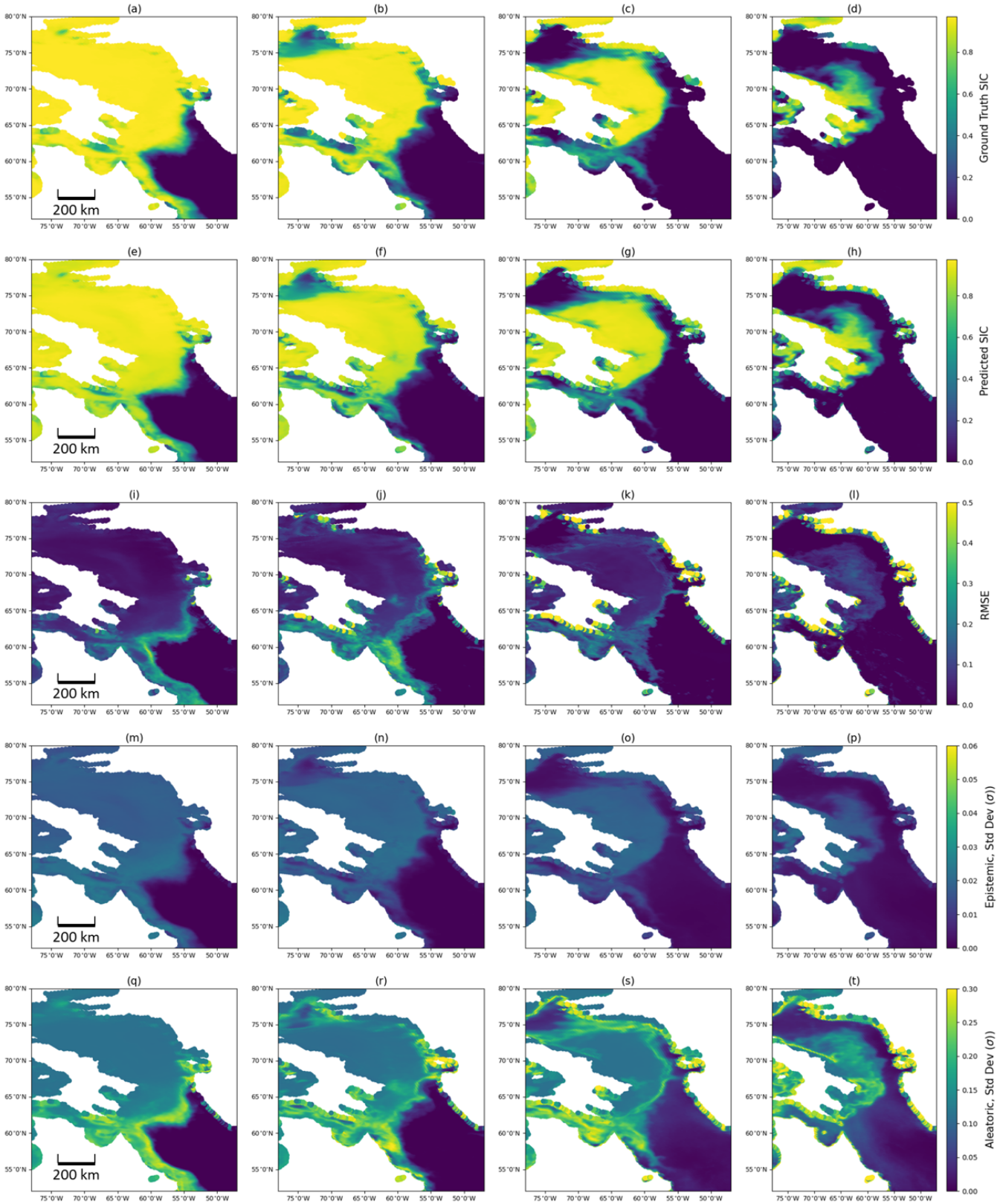

Fig. 3: Ground truth SIC, Estimated SIC, RMSE and uncertainties averaged over June, July, and August 2015 using data collected in Western Arctic region. All models use HBBB model with all 8 features as input data. The first row corresponds to ground truth SIC. The second row corresponds to predicted SIC. The third row corresponds to RMSE values as calculated between the predicted SIC and ground truth SIC. The fourth and fifth rows correspond to the epistemic and aleatoric uncertainties.

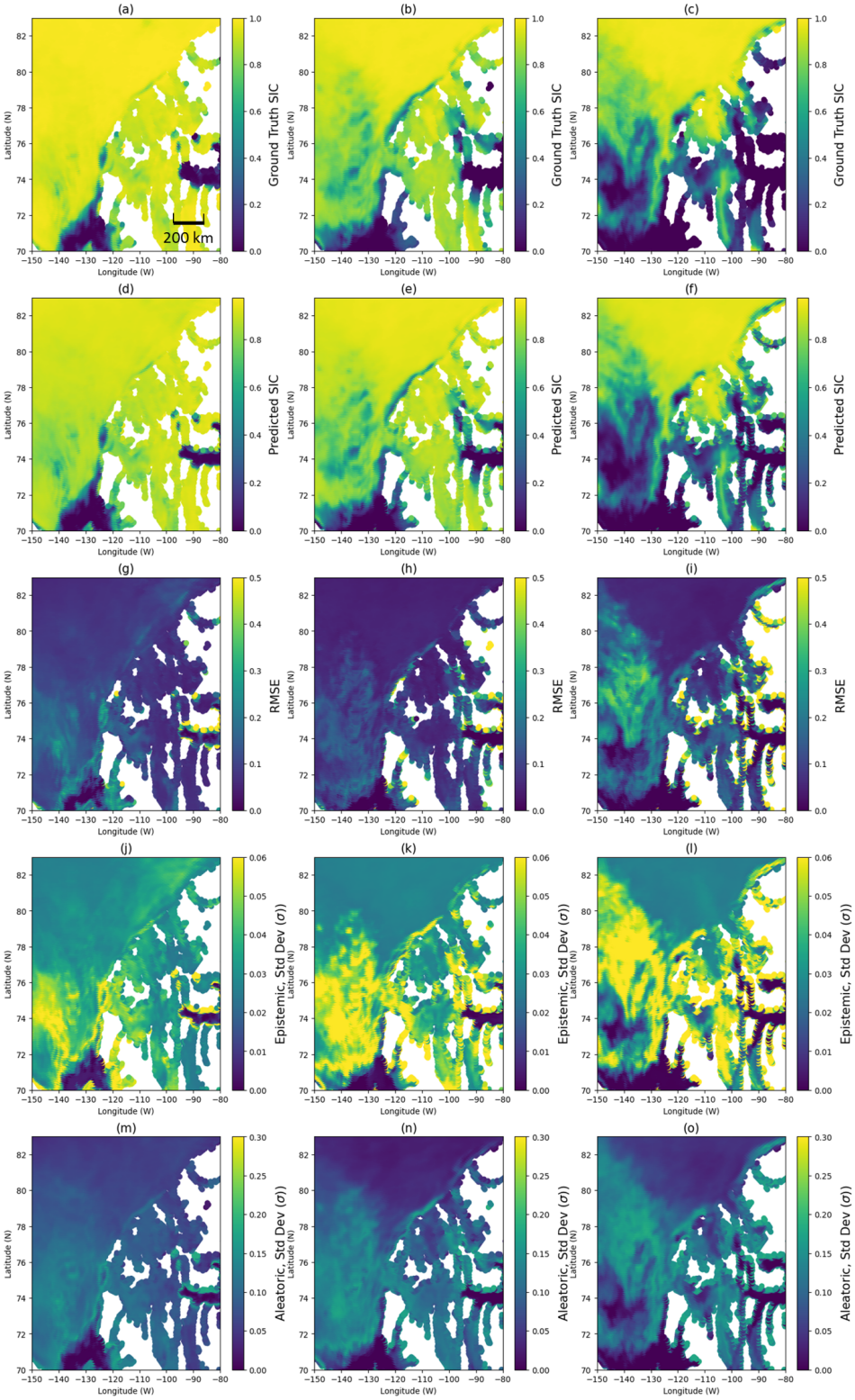

Fig. 4: Ground truth SIC, Estimated SIC, RMSE and uncertainties averaged over June, July, and August 2015 using data collected in the Western Arctic region. All models use the HBBB model with all 8 features as input data. The first row corresponds to ground truth SIC. The second row corresponds to predicted SIC. The third row corresponds to RMSE values as calculated between the predicted SIC and ground truth SIC. The fourth and fifth rows correspond to the epistemic and aleatoric uncertainties.

TABLE IV: Results Summary Between Methods for May of 2015

Model	RMSE		Epistemic		Aleatoric	
	May average	Annual average	May average	Annual average	May average	Annual average
MLP	0.283	0.257	N/A	N/A	N/A	N/A
BBB	0.260	0.237	0.066	0.042	N/A	N/A
$HBBB$	**0.127**	**0.105**	0.012	0.0095	0.156	0.133
$HBBB_{only-tb}$	0.189	0.171	0.030	0.018	0.145	0.148
$HBBB_{no-tb}$	0.510	0.439	0.080	0.080	0.448	0.441

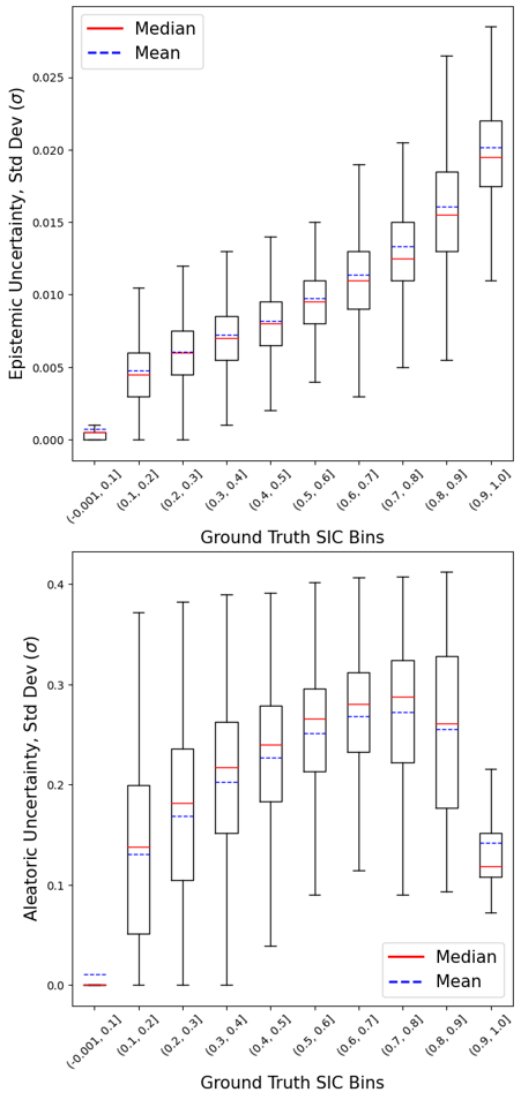

Fig. 5: Box whisker plots of ground truth SIC (from NT2) vs epistemic uncertainty (top) and aleatoric uncertainty (bottom) from the $HBBB$ model for May 2015. The ground truth SIC is divided into 10 equal sized bins. The box whisker plots show the mean, median, interquantile range, and extremes of uncertainties for each SIC bin.

D. Analysis on the influence of input feature combinations

Results calculated from testing samples with HBBB using different combinations of features are presented in Table IV. It shows that compared to the model that only uses brightness temperature as input features (i.e., $HBBB_{only-tb}$), incorporating atmospheric variables further improves estimation accuracy and decrease uncertainties. The figure illustration of SIC prediction of $HBBB_{only-tb}$ (Fig. 7(a)) shows the model can predict the structure of the large ice covered region, although there is also some SIC predicted in the southern portion of the domain, which is expected to be open water. Observing the brightness temperatures, TBH and TBV, in Fig. 6(a), (b), (c), and (d), it can be seen there are TB signatures in the open water that are similar to those in the marginal ice zone. This overlap in feature space is why the model has difficulty differentiating between intermediate ice concentrations and weather-impacted TBs without the atmospheric data.

Following these observations from the model using only TBs as input features, we turn to those from the model only features from the reanalysis data, $HBBB_{no-tb}$. In this case the SIC prediction (Fig. 7(b)) shows the model cannot capture the same detail as that of $HBBB$ (Fig 3(b)), with high levels of RMSE. As per Figs. 6(e)-(h), none of the features from the reanalysis data delineate the SIC and open water regions as clearly as the TBs (Figs. 6(a)-(d)). This is because none of the atmospheric variables are direct measurements from the sea ice, whereas TBs are. In addition to the high RMSE, $HBBB_{no-tb}$ model also suffers from high epistemic uncertainty values.

We now turn our attention to look at specific details of the features as they are related to the monthly trends. Heatmaps of TBH and TBV (Figs. 8(a)-(d)), show the monthly total (epistemic and aleatoric) uncertainty over the test year. It can be seen that the features have their highest total uncertainties in the months of April, May, and June. WS values of 4.3-6.4 m/s are associated with highest total uncertainty, while WV has the highest total uncertainties for values between 0.4-4.3 mm. The final two features of CW and AT have high cumulative uncertainty in bins of 0-0.1 mm and 271-278 K respectively. These observations suggest that these particular values of features contribute to conditions where the AT is around the freezing point in spring, or melt onset. Additionally, all features, except CW, have some distribution of uncertainty in various bins across months. Cloud water on the other hand,

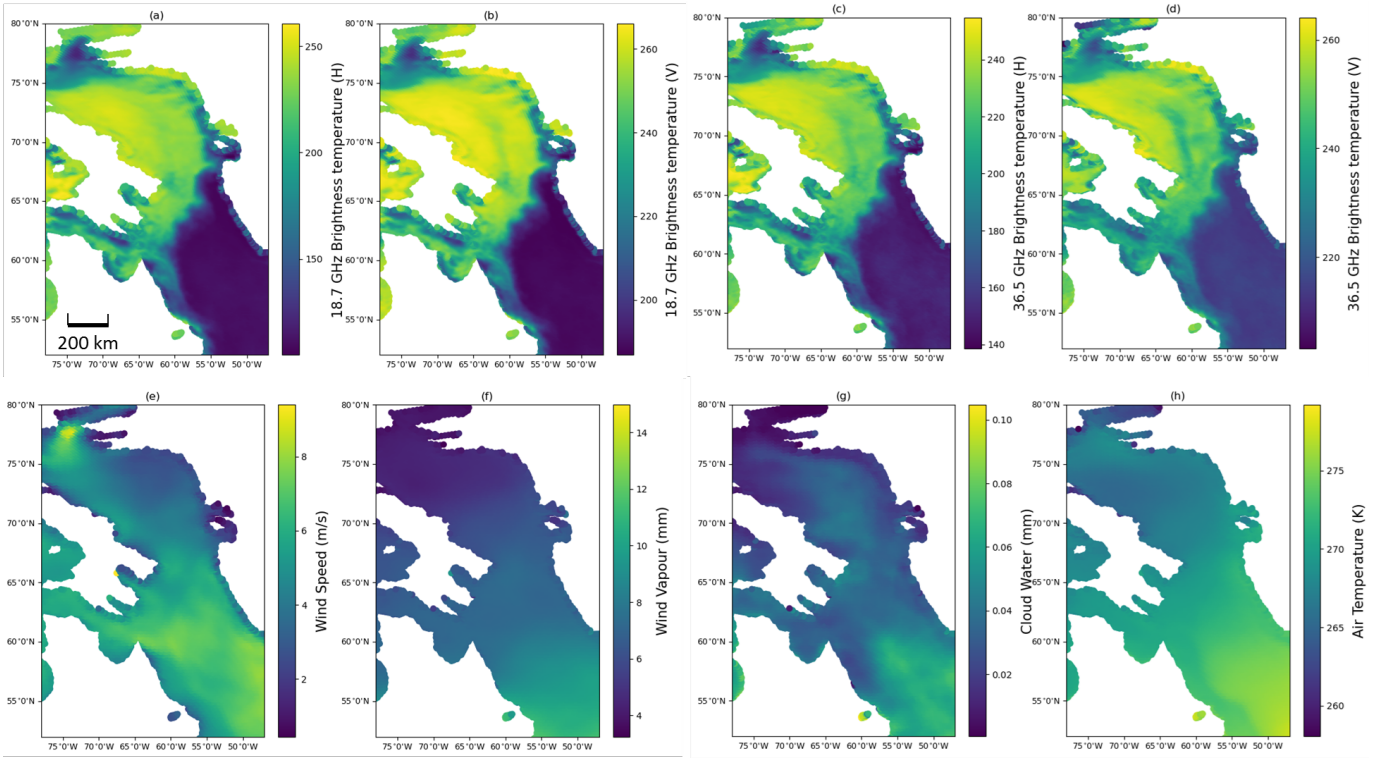

Fig. 6: Maps of input features over the month of May 2015. The features are (a) Brightness Temperature (H) (TBH), (b) Brightness Temperature (V) (TBV), (c) Windspeed (WS), (d) Water Vapour (WV), (e) Cloud Water (CW), (f) Air Temperature (AT).

has highest uncertainty in the bins of 0-0.1 mm regardless of the month. This is consistent with the view [11] that CW is of limited use in TB corrections for sea ice concentration retrievals because CW estimates from reanalysis or numerical weather prediction are not yet sufficiently reliable in ice-covered regions.

E. Comparison with Landsat SIC

To further validate the effectiveness of estimation accuracy and uncertainties, the estimates derived from the NT2 algorithm and the HBBB model are compared with the high-resolution (30 m) binary sea ice cover maps estimated from satellite observations acquired in the visible frequency range by the joint NASA and United States Geological Survey (USGS) Landsat sensor (Landsat-8) [59]. To derive SIC measurements from Landsat scenes, the number of sea ice and open water pixels within the resolution of 12.5 km are first averaged. Then, a nearest neighbor interpolation scheme is used to downsample the HBBB SIC to the respective Landsat SIC for comparison. As shown by the examples in Fig. 9, the RMSE values calculated between NT2-derived SIC VS Landsat-derived SIC and HBBB-derived SIC VS Landsat-derived SIC are very close to each other. This demonstrates that the proposed model is capable of estimating reasonable SIC, compared with independent validation data (see Fig.9).

VII. Discussion

The analysis of predictions from $HBBB$ over the annual cycle shows the highest values for RMSE and uncertainty are in May and April, respectively, which coincides with melt onset in this region [57]. Within this period, significant changes in brightness temperatures are known to lead to problems with SIC retrievals [12], [18], reflected here as higher RMSEs and greater uncertainties. The model also showed promise in predicting SIC in thin ice periods, as is the case for November and December. The lowest values of SIC uncertainty coincide with times of low SIC, without an apparent need for weather filters or TB corrections, as used in other studies [11], [15]–[17], [19], A more thorough investigation of weather impacts will be done in a future study. The seasonal variability of uncertainty is consistent with results from uncertainty in the NT2 algorithm [31]. At these times the uncertainty is relatively high, but the RMSE is around 10%. Estimated aleatoric uncertainties were higher for intermediate ice concentration values, similar to an alternative approach [19] in which SIC uncertainty is based on tie-point standard deviations and smearing due to the large spatial footprint associated with the PM measurements.

The comparison between methodologies showed the use of heteroscedastic loss as opposed to homoscedastic loss in the model produces lower model uncertainty as well as greater accuracy to predictions. Other methods such as Monte Carlo

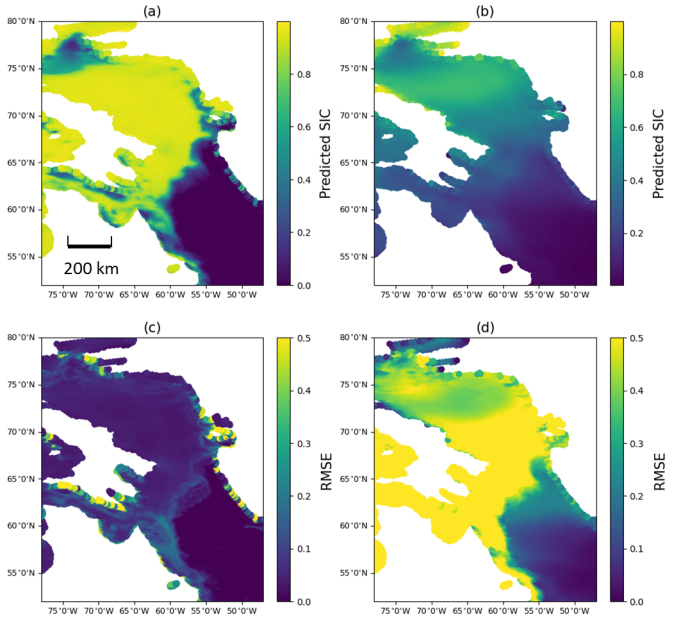

Fig. 7: Spatial maps of results from the HBBB model averaged over the month of May 2015 utilizing different combinations of input features. The first row corresponds to predicted sea ice concentrations (SIC), the second row corresponds to the root mean squared error (RMSE) values as calculated between the predicted SIC and ground truth SIC. Columns in the figure correspond to the combinations of input features. The first column are the results from the $HBBB_{only-TB}$ model using only TBH and TBV from AMSR2 as input. The second column are the $HBBB_{no-TB}$ model results using WS, WV, CW, and AT as input.

dropout for measuring epistemic uncertainty were explored but did not show capabilities to effectively predict SIC and estimate epistemic uncertainty for these predictions. Ensemble methods were also considered, but required greater computational costs then that of a Bayesian neural network, and due to limited computational power were not used.

The analysis on input feature combinations showed the 4 features of wind speed, column water vapour, liquid water, and air temperature are able to correct the over-prediction of sea ice concentration in the open water regions as reflected by the reduction in RMSE. Both $HBBB_{no-tb}$ and $HBBB_{only-tb}$ models have non-zero aleatoric uncertainty over the open water, which is reduced to zero when all 8 features are used as input.

VIII. Conclusions

This study has conveyed the capabilities of a MLP model using PM-TB and atmospheric variables to predict SIC, while quantifying both epistemic (model) and aleatoric (data) uncertainty. The model was evaluated over an annual cycle in a seasonal ice zone. It was found uncertainties vary seasonally, where highest uncertainties correspond to times of expected melt onset in the region. The spatial distribution of uncertainty showed highest values were in the MIZ and along the land-ocean boundary. Both brightness temperatures and atmospheric variables were found necessary to produce realistic distributions of SIC.

An advantage of the approach presented in this study is its data-driven nature. Given new data, uncertainty estimates are specific to that data instance, which means the spatial uncertainty maps reflect local conditions, such as changes in surface conditions or local atmospheric moisture. Once the model is trained, inference, which is the production of a SIC estimate and its uncertainty, is very efficient, with daily results produced for entire year of 2015 in minutes. Additionally, the decomposition of uncertainty into an epistemic (model) contribution and an aleatoric (data) term, is novel for this application domain, though it has been used in related studies [34] and is worth further investigation. In a future study we will also compare this method to others in the literature to better under the advantages and disadvantages of the approach over a wider range of conditions, considering further validation with independent data. The role of air temperature and its correlation with ice concentration and surface melt will also be considered.

References

[1] A. Schweiger, "Accelerated sea ice loss in the Wandel Sea points to a change in the Arctic's last ice area," *Earth and Space Science Open Archive*, 2021.

[2] N. S. Steiner, J. Bowman, K. Campbell, *et al.*, "Climate change impacts on sea-ice ecosystems and associated ecosystem services," *Elementa: Science of the Anthropocene*, vol. 9, no. 1, Oct. 2021, ISSN: 2325-1026.

[3] A. Lovecraft, "The human geography of Arctic sea ice: Introduction," *Polar Geography*, vol. 36, no. 1-2, pp. 1–4, 2013.

[4] L. Wang, K. Scott, L. Xu, and D. Clausi, "Sea ice concentration estimation during melt from dual-pol SAR scenes using deep convolutional neural networks: A case study," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, pp. 1–10, Apr. 2016.

[5] Q. Yan and W. Huang, "Sea ice sensing from GNSS-R data using convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1510–1514, 2018.

[6] F. Ulaby, M. R.K., and A. Fung, *Microwave Remote Sensing - Active and Passive*. Artech House, 1981, ISBN: 9780198520115.

[7] F. D. Carsey, *Microwave Remote Sensing of Sea Ice, Volume 68* (Geophysical Monograph Series). American Geophysical Union, 1992, ISBN: 9780875900339.

[8] J. C. Comiso, "Enhanced sea ice concentrations from passive microwave data," *NASA Goddard Space Flight Center, Greenbelt, Maryland*, vol. 20, 2007.

[9] T. Markus, D. J. Cavalieri, and A. Ivanoff, *Algorithm theoretical basis document: Sea ice products: Updated December 2011*, 2011.

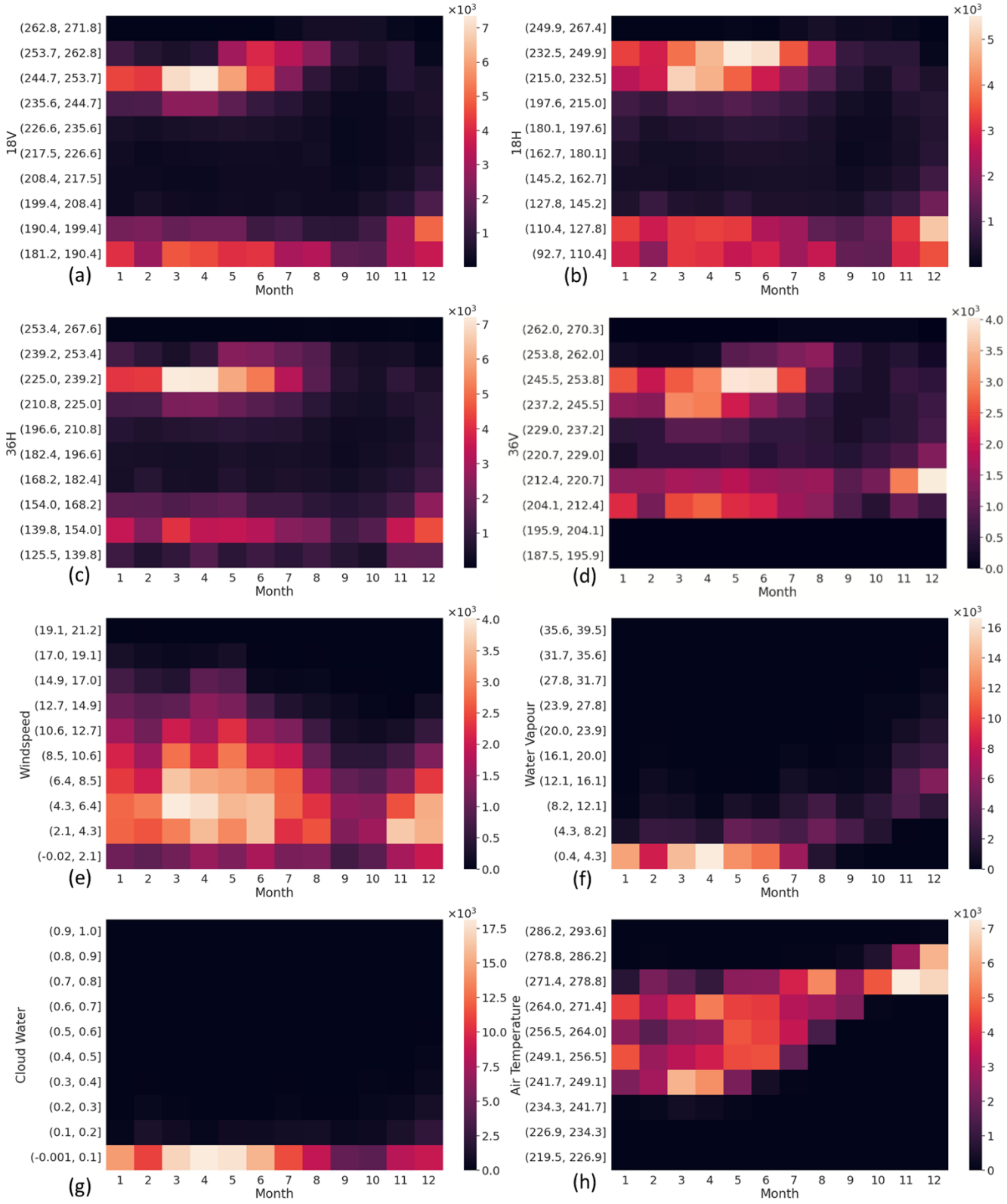

Fig. 8: Heatmaps of input features, plotted with bins of their respective values against month in the Eastern Arctic. The colorbars represent cumulative total uncertainty as outputted by the HBBB model. For most features, April and/or May distinctly show the highest cumulative uncertainty.

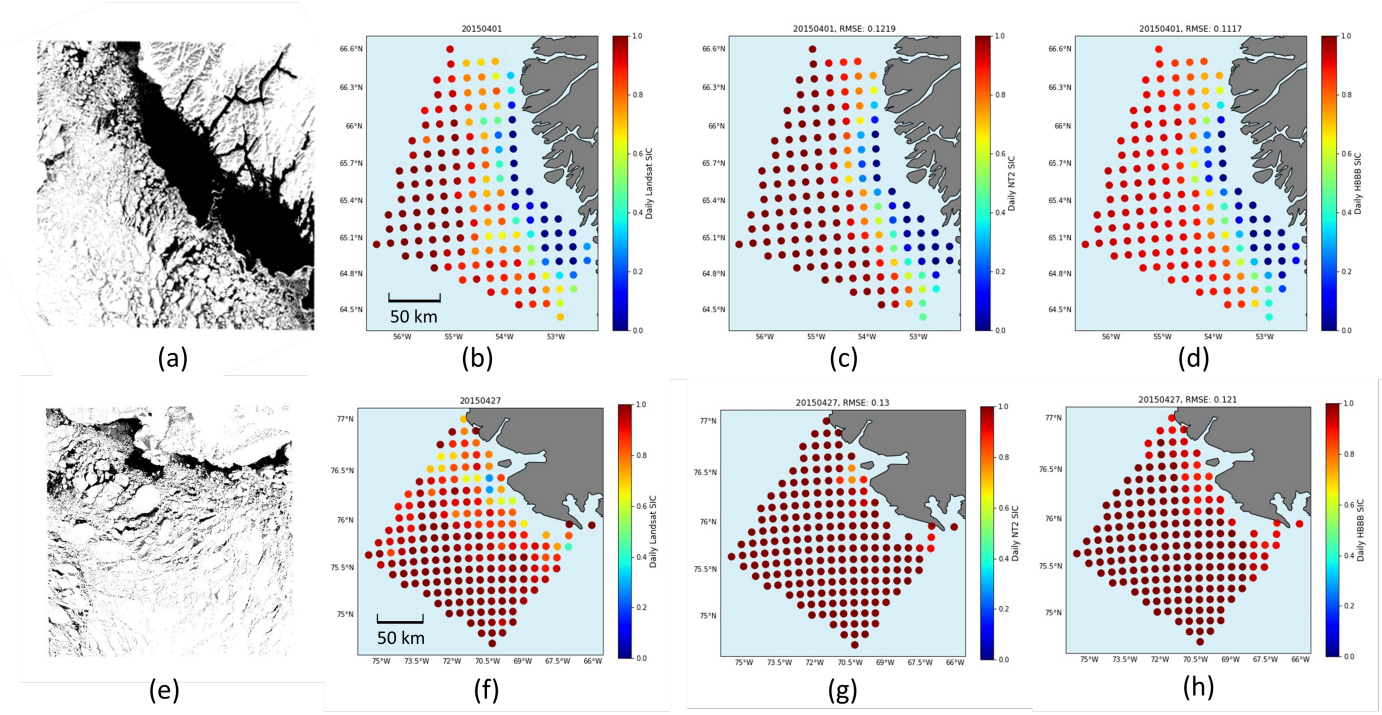

Fig. 9: Two examples of the Landsat image (first column) and SIC maps derived from Landsat (second column), PM NT2 (third column), and HBBB (fourth column) in the area covered by (a) on 20150401 and (b) 20150427 along the Greenland coast. The RMSE values are calculated between the Landsat-derived and NT2-derived SICs, and between the Landsat-derived and HBBB-derived SICs. Regions without data are shown in light blue, while the lands are in grey.

[10] N. Ivanova, O. Johannessen, L. Pedersen, and R. Tonboe, "Retrieval of Arctic sea ice parameters by satellite passive microwave sensors: A comparison of eleven sea ice concentration algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7233–7246, 2014.

[11] S. Andersen, R. Tonboe, S. Kern, and H. Schyberg, "Improved retrieval of sea ice total concentration from spaceborne passive microwave observations using numerical weather prediction model fields: An intercomparison of nine algorithms," *Remote Sens. Environ.*, vol. 104, pp. 374–392, Oct. 2006.

[12] W. Meier and D. Notz, "A note on the accuracy and reliability of satellite-derived passive microwave estimates of sea-ice extent," CLIC International Project Office, Tech. Rep., Oct. 28, 2010.

[13] H. Han and H. Kim, "Evaluation of summer passive microwave sea ice concentrations in the Chukchi Sea based on KOMPSAT-5 SAR and numerical weather prediction data," *Remote Sens. Environ.*, vol. 209, pp. 343–362, 2018.

[14] S. Kern, T. Lavergne, D. Notz, L. T. Pedersen, and R. Tonboe, "Satellite passive microwave sea-ice concentration data set inter-comparison for Arctic summer conditions," *The Cryosphere*, vol. 14, no. 7, pp. 2469–2493, 2020.

[15] S. Andersen, R. Tonboe, L. Kaleschke, G. Heygster, and L. T. Pedersen, "Intercomparison of passive microwave sea ice concentration retrievals over the high-concentration Arctic sea ice," *J. Geophys. Res.: Oceans*, vol. 112, 2007.

[16] P. Gloersen and D. J. Cavalieri, "Reduction of weather effects in the calculation of sea ice concentration from microwave radiances," vol. 91, no. C3, pp. 3913–3919, Mar. 1986.

[17] D. J. Cavalieri, K. M. St. Germain, and C. T. Swift, "Reduction of weather effects in the calculation of sea-ice concentration with the DMSP SSM/I," *Journal of Glaciology*, vol. 41, no. 139, pp. 455–464, 1995.

[18] N. Ivanova, L. T. Pedersen, R. T. Tonboe, *et al.*, "Inter-comparison and evaluation of sea ice algorithms: Towards further identification of challenges and optimal approach using passive microwave observations," *The Cryosphere*, vol. 9, no. 5, pp. 1797–1817, 2015.

[19] R. T. Tonboe, S. Eastwood, T. Lavergne, *et al.*, "The EUMETSAT sea ice concentration climate data record," *The Cryosphere*, vol. 10, no. 5, pp. 2275–2290, 2016.

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–44, May 2015.

[21] L. Wang, K. A. Scott, and D. A. Clausi, "Sea ice concentration estimation during freeze-up from SAR

imagery using a convolutional neural network," *Remote Sens.*, vol. 9, no. 5, 2017.

[22] C. L. V. Cooke and K. A. Scott, "Estimating sea ice concentration from SAR: Training convolutional neural networks with passive microwave data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4735–4747, 2019.

[23] K. Radhakrishnan, K. Scott, and D. Clausi, "Sea ice concentration estimation: Using passive microwave and SAR data with a U-Net and curriculum learning," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 5339–5351, Apr. 2021.

[24] A. Soleymani and K. A. Scott, "Evaluation of a neural network on sea ice concentration estimation in MIZ using passive microwave data," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 5656–5659.

[25] Y. J. Kim, H. Kim, D. Han, S. Lee, and J. Im, "Prediction of monthly Arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks," *The Cryosphere*, vol. 14, no. 3, pp. 1083–1104, 2020.

[26] H. Han, S. Lee, H.-C. Kim, and M. Kim, "Retrieval of summer sea ice concentration in the Pacific Arctic Ocean from AMSR2 observations and numerical weather data using random forest regression," *Remote Sens.*, vol. 13, no. 12, 2021.

[27] C. J. Merchant, F. Paul, T. Popp, *et al.*, "Uncertainty information in climate data records from earth observation," *Earth System Science Data*, vol. 9, no. 2, pp. 511–527, 2017.

[28] N. Asadi, K. A. Scott, A. S. Komarov, M. Buehner, and D. A. Clausi, "Evaluation of a neural network with uncertainty for detection of ice and water in SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 247–259, 2021.

[29] N. Saberi, K. A. Scott, and C. Duguay, "Incorporating aleatoric uncertainties in lake ice mapping using RADARSAT-2 SAR images and CNNs," *Remote Sens.*, vol. 14, no. 3, 2022.

[30] X. Chen, K. A. Scott, L. Xu, M. Jiang, Y. Fang, and D. A. Clausi, "Uncertainty-incorporated ice and open water detection on dual-polarized SAR sea ice imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023. DOI: 10.1109/TGRS.2022.3233871.

[31] L. Brucker, D. J. Cavalieri, T. Markus, and A. Ivanoff, "NASA Team 2 sea ice concentration algorithm retrieval uncertainty," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7336–7352, 2014. DOI: 10.1109/TGRS.2014.2311376.

[32] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural networks," ser. ICML'15, JMLR.org, 2015, pp. 1613–1622.

[33] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" *CoRR*, vol. abs/1703.04977, 2017.

[34] P. Ortiz, M. Orescanin, V. Petković, S. W. Powell, and B. Marsh, "Decomposing satellite-based classification uncertainties in large earth science datasets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022. DOI: 10.1109/TGRS.2022.3152516.

[35] P. Minnett, A. Alvera-Azcárate, T. Chin, *et al.*, "Half a century of satellite remote sensing of sea-surface temperature," *Remote Sens. Environ.*, vol. 233, p. 111 366, 2019.

[36] F. Wentz and T. Meissner, "AMSR Ocean Algorithm, Algorithm Theoretical Basis Document," Remote Sensing Systems, Tech. Rep., 2002.

[37] D. J. Cavalieri, K. M. S. Germain, and C. T. Swift, "Reduction of weather effects in the calculation of sea-ice concentration with the DMSP SSM/I," *Journal of Glaciology*, vol. 41, no. 139, pp. 455–464, 1995.

[38] G. Spreen, L. Kaleschke, and G. Heygster, "Sea ice remote sensing using AMSR-E 89-GHz channels," *Journal of Geophysical Research: Oceans*, vol. 113, no. C2, 2008.

[39] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by backpropagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[40] P. Wasserman and T. Schwartz, "Neural networks II. what are they and why is everybody so interested in them now?" *IEEE Expert*, vol. 3, no. 1, pp. 10–15, 1988.

[41] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[42] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, vol. 4, no. 2, pp. 251–257, 1991.

[43] B. Kompa, J. Snoek, and A. L. Beam, "Communicating uncertainty in medical machine learning," *npj Digital Medicine*, vol. 4, no. 1, pp. 1–6, 2021.

[44] L. Ding, D. Li, B. Liu, *et al.*, "Capture uncertainties in deep neural networks for safe operation of autonomous driving vehicles," *CoRR*, vol. abs/2108.05118, 2021.

[45] J. Haas and B. Rabus, "Uncertainty estimation for deep learning-based segmentation of roads in synthetic aperture radar imagery," *Remote Sens.*, vol. 13, no. 8, 2021.

[46] M. Abdar, F. Pourpanah, S. Hussain, *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243–297, 2021.

[47] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," 2015. DOI: 10.48550/ARXIV.1506.02142. [Online]. Available: https://arxiv.org/abs/1506.02142.

[48] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation

using deep ensembles," ser. NIPS'17, Curran Associates Inc., 2017, pp. 6405–6416.

[49] K. Laidre, H. Stern, K. Kovacs, *et al.*, "Arctic marine mammal population status, sea ice habitat loss, and conservation recommendations for the 21st century," *Conservation Biology*, vol. 29, pp. 724–737, 3 2015.

[50] L. Pizzolato, S. Howell, J. Dawson, F. Laliberté, and L. Copland, "The influence of declining sea ice on shipping activity in the Canadian Arctic," *Geophysical Research Letters*, vol. 43, pp. 12, 146–12, 154, 23 2015.

[51] W. Halliday, J. Dawson, D. Yurkowski, *et al.*, "Vessel risks to marine wildlife in the Tallurutiup Imanga national marine conservation area and the eastern entrance to the Northwest Passage," *Environmenal Science and Policy*, vol. 127, pp. 181–192, 2022.

[52] H. Hersbach, B. Bell, P. Berrisford, *et al.*, *ERA5 hourly data on single levels from 1979 to present*, version 10-meter meridional-component windspeed, 10-m zonal-component windspeed, integrated total column water vapour, integrated total column liquid water, 2-meter air temperature, Copernicus Climate Change Service (C3S) Climate Data Store (CDS).

[53] W. N. Meier, T. Markus, and J. C. Comiso, *AMSRE AMSR2 unified l3 daily 25 km brightness temperatures, sea ice concentration, motion snow depth polar grids, version 1*. NASA National Snow and Ice Data Center Distributed Active Archive Center, 2018.

[54] R. Scarlat, G. Heygster, and L. Pedersen, "Experiences with an optimal estimation algorithm for surface and atmospheric parameter retrieval from passive microwave data in the Arctic," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 9, pp. 3934–3947, 2017.

[55] T. Markus and D. J. Cavalieri, "An enhancement of the NASA Team sea ice algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1387–1398, 2000. DOI: 10.1109/36.843033.

[56] Q. V. Le, A. J. Smola, and S. Canu, "Heteroscedastic gaussian process regression," ser. ICML '05, Bonn, Germany: Association for Computing Machinery, 2005, pp. 489–496, ISBN: 1595931805.

[57] A. Bliss, M. Steele, G. Peng, W. Meier, and S. Dickinson, "Regional variability of Arctic sea ice seasonal change climate indicators from a passive microwave climate data record," *Environmental Research Letters*, vol. 14, Apr. 2019.

[58] M. Leppranta, "Drift ice material," in *The Drift of Sea Ice*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 11–63.

[59] S. Kern, *Landsat surface type over water from supervised classification of surface broadband albedo estimates*, This is a data set created for the evaluation of sea-ice concentration products derived from satellite passive microwave observations., May 2021. DOI: 10.25592/uhhfdm.9181. [Online]. Available: https://doi.org/10.25592/uhhfdm.9181.