

Article

# Artificial Neural Network for the Short-Term Prediction of Arctic Sea Ice Concentration

Minjoo Choi <sup>1,\*</sup>, Liyanarachchi Waruna Arampath De Silva <sup>2</sup> and Hajime Yamaguchi <sup>2</sup><sup>1</sup> Mathematics and Cybernetics, SINTEF Digital, P.O. Box 124 Blindern, NO-0314 Oslo, Norway<sup>2</sup> Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa 277-8561, Japan; waruna@edu.k.u-tokyo.ac.jp (L.W.A.D.S); h-yama@edu.k.u-tokyo.ac.jp (H.Y.)

\* Correspondence: minjoo.choi@sintef.no; Tel.: +47-405-73-137

Received: 15 March 2019; Accepted: 3 May 2019; Published: 6 May 2019



**Abstract:** In this paper, we applied an artificial neural network (ANN) to the short-term prediction of the Arctic sea ice concentration (SIC). The prediction was performed using encoding and decoding processes, in which a gated recurrent unit encodes sequential sea ice data, and a feed-forward neural network model decodes the encoded input data. Because of the large volume of Arctic sea ice data, the ANN predicts the future SIC of each cell individually. The limitation of these singular predictions is that they do not use information from other cells. This results in low accuracy, particularly when there are drastic changes during melting and freezing seasons. To address this issue, we present a new data scheme including global and local SIC information, where the global information is represented by sea ice statistics. We trained ANNs using different data schemes and network architectures, and then compared their performances quantitatively and visually. The results show that, compared with a data scheme that uses only local sea ice information, the newly proposed scheme leads to a significant improvement in prediction accuracy.

**Keywords:** artificial neural network; gated recurrent unit; Arctic sea ice prediction; short-term prediction

## 1. Introduction

The Arctic region has attracted growing interest because of the declining Arctic sea ice cover. This allows for better accessibility to natural resources in the region and shorter commercial shipping routes, including the Northern Sea Route (NSR) and the Northwest Passage, crossing the Arctic seas. According to Schwarz [1], using the NSR to travel between Yokohama and Rotterdam can reduce the distance by 40% compared with using the conventional southern sea route.

Arctic sea ice conditions can change over short to medium-term timeframes because of dynamic and thermodynamic processes. Therefore, precise sea ice prediction is critical for ships operating in the Arctic seas. This prediction is the basis of operational decisions that determine the efficiency and safety of operating the ships. Although Arctic sea ice prediction may be needed for a variety of purposes, the prime focus of this paper is on the commercial use of the Arctic seas. Arctic sea ice prediction can be categorized as short-, mid-, and long-term [2]. Long-term predictions include several years. These are useful for making decisions regarding the construction of new icebreakers and infrastructures. Several general circulation climate models are available for this purpose, such as the Hadley Centre Global Environmental Model (HadGEM) of The Met Office's Hadley Centre [3], the Community Earth System Model (CESM) of the National Center for Atmospheric Research (NCAR) [4], and the Model for Interdisciplinary Research on Climate (MIROC) of the Japanese research community [5]. Mid-term predictions cover several months and are useful for making decisions on whether to use the NSR or the conventional southern sea route during the coming summer navigation season. These predictions

are performed using statistical or ice-ocean coupled models. Kimura et al. [6] presented a statistical method that predicted the summer sea ice cover using the previous winter sea ice cover and ice motion. A fully coupled model including ice, ocean, atmosphere, and land produced by the National Centers for Environmental Prediction (NCEP) Climate Forecast System version 2 (CFSv2) [7] was used to predict the seasonal sea ice. An ice-ocean coupled model with ensemble prediction was also used for seasonal sea ice predictions [8]. Finally, short-term predictions cover several days. These are useful for selecting a safe and economic path when a ship enters ice-covered areas in the Arctic ocean. A few operational short-term prediction models have been developed for this purpose. Examples are the Japanese ice-ocean coupled model icePOM [2], the U.S. Navy Arctic Cap Nowcast/Forecast System (ACNFS) [9], and the Norwegian TOPAZ4, which is a coupled ocean-sea ice data assimilation system for the North Atlantic and Arctic [10].

With the recent advances in modeling and data processing technologies, artificial neural networks (ANNs) have shown great success in a variety of applications, such as object recognition [11], speech recognition [12], super-resolution imaging [13], and natural language processing [14]. The prime advantage of ANNs is the ability to learn non-linear and complex relationships between inputs and outputs. Because of the recent success in such areas, ANN approaches have been applied to Arctic sea ice concentration (SIC) prediction. Chi and Kim [15] applied a long short-term memory (LSTM) network [16] to the mid-term prediction of Arctic SIC. The LSTM is a recurrent neural network that uses the output from a previous step as the input for the current step. It has a memory cell, which is referred to as the state cell, that is updated at each step using a new input. Such an updating process allows for the LSTM to use only the necessary information rather than all the inputs during training. This makes the training process more efficient and improves prediction accuracy. Unlike existing approaches that use a variety of atmospheric and oceanic data, the authors use only observed SIC data in the prediction. The authors compared the performance of the LSTM with that of an autoregressive (AR) model, which is a statistics-based model for time-series data, and the LSTM outperformed the AR. Kim et al. [17] used a feed-forward neural network for mid-term prediction of SIC in the Kara and Barents seas. The authors selected the following four atmospheric factors as input factors among 19 candidates using the correlation coefficient: (i) near-surface specific humidity; (ii) near-surface air temperature; (iii) surface downwelling longwave radiation; and (iv) surface upwelling longwave radiation. The authors created model input through an ensemble of the data that were provided by multiple regional climate models using the Bayesian model averaging method [18].

Consistent with previous research, we applied an ANN to Arctic SIC prediction. The ANN was for short-term prediction and only used observed SIC data, as described by Chi and Kim [15]. Because of the large volume of Arctic SIC data, the ANN predicted the future SIC of each cell individually. The limitation of such an individual prediction is that it does not use information from other cells. This results in low accuracy, particularly when there are drastic changes during the melting and freezing seasons. To address this issue, we present a new data scheme that enables such individual cell prediction to use global SIC information and the cell's local SIC information. In Section 2, the ANN and the new data scheme are described. In Section 3, we show the test results of the ANN, in comparison with an ANN that was trained by a data scheme that did not include global SIC information. In Section 4, we draw conclusions and provide suggestions for future research.

## 2. Materials and Methods

We trained an ANN for short-term prediction of SIC in the Arctic grid cells. The grid size is  $233 \times 274$  cells and each cell's area is  $25 \times 25$  km<sup>2</sup>. The SIC data used in this study are based on retrievals using brightness temperature from the following passive microwave radiometers: Special Sensor Microwave/Imager (SSM/I), Special Sensor Microwave Imager/Sounder (SSMIS), Advanced Microwave Scanning Radiometer for EOS (AMSR-E), and Advanced Microwave Scanning Radiometer 2 (AMSR2). SSM/I and SSMIS SIC data were obtained from the National Snow & Ice Data Center (NSIDC) [19]. AMSR-E SIC data were also obtained from the NSIDC [20]. AMSR2 SIC data were obtained from

the Arctic Data Archive System (ADS) [21]. The horizontal resolution of the SSM/I SIC data is about 25 km, and that of AMSRE and AMSR2 SIC data is about 12.5 km and 10 km, respectively. Therefore, we interpolated this data into the 25-km grid using the inverse distance interpolation method.

For efficient computation, we selected 20,661 cells in which sea ice had been observed from 1 January 2000 to 31 December 2016. The cells are referred to as ‘dynamic cells’ and account for around 32% of the entire grid cells. We assumed that sea ice would appear only in the dynamic cells, believing that the prediction errors made from this assumption would not be significant. In Figure 1, the white cells indicate the dynamic cells. The black cells indicate ‘ground cells’, which are continental areas and have a positive elevation. The blue cells indicate ‘open-sea cells’, which have a negative elevation, and sea ice was not observed in these areas during this period.

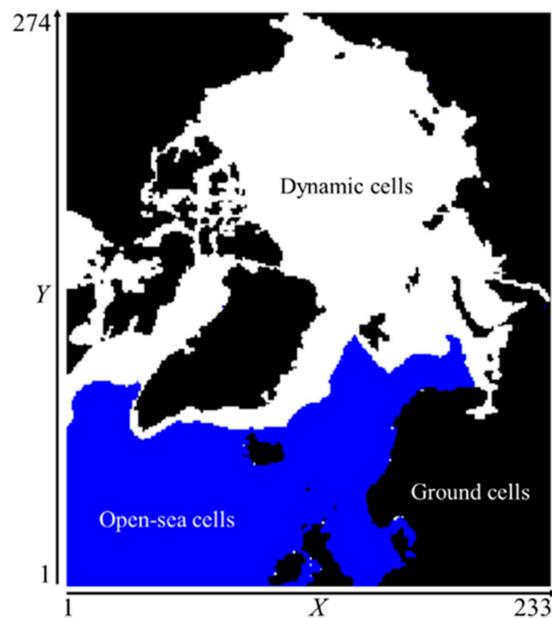


Figure 1. Dynamic, open-sea, and ground cells.

The ANN predicts the SIC of each dynamic cell individually. The prediction period is 15 days using sea ice information from the previous 120 days and geographical information from the cell. The input length, 120 days, was determined using trial experiments, in which we investigated performance using 15, 30, 60, 90, 120, and 150 days of input length, and the best performance was found in 120 days. Figure 2 describes the input and output of the model, in which the values in the white cells indicate local information, and the values in the grey cells indicate global information. The local information is relevant only for an individual cell, such as the geographical information and SIC of the cell. The geographical information indicates the cell’s coordinates and elevation. This allows for one model to be used to predict multiple cells. We used X–Y coordinates instead of longitude and latitude to avoid the discontinuity between the end of the longitude lines. For example, if we represent the longitude lines 179 degrees east and 179 degrees west by the values 179 and  $-179$ , respectively, although these lines are close to each other geographically, they are quite different in their numerical value.

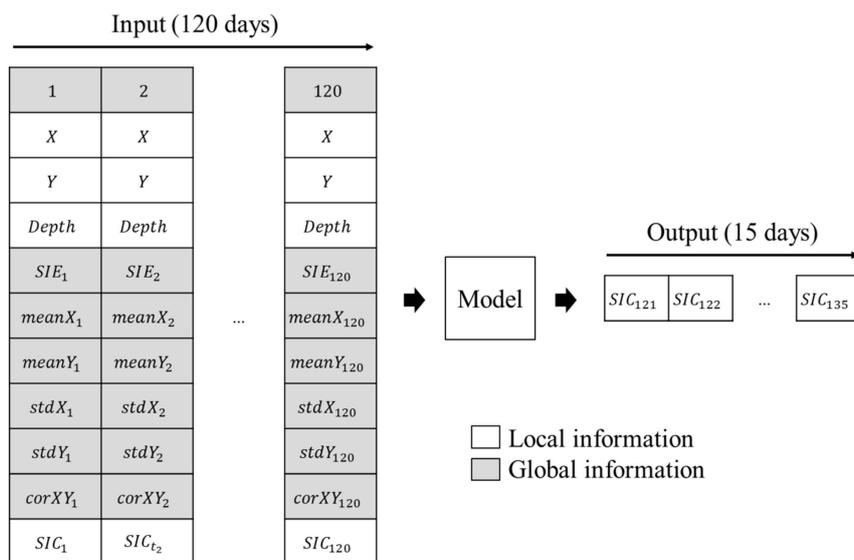


Figure 2. Artificial neural network input and output.

The global information comprised grid-level information. We used the time step  $t$  and six SIC statistics as the global information. The time steps are integers that are the date gaps from a reference date, 1 January 2000. For instance, 2 January 2000 and 3 January 2000 can be converted to integers 1 and 2, respectively. When computing the statistics, we first identified ice-covered cells with a SIC of more than 0.15. We then computed the statistics including the sea ice extent, mean, standard deviation, and correlation of ice-covered cells. The sea ice extent,  $SIE$ , describes the entire sea ice area at time step  $t$  and is the aggregated area of ice-covered cells. This information can be used in the mode description of Arctic sea ice, such as freezing, frozen, melting, and melted modes. The mean of the X–Y coordinates of ice-covered cells,  $meanX_t$  and  $meanY_t$ , describes the center of sea ice distribution at time step  $t$ . The standard deviation,  $stdX_t$  and  $stdY_t$ , describes the amount of variation in the distribution along the X and Y axes at time step  $t$ . The correlation of X and Y,  $corXY_t$ , describes the shape of the distribution at time step  $t$ . Figure 3 illustrates different distributions with different statistical values.

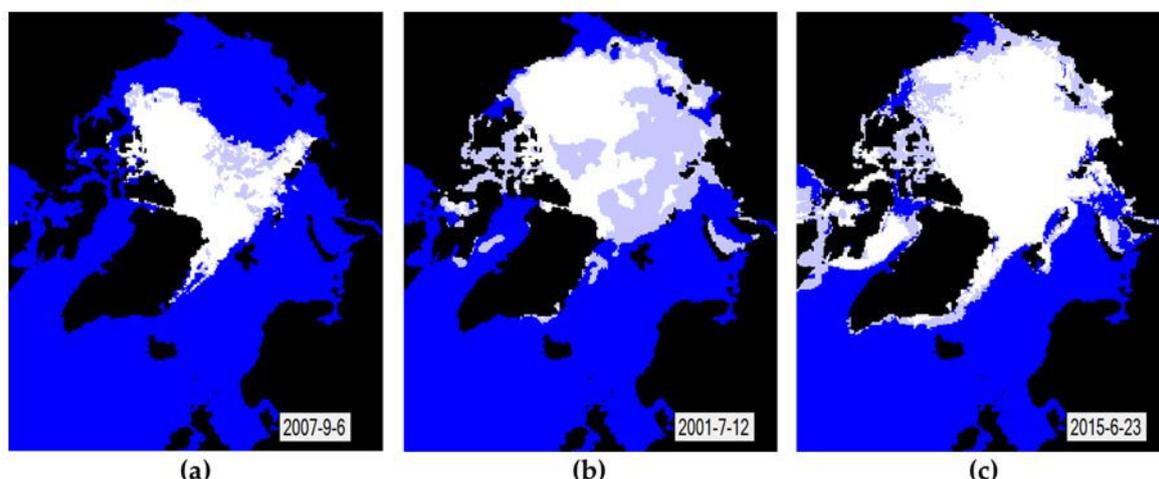


Figure 3. Distributions with different sea ice extent ( $SIE$ ), mean, standard deviation ( $std$ ), and correlation values ( $cor$ ): (a) Sea ice distribution with  $SIE = 3914375$ ,  $meanX_t = 123.4$ ,  $meanY_t = 174.3$ ,  $stdX_t = 27.1$ ,  $stdY_t = 26.5$ ,  $corXY_t = -0.35$ ; (b) Sea ice distribution with  $SIE_t = 9048750$ ,  $meanX_t = 127.4$ ,  $meanY_t = 187.9$ ,  $stdX_t = 40.4$ ,  $stdY_t = 36.8$ ,  $corXY_t = 0$ ; (c) Sea ice distribution with  $SIE_t = 9706875$ ,  $meanX_t = 117.6$ ,  $meanY_t = 181.6$ ,  $stdX_t = 46.9$ ,  $stdY_t = 41.1$ ,  $corXY_t = 0.28$ .

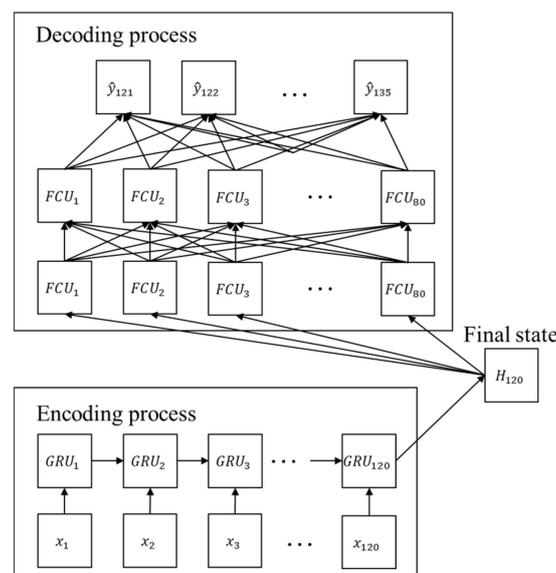
The input values are normalized between 0 and 1, using the minimum and maximum values, to ensure that each factor has an equal contribution. Table 1 describes the values and Equation (1) defines the normalization, where  $x_i$  is  $i$ -th input factor and  $min_i$  and  $max_i$  are the minimum and maximum values of the factor. The maximum value of time step  $t$ , which is 19 May 2027, was determined to consider the future use of the ANN. The minimum and maximum values of the other factors were determined using the features of the given data, which include geographical data and 17 years of previous observations, from 1 January 2000 to 31 December 2016.

$$f(x_i) = \frac{x_i - min_i}{max_i - min_i} \tag{1}$$

**Table 1.** Minimum and maximum value of input factors.

	$t$	$X$	$Y$	$depth$	$SIE_t$	$meanX_t$	$meanY_t$	$stdX_t$	$stdY_t$	$corXY_t$	$SIC_t$
Min.	0	0	0	-4500	0	108.6	164.0	12.0	13.6	-0.5	0
Max	10,000	233	274	0	$1.5 \times 10^7$	139.8	201.4	59.5	54.6	0.4	1

The prediction comprises encoding and decoding processes. We use a gated recurrent unit (GRU) for encoding. The GRU is a recursive neural network that has often been used with sequential data, such as natural language processing. It showed a similar performance to that of a LSTM, but has a simpler architecture that allows for efficient training of the network [22]. Similar to a LSTM, the GRU has a hidden state, which is a set of variables, and is updated gradually using a new input at each time step. We used 30 real variables as the hidden state. In the final hidden state, the encoded sequential input data are used as an input in the decoding process, in which a feed-forward neural network determines the future SIC. This comprises two hidden layers and 60 fully connected units in each layer. These numbers, which define the network architecture, are determined based on trial experiments. Figure 4 illustrates the encoding and decoding processes. The subscripts of  $x$  and  $\hat{y}$  indicate a time step.



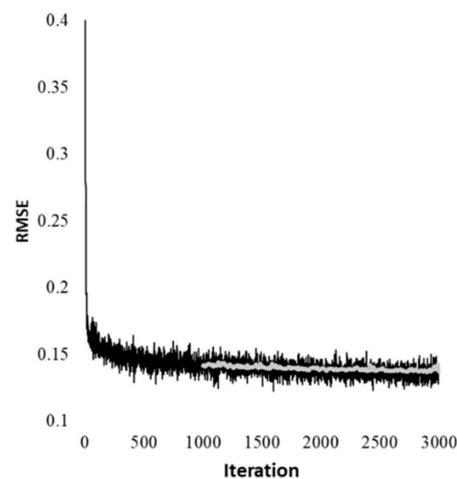
**Figure 4.** Illustration of the encoding and decoding processes used to make the prediction.

We used 80,000 samples for training and 20,000 samples for validation using SIC observation data from 1 January 2000 to 31 December 2016. Because the ANN predicts the future SIC of each cell individually, 20,661 samples can be created from a timeframe. In creating a sample, we first determined

a time period of 135 days within the above 17-year time frame. The data from the first 120 days were used to create an input and the data from the last 15 days were used as reference data for computing prediction errors. The error is defined by the root-mean-square error (RMSE) and is minimized during the training process. We used a mini-batch set for efficient convergence, in which the batch size was 2000 samples and the training samples were randomly shuffled when all the samples were used. Equation (2) defines the objective function of the training, in which set  $S$  is samples in a mini-batch set and set  $T$  is the prediction time steps, which are 15 days.  $|S|$  and  $|T|$  indicate the size of sets  $S$  and  $T$ , respectively;  $y_{st}$  is an observation value of sample  $s$  at time step  $t$ ;  $\hat{y}_{st}$  is a prediction value of sample  $s$  at time step  $t$ .

$$\text{Minimize RMSE} = \sqrt{\sum_{s \in S} \sum_{t \in T} \frac{(y_{st} - \hat{y}_{st})^2}{|S||T|}} \quad (2)$$

The training process was terminated after 3000 iterations. We used the following two RMSEs: (i) a training RMSE; and (ii) a validation RMSE. These RMSEs were computed using the training and validation samples, respectively. The former was used to compute gradients at each training step and the latter was used to validate the network model. The validation started at 1000 iterations and the model was kept if the validation RMSE was better than the known best value. Figure 5 illustrates the convergence of the RMSEs in the network model.



**Figure 5.** Convergence of the root-mean-square errors (RMSEs) during the training process. The black line indicates the training RMSE and the grey line indicates the validation RMSE.

### 3. Results and Discussion

We tested the model using 45 periods from 1 January 2017 to 6 November 2018, which were not used in the training process (Table 2).

**Table 2.** Test periods.

	Period 1	Period 2	Period 3	...	Period 45
Input period	3 Sep. 2016– 31 Dec. 2016	18 Sep. 2016– 15 Jan. 2017	3 Oct. 2016– 30 Jan. 2017	...	25 June 2018– 22 Oct. 2018
Prediction period	1 Jan. 2017– 15 Jan. 2017	16 Jan. 2017– 30 Jan. 2017	31 Jan. 2017– 14 Feb. 2017	...	23 Oct. 2018– 06 Nov. 2018

To determine the effect of global SIC information on model performance, we performed controlled experiments in the test. We trained an additional model without using the global SIC information and compared its performance with that of the model presented in Section 2. We called the above control model ‘GRU 1’ and the latter experimental model ‘GRU 2’. Moreover, we compared the GRU models

with LSTM models. Similarly, the LSTM models were trained using the different data schemes, and we called the control model ‘LSTM 1’ and the experimental model ‘LSTM 2’. Table 3 compares the data schemes in GRU 1, LSTM 1, GRU 2, and LSTM 2.

**Table 3.** Data schemes in GRU 1, GRU 2, LSTM 1, and LSTM 2.

Model	$t$	$X$	$Y$	$depth$	$SIE_t$	$meanX_t$	$meanY_t$	$stdX_t$	$stdY_t$	$corXY_t$	$SIC_t$
GRU 1	Yes	Yes	Yes	Yes	No	No	No	No	No	No	Yes
LSTM 1	Yes	Yes	Yes	Yes	No	No	No	No	No	No	Yes
GRU 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LSTM 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

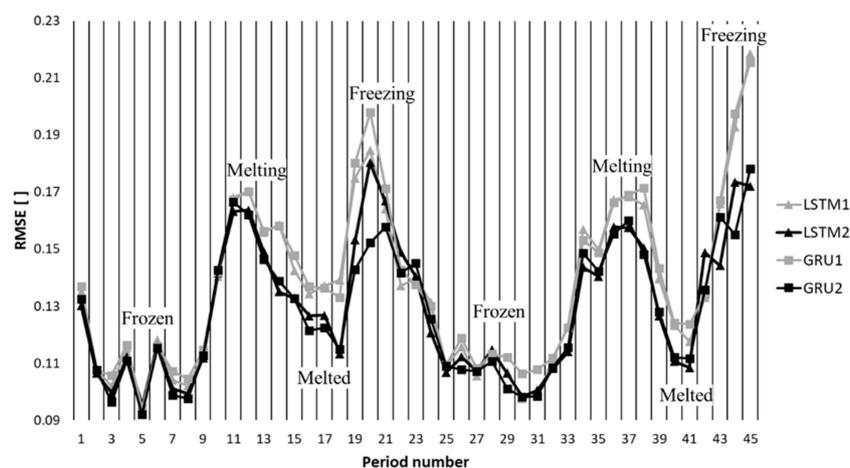
We used three types of errors in the test: (i) RMSE; (ii) mean absolute error in SIC (MAESIC); and (iii) mean absolute error in SIE (MAESIE). Equations (3)–(5) define the three errors, respectively, where set  $C$  is the dynamic cells and set  $T$  is the prediction time steps.  $|C|$  and  $|T|$  are the size of the sets,  $y_{ct}$  is an observed SIC in cell  $c$  at time step  $t$ , and  $\hat{y}_{ct}$  is a predicted SIC in cell  $c$  at time step  $t$ . Similarly,  $z_{ct}$  is an observed SIE in cell  $c$  at time step  $t$ , and  $\hat{z}_{ct}$  is a predicted SIE in cell  $c$  at time step  $t$ .

$$RMSE = \sqrt{\sum_{c \in C} \sum_{t \in T} \frac{(y_{ct} - \hat{y}_{ct})^2}{|C||T|}} \quad (3)$$

$$MAESIC = \sum_{c \in C} \sum_{t \in T} \frac{|y_{ct} - \hat{y}_{ct}|}{|C||T|} \quad (4)$$

$$MAESIE = \sum_{c \in C} \sum_{t \in T} \frac{|z_{ct} - \hat{z}_{ct}|}{|T|} \quad (5)$$

Figures 6–8 describe the test results. The horizontal axis indicates the period number and the vertical axis indicates the errors defined in Equations (3)–(5). In each figure, there are four lines defined by two marker types and two line colors. The lines with triangles are the errors of the LSTM models and the lines with the squares are the errors of the GRU models. The line colors are used to represent the data schemes, in which if the model used both global and local information, the line color is black. Otherwise, the line color is grey.



**Figure 6.** RMSE in the test periods.

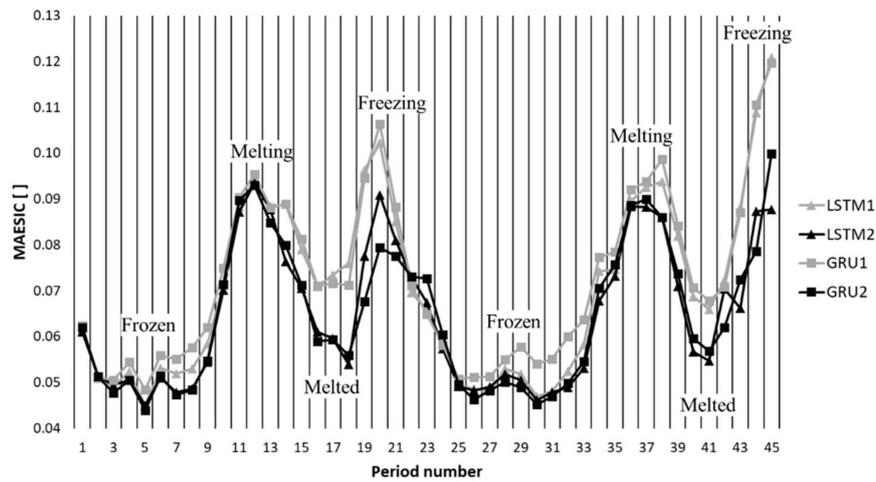


Figure 7. Mean absolute error in sea ice concentration (MAESIC) in the test periods.

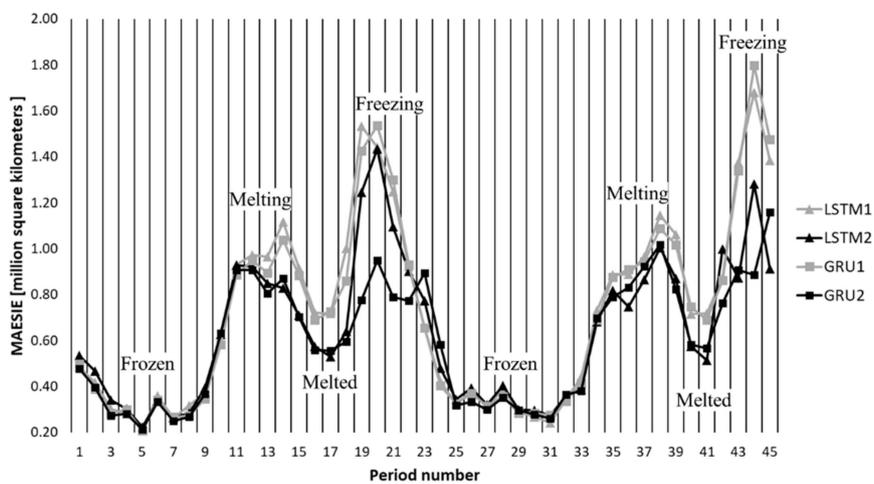


Figure 8. Mean absolute error in sea ice extent (MAESIE) in the test periods.

Table 4 describes the test results in a quantitative manner. In both GRU and LSTM models, the use of global information improved prediction accuracy. On average over the periods, the improved accuracy of the GRU 2 was 0.01 in RMSE, 0.01 in MAESIC, and 0.11 million km<sup>2</sup> in MAESIE. Similarly, the improved accuracy of LSTM 2 was 0.01 in RMSE, 0.01 in MAESIC, and 0.08 million km<sup>2</sup> in MAESIE. These numbers increased to 0.02, 0.02, and 0.51 million km<sup>2</sup>, respectively, in the GRU models and 0.02, 0.01, and 0.26 km<sup>2</sup>, respectively, in the LSTM models during the freezing seasons, which are periods 19–22 and 43–45.

Table 4. Test results.

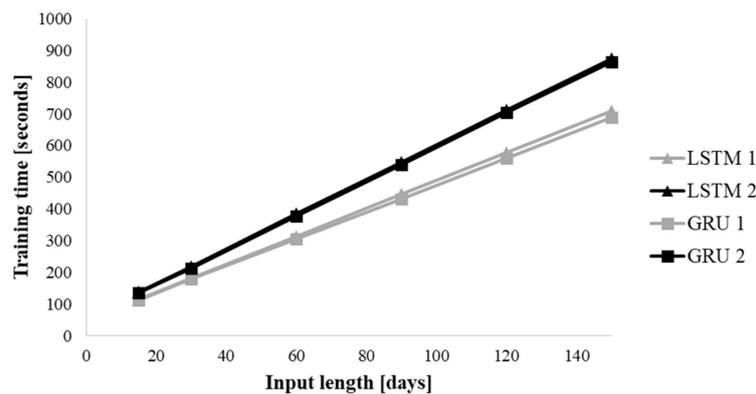
Period	Model	RMSE	MAESIC	MAESIE
1–45	GRU 1	0.14	0.07	0.71 × 10 <sup>6</sup> km <sup>2</sup>
	GRU 2	0.13	0.06	0.60 × 10 <sup>6</sup> km <sup>2</sup>
	LSTM 1	0.14	0.07	0.72 × 10 <sup>6</sup> km <sup>2</sup>
	LSTM 2	0.13	0.06	0.64 × 10 <sup>6</sup> km <sup>2</sup>
19–22, 43–45	GRU 1	0.18	0.09	1.40 × 10 <sup>6</sup> km <sup>2</sup>
	GRU 2	0.16	0.07	0.89 × 10 <sup>6</sup> km <sup>2</sup>
	LSTM 1	0.18	0.09	1.37 × 10 <sup>6</sup> km <sup>2</sup>
	LSTM 2	0.16	0.08	1.11 × 10 <sup>6</sup> km <sup>2</sup>

For statistical comparison, we conducted *t*-tests to determine significant differences between the different data schemes and model types (Table 5). In the comparison between the data schemes, all *p*-values were 0, which indicates that there is significance in the selection of the data scheme. In the comparison between the model types, it is difficult to conclude that there is significance in the selection of model type because some of the *p*-values were over 0.05.

**Table 5.** *T*-test results of comparisons between the models.

Model Pair	<i>p</i> -value		
	RMSE	MAESIC	MAESIE
GRU 1–GRU 2	0	0	0
LSTM 1–LSTM 2	0	0	0
GRU 1–LSTM 1	0.02	0	0.13
GRU 2–LSTM 2	0.07	0.45	0.02

All the models were trained using an Intel Core i7-8700K (3.70 GHz) with 16-GB memory and NVIDIA GeForce GTX 1080 (2560 CUDA cores and 8- GB memory). Figure 9 presents the training times of each model using different input lengths. Although the GRU is known to involve faster training, there was no significant difference in this experiment. If the model type is same, using global SIC information increases the training time. Moreover, there is a clear linear relationship between training time and input length.



**Figure 9.** Training time of the models.

Moreover, we conducted visualization of the evolving SIC of the GRU models. Figure 10, for example, illustrates the development of SIC in periods 20 and 45. The first-row images show the observations and the second- and third-row images present the predictions of GRU 1 and GRU 2, respectively. In each image, the white cells indicate the area where the SIC is more than 0.15. Figure 11 illustrates the prediction errors. The first- and second-row images show the errors of GRU 1 and GRU 2, respectively. Although the differences in RMSE and MAESIC between GRU 1 and GRU 2 seem to be not significant in Table 4, the visual comparison shows more apparent differences, with the prediction of GRU 2 being remarkably better than that of GRU 1. This discrepancy between the quantitative and visual comparisons may be due to the central area SIC. Both GRU 1 and GRU 2 can predict the SIC accurately because the SIC is relatively static. One interesting observation is how GRU 2 predicts future SIC in uncertain areas. Unlike existing numerical models, the model fills in the areas with ambiguous values, such as the value between 0.2 and 0.6, to minimize the errors.

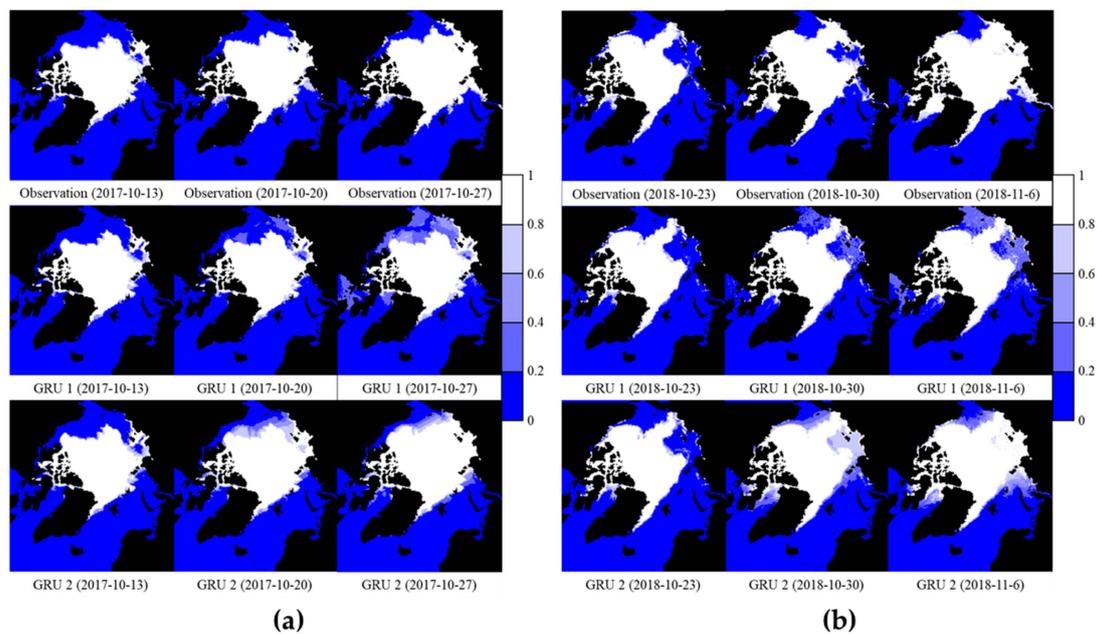


Figure 10. Development of SIC: (a) Period 20; (b) Period 45.

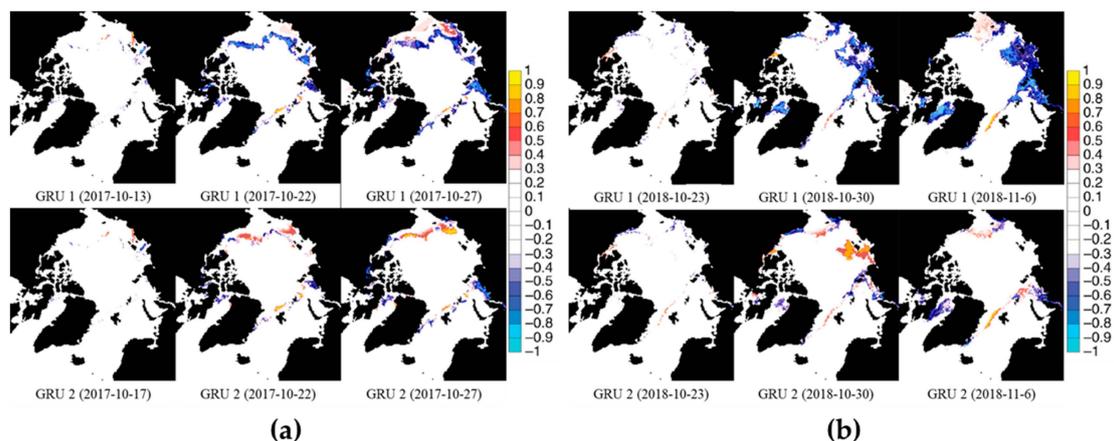


Figure 11. SIC deviation: (a) Period 20; (b) Period 45.

#### 4. Conclusions

In this paper, we presented an ANN for the short-term prediction of Arctic SIC. The ANN determines the future SIC using encoding and decoding processes, in which a GRU encodes sequential input data as a hidden state, and a feed-forward neural network determines future SIC using the hidden state. The ANN predicts each cell individually because of the large volume of Arctic sea ice data. This results in low accuracy, as each prediction is made using only local sea ice information. To overcome this, the ANN was trained using global SIC information that includes statistics describing SIC distribution at each input time step. We compared the performance of the ANN with that of other ANNs that use only local SIC information or a LSTM network as the encoder. The results of the *t*-tests showed that the selection of the data scheme has a significant impact on performance, but was difficult to determine whether the encoder choice between the GRU or the LSTM network has a significant impact.

The selection of training, validation, and test samples for an ANN affects the model's performance significantly. In this paper, we created the samples in a random manner. However, this may not be the best method when using Arctic sea ice data because of the different characteristics of cells according

to the regions and seasons. For example, we are more interested in the boundary area of the Arctic than in the central area because of the different changes in SIC. The SIC of the boundary area is more dynamic and uncertain than that of the central area. Thus, most prediction errors arise in the boundary area. However, it is difficult to define the location of the boundary area because it varies over time. For example, the boundary area in a summer season is located relatively closer to the Arctic center than in a winter season. The annual decline in Arctic sea ice is also related to the location of the boundary area. Such different characteristics of cells should be considered in the sampling process. We believe that a better selection of the samples can improve prediction accuracy.

We used a GRU for encoding and a feed-forward neural network for decoding. However, there are other options that can be chosen when selecting the network architecture. For example, we can use GRUs in both encoding and decoding processes. This allows the decoding to use the prediction values from the previous steps in the next-step prediction as well as the final hidden state. Moreover, although we determined the architectural parameters (e.g., number of hidden state variables in the encoder, and number of layers and neurons in the decoder) based on several trials, the parameters could be further optimized, which may improve the model's performance.

Researchers have shown atmospheric and oceanic factors that are correlated with the Arctic SIC [17]. We believe that using such correlated factors as well as the SIC could improve the results of our approach.

**Author Contributions:** Conceptualization, methodology, validation, visualization and original draft writing by M.C.; Data curation and review & editing writing by L.W.A.D.S.; Supervision by H.Y.

**Funding:** This research was supported by the Arctic Challenge for Sustainability Research Project (ArCS) by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Schwarz, J. Can the Northern Sea Route be Profitable? In *Northern Sea Route, Future and Perspective, Proceedings of the INSROP Symposium Tokyo, '95, Tokyo, 1–6 October 1995*; Kitagawa, H., Ed.; Ship and Ocean Foundation: Tokyo, Japan, 1996; pp. 605–614.
2. De Silva, L.W.A.; Yamaguchi, H.; Ono, J. Ice-ocean Coupled Computations for Sea-ice Prediction to Support Ice Navigation in Arctic Sea Routes. *Polar Res.* **2015**, *34*, 18. [[CrossRef](#)]
3. Johns, T.C.; Durman, C.F.; Banks, H.T.; Roberts, M.J.; McLaren, A.J.; Ridley, J.K.; Searl, Y. The New Hadley Centre Climate Model (HadGEM1): Evaluation of Coupled Simulations. *J. Climate.* **2006**, *19*, 1327–1353. [[CrossRef](#)]
4. Kay, J.E.; Deser, C.; Phillips, A.; Mai, A.; Hannay, C.; Strand, G.; Arblaster, J.M.; Bates, S.C.; Danabasoglu, G.; Edwards, J. The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bull. Amer. Meteor. Soc.* **2015**, *96*, 1333–1349. [[CrossRef](#)]
5. Watanabe, M.; Suzuki, T.; O'ishi, R.; Komuro, Y.; Watanabe, S.; Emori, S.; Takemura, T.; Chikira, M.; Ogura, T.; Sekiguchi, M.; et al. Improved Climate Simulation by MIROC5: Mean States, Variability, and Climate Sensitivity. *J. Climate.* **2010**, *23*, 6312–6335. [[CrossRef](#)]
6. Kimura, N.; Nishimura, A.; Tanaka, Y.; Yamaguchi, H. Influence of Winter Sea-ice Motion on Summer Ice Cover in the Arctic. *Polar Res.* **2013**, *32*, 20193. [[CrossRef](#)]
7. Wang, W.; Chen, M.; Kumar, A.; Wang, W.; Chen, M.; Kumar, A. Seasonal Prediction of Arctic Sea Ice Extent from a Coupled Dynamical Forecast System. *Month. Weather Rev.* **2013**, *141*, 1375–1394. [[CrossRef](#)]
8. Zhang, J.; Steele, M.; Lindsay, R.; Schweiger, A.; Morison, J. Ensemble 1-Year Predictions of Arctic Sea Ice for the Spring and Summer of 2008. *Geophys. Res. Lett.* **2008**, *35*, L0850. [[CrossRef](#)]
9. Hebert, D.A.; Allard, R.A.; Metzger, E.J.; Posey, P.G.; Preller, R.H.; Wallcraft, A.J.; Phelps, M.W.; Smedstad, O.M. Short-term Sea Ice Forecasting: An Assessment of Ice Concentration and Ice Drift Forecasts using the U.S. Navy's Arctic Cap Nowcast/Forecast System. *J. Geophys. Res. Oceans* **2015**, *120*, 8327–8345. [[CrossRef](#)]

10. Sakov, P.; Counillon, F.; Bertino, L.; Lisæter, K.A.; Oke, P.R.; Korabiev, A. TOPAZ4: An Ocean-sea Ice Data Assimilation System for the North Atlantic and Arctic. *Ocean Sci.* **2012**, *8*, 633–656. [[CrossRef](#)]
11. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems, Stateline, NV, USA, 3–8 December 2012; pp. 1097–1105.
12. Dahl, G.E.; Yu, D.; Deng, L.; Acero, A. Context-dependent Pre-trained Deep Neural Networks for Large-vocabulary Speech Recognition. *IEEE Transact. Audio Speech Lang Process.* **2012**, *20*, 30–42. [[CrossRef](#)]
13. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Aitken, A.P.; Tejani, A.; Totz, J.; Wang, Z.; Shi, W. Photo-realistic Single Image Super-resolution using a Generative Adversarial Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
14. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint* **2018**, arXiv:1810.04805.
15. Chi, J.; Kim, H.C. Prediction of Arctic Sea Ice Concentration Using a Fully Data Driven Deep Neural Network. *Remote Sens.* **2017**, *9*, 1305.
16. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comp.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
17. Kim, J.; Kim, K.; Cho, J.; Kang, Y.; Yoon, H.J.; Lee, Y.W. Satellite-based Prediction of Arctic Sea Ice Concentration Using a Deep Neural Network with Multi-model Ensemble. *Remote Sens.* **2019**, *11*, 19. [[CrossRef](#)]
18. Hoeting, J.A.; Madigan, D.; Raftery, A.E.; Volinsky, C.T. Bayesian Model Averaging: A Tutorial. *Stat. Sci.* **1999**, *14*, 382–401.
19. Cavalieri, D.; Parkinson, C.; Gloersen, P.; Zwally, H.J. *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: Boulder, CO, USA, 1996. [[CrossRef](#)]
20. Cavalieri, D.J.; Markus, T.; Comiso, J.C. *AMSR-E/Aqua Daily L3 12.5 km Brightness Temperature, Sea Ice Concentration, & Snow Depth Polar Grids, Version 3*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2014. [[CrossRef](#)]
21. Hori, M.; Yabuki, H.; Sugimura, T.; Terui, T. AMSR2 Level 3 product of Daily Polar Brightness Temperatures and Product, Version 1.00, Japan, Arctic Data archive System (ADS), 2012. Available online: [https://ads.nipr.ac.jp/portal/kiwa/Summary.action?owner\\_site=ADS&selectFile=A20170123-003&version=1.00&scr=list\\_home](https://ads.nipr.ac.jp/portal/kiwa/Summary.action?owner_site=ADS&selectFile=A20170123-003&version=1.00&scr=list_home) (accessed on 19 April 2019).
22. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-decoder for Statistical Machine Translation. *arXiv preprint* **2014**, arXiv:1406.1078.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).