https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023


Geoscientific Model Development Discussions

# IceTFT v 1.0.0: Interpretable Long-Term Prediction of Arctic Sea Ice Extent with Deep Learning

Bin Mu [1,*] , Xiaodan Luo [1,*], Shijin Yuan [1], and Xi Liang [2]

[1]School of Software Engineering, Tongji University, Shanghai 201804, China
[2]Key Laboratory of Research on Marine Hazards Forecasting, National Marine Environmental Forecasting Center, Beijing, China
[*]These authors contributed equally to this work.

**Correspondence:** Shijin Yuan (yuanshijin2003@163.com)

**Abstract.**

Annual reductions in Arctic sea ice extent (SIE) due to global warming. According to International Panel on Climate Change (IPCC) climate model projections, the summer Arctic will be nearly sea ice free in the 50s of the 21st century, resulting in sea level rise and thus affecting human life. Therefore, it is important to predict SIE accurately. For the most current studies, the majority of deep learning-based SIE prediction models focus on single-step prediction, and they not only have short lead times but also have limited forecasting skills. In addition, these models often lack interpretability. In this study paper, we construct the Ice Temporal Fusion Transformer (IceTFT) model, which consists mainly of the variable selection network (VSN), the long short-term memory (LSTM) encoder, and multi-headed attention mechanism. Then we select 11 predictors for IceTFT model, including SIE, atmospheric, and ocean variables according to the physical mechanisms influencing sea ice development. And the VSN in IceTFT can automatically adjust the weights of predictors and filter spuriously correlated variables. We also evaluate the IceTFT model from the division of the training set, the slicing methods of input data, and the length of input. The IceTFT model directly generates 12-month SIE with average monthly prediction errors of less than 0.21 $10^6$ $km^2$. And it predicts the September SIE nine months in advance with prediction error of less than 0.1 $10^6$ $km^2$, which is superior to the models from Sea Ice Outlook (SIO). Furthermore, we analyze the sensitivity of the selected predictors to the SIE prediction. It verifies that the IceTFT model has some physical interpretability. And the variable sensitivities also provide some reference for understanding the mechanisms governing sea ice development and selecting the assimilation variables in dynamic models.

## 1 Introduction

Arctic sea ice is one of the vital components of the global climate system. Due to global warming, the pace of the Arctic temperature rise has accelerated. This phenomenon, known as Arctic amplification, has hastened the melting of Arctic sea ice, which may have a potential influence on weather patterns and the climate of the Northern Hemisphere (Liu et al., 2013; Cohen

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



et al., 2014). According to International Panel on Climate Change (IPCC) climate model projections, the summer Arctic will be nearly sea ice free in the 50s of the 21st century (Stroeve et al., 2012; Overland and Wang, 2013; Voosen, 2020), which will have a significant influence on global climate change. Therefore, it is important to predict the development of Arctic sea ice, and it can provide a strong reference for studying and predicting global climate change trends.

However, in recent years, the Arctic sea ice extent (SIE) has been observed to be decreasing rapidly. From 1979 to 2013, the average September SIE decreased by around 14% per decade, and by approximately 50% by 2020 (Johannessen et al., 2020; Ramsayer, 2020). It is the second lowest in the satellite record (1979–2020), according to data from the National Snow and Ice Data Center (NSIDC) Sea Ice Index V3 (Fetterer and Windnagel., 2017). The rapid melting has made it more difficult to make an accurate SIE prediction. In addition, the melting of Arctic sea ice is influenced by a variety of physical factors. Consequently, the accurate forecast for SIE has become a challenging focus.

Since 2008, the annual Sea Ice Outlook (SIO) has been provided by the Sea Ice Prediction Network (SIPN) for the global research community to share predictions and ideas about the September SIE. The participants are required to submit their forecasts in early June, July, and August. The medians for errors between the September SIE forecast and observed in the last three years are showed in Fig.1. Figure 1a (b, c) depicts the errors were submitted by contributions in June (July, August). SIPN represents the current forecast level and community knowledge of the state and evolution of Arctic sea ice on the sub-seasonal-to-seasonal (S2S) timescale Wei et al. (2021). For example, the average SIE in September 2020 is 4.001 $10^6$ $km^2$ which is 2.41 $10^6$ $km^2$ less than the 1981–2010 climatological average. From the 2020 SIO contributions, the predicted medians for June, July, and August are 4.33, 4.36, and 4.30 $10^6$ $km^2$ respectively (Meier et al., 2021). There is still a certain gap between these forecast and observed. The major methods of Arctic SIE prediction include heuristic, statistical, dynamic models, machine learning models and others according to SIO. And we found more submissions for statistical approaches and dynamical models, while fewer for other methods. It is noteworthy that the number of applications using machine learning to predict SIE has increased since 2021, and it was rare to find one or two contributions before that time.

Fig.1 also shows the forecast skills in SIO over 2019–2021. Surprisingly, forecast skill did not improve significantly as forecast lead time was reduced. These results are consistent with other study (Stroeve et al., 2014). The medians of statistical approaches and machine learning are relatively close. They are evenly matched in forecast skills, and they both have slightly higher skills than dynamical models. As for dynamic models, they are the common forecast systems with physical significance, but uncertainties exist for initial fields and model parameters. Some studies applied data assimilation with different methods to improve forecast skill. For example, the assimilation of observed SIC has positive impact on the global sea ice cover (Mathiot et al., 2012; Massonnet et al., 2013). Toyoda et al. (2016) used a three-dimensional variational method to assimilate multivariate data, which improve the global representation of both the ocean and sea ice fields. In addition, Fritzner et al. (2019) analyzed the impact of assimilation of sea ice thickness (SIT) and snow depth on coupled ocean–sea-ice model, and they found that the assimilation of SIT can improve the sea ice forecast skill strongly than that of snow. These attempts demonstrated that the dynamic models partly produce effective forecasts, but they are often no better than simple statistical forecasts at lead times of two months or longer(Wayand et al., 2019; Blanchard-Wrigglesworth et al., 2015). However, due to the complexity of

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



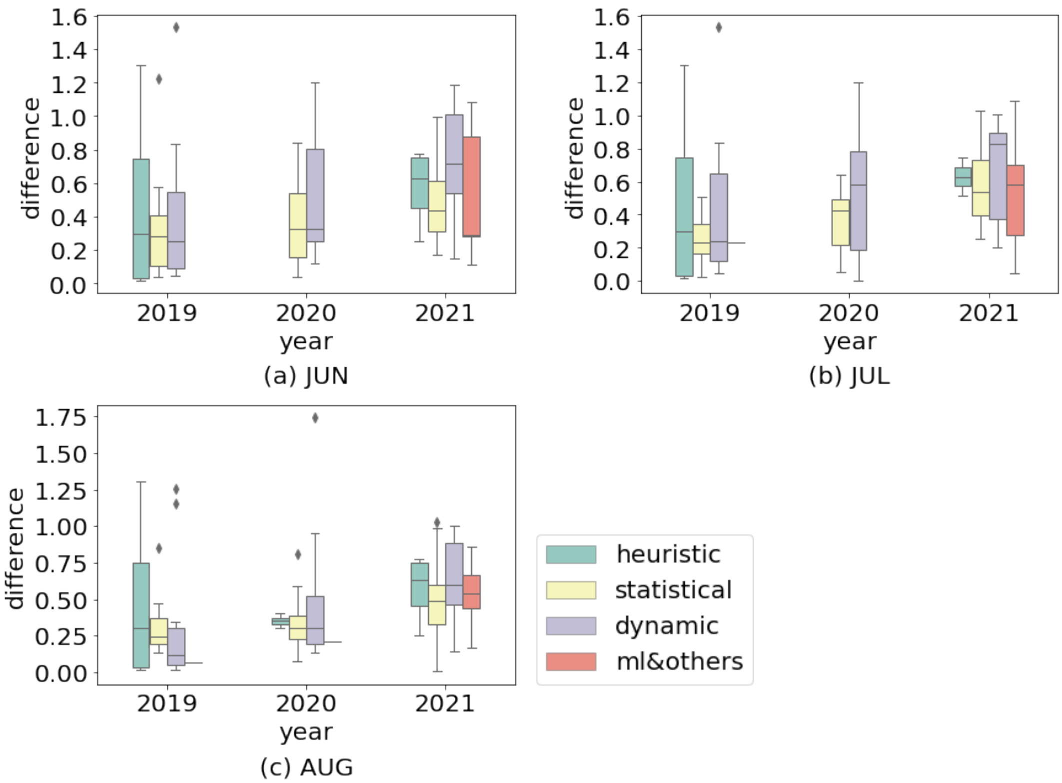


**Figure 1.** The medians for bias between September Arctic SIE observations and forecasts from June (a), July (b), and August (c) in 2019, 2020, and 2021, respectively.

Arctic sea ice melt mechanisms, statistical models cannot capture the non-linear relationships between variables, and they are dependent on the precision of the observed data, making it difficult to predict abrupt changes.

Up to now, deep learning (DL) methods have been progressively employed to predict Arctic sea ice. Such as Chi and Kim (2017) firstly applied long short-term memory (LSTM) to one-month Arctic SIC prediction model, and only used observed sea ice data by remote sensing sensors as an input of that. Then, they used a recursive approach to make the prediction model provide 12-month predictions. Kim et al. (2020) proposed a novel one-month sea ice concentration (SIC) prediction model with convolutional neural networks (CNNs) by incorporating SIC, atmospheric, and oceanic variables. Due to the CNN cannot capture the time series dependence, they trained 12 models to provide each month prediction of the year and confirmed the superiority of that. Study of Liu et al. (2021) used convolutional LSTM (ConvLSTM) model, which can learn space-time features to predict daily SIC. It demonstrated that ConvLSTM can provide more accurate prediction than CNN, but the study is still based on a single-step model. Andersson et al. (2021) proposed a new deep learning model named IceNet, which learning from climate simulations and sea ice observational data, it outperforms a leading physics-based model in seasonal forecast of Arctic sea ice, particularly for summer predictions. However, it is redundant for training several models to provide months of

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023


Geoscientific Model Development Discussions

predictions. Ren et al. (2022) proposed a purely data-driven model for daily SIC prediction, SICNet. They only used the SIC as an input to the model and used the iteration method to get a weekly SIC forecast. In these studies, they only focus on one-step models. To provide long-term predictions, they used recursive approach which can result in increasing errors, or trained more models, which can lead to a growing cost and time of calculation. Therefore, these studies make it very evident that there has been less research on long-term prediction than on short-term forecasting. That ignores the periodicity of SIE, which reaches its maximum in March and minimum in September(Kwok and Untersteiner, 2011). In addition, the explainability of the deep learning model has received minimal attention. Chi et al. (2021) used ConvLSTM with a new perceptual loss function to predict SIC, and they used various combinations of input variables to identify the essential variables to accurately predict. In this study, various variables were used as the input of the proposed model for different channels, which does not shed light on how the model utilizes the full channels of input. These channels data may have an incomprehensible impact. Although the model in Andersson et al. (2021) has been pre-trained on climate simulation, the impact of that on prediction is also unexplained. Compared to dynamic models, DL models are considered as “black box” due to lack of physical mechanisms.

To improve the long-term forecast skill for SIE and analyze the impact of different factors on SIE forecasting, we introduce a new SIE prediction model, IceTFT, which is an interpretable model based on the Temporal Fusion Transformer (TFT) (Lim et al., 2021). The IceTFT model directly predicts 12 months of averaged Arctic SIE through multi-horizon forecasting. The long-term forecasting skills of SIE can be improved by using static metadata, the LSTM encoder, and the multi-headed attention mechanism in IceTFT to learn temporal long-term dependencies. Due to the complexity of the physical mechanisms affecting Arctic sea ice development, we select several atmospheres, ocean variables as input to improve the forecast skill. And the design of the variable selection network (VSN) in the IceTFT model species adjusts the weights of the variables by calculating their contribution to the prediction, that increases the interpretability of the model and assists in analyzing the effects of different variables on SIE predictions.

In this study, we set up several controlled experiments for different training sets, slicing methods of input data, and the length of input to investigate their effect on SIE forecasting. And to evaluate the forecast skills for SIE minima in IceTFT model, we compare them with those of various models in SIO. In addition, a sensitivity analysis of the variables in the IceTFT model was carried out to evaluate the effects of different variables and identify the physical mechanisms affecting sea ice development.

The contributions of this paper are as follows: 1) The IceTFT model directly provides long-term prediction forecasts of SIE for up to 12 months, avoiding the accumulation of errors in iterative predictions and the waste of computational resources caused by training multiple models. 2) The IceTFT model has a long lead time. It can forecast the September SIE forecast 9 months in advance with a small error. 3) The IceTFT model can automatically filter out dummy correlation variables, avoiding interference in the forecast resulting from noise in the input data. 4) The IceTFT model is interpretable, breaking the tradition of using ’black box’ machine learning models for SIE forecasting. At the same time, sensitivity analysis can be used to investigate the contribution of different input variables to SIE forecasts and uncover the physical mechanisms of sea ice development.

The remaining part of the paper proceeds as follows: Sect. 2 introduces the proposed structure model, IceTFT. Sect. 3 deals with the atmospheric and oceanographic variables that we have selected. Sect. 4 presents the experiment designs and evaluation, and the experiment results and analyzes affecting SIE prediction are introduced in Sect. 5.

## 2 IceTFT model

The sea ice dataset is a time series with substantial periodicity, and time series forecast models analyze historical data to make forecasts for next time steps. DL has good performance in this area, however, previous research mostly used CNN which cannot capture the characteristics of time series to predict SIC. Despite the fact that ConvLSTM, a component of LSTM that may capture time-dependence, has been employed in a few studies, it is still inadequate for capturing long-term dependencies. Transformer model was originally proposed by Google team in 2017(Vaswani et al., 2017), which initiated the attention mechanism to replace the traditional recurrent neural network to extract sequence information. It adds sine and cosine functions of different frequencies as positional encoding to the normalized input sequence, and that makes the attention mechanism in Transformer model fully capture the time series dependence, and it has higher performance than other traditional recurrent neural networks. Based on Transformer model, Temporal Fusion Transformer (TFT) was proposed for multi-step forecasting (Lim et al., 2021). TFT not only uses a sequence-to-sequence layer to learn both short-term and long-term temporal relationships in local, but also uses multi-head attention block to capture long-term dependencies. Due to the pronounced periodicity of sea ice, which has one peak and one trough in a yearly cycle, these two peaks are usually critical to the forecast. Therefore, TFT, which can capture long-term temporal features, is suitable for sea ice prediction.

Compared with traditional multi-step forecasting methods based on recursive approach, TFT uses LSTM encoders to summarize past inputs and generate context vectors so that it can produce multiple step forecasts at each time step directly. In the current study, most researchers use recursive methods to achieve multi-step prediction based on single-step prediction, resulting in the accumulation of errors and an increase in prediction error as the prediction step increases. This design in TFT reduces the long-term forecast error, and provides a feasible method to improve the long-term forecast skill of sea ice. Moreover, sea ice melting can be affected by several physical factors, and the various mechanisms responsible for sea ice variability remain unexplained. Regardless of sea ice prediction or data assimilation, it is a challenge to choose which variables to improve forecast accuracy. However, the VSN in the TFT model can modify the weights of input variables during the training process and automatically complete the selection of variables. It avoids the manual selection of variables, and the variable selection of the model also makes the calculation of the model interpretable.

Therefore, we construct the SIE yearly forecast structure model named IceTFT, which is modified in the design of input and loss function based on TFT. The original design of TFT used known future data to support in the prediction of the primary time series data. To help the model learn the physical mechanics underlying the SIE, we modified this part to use atmospheric and oceanographic variables with the same moment as the SIE. In addition, TFT is relies on positional encoding to capture temporal features. When time series data is rolled into the model, it may lose temporal information of input data. To solve this problem, we set time static metadata to provide temporal features that help the model better capture the periodicity of sea ice during the training process. The IceTFT architecture of it is shown on Fig.2. Three types of dataset are the inputs of the IceTFT, and each type is selected by a separate variable selection network to filter unnecessary noises. Through the use of the GRN, the VSN calculates the weight of each variable of prediction contribution, allowing the model to focus on the most significant features rather than overfitting irrelevant features. And it can filter spuriously correlated variables to improve the accuracy of

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



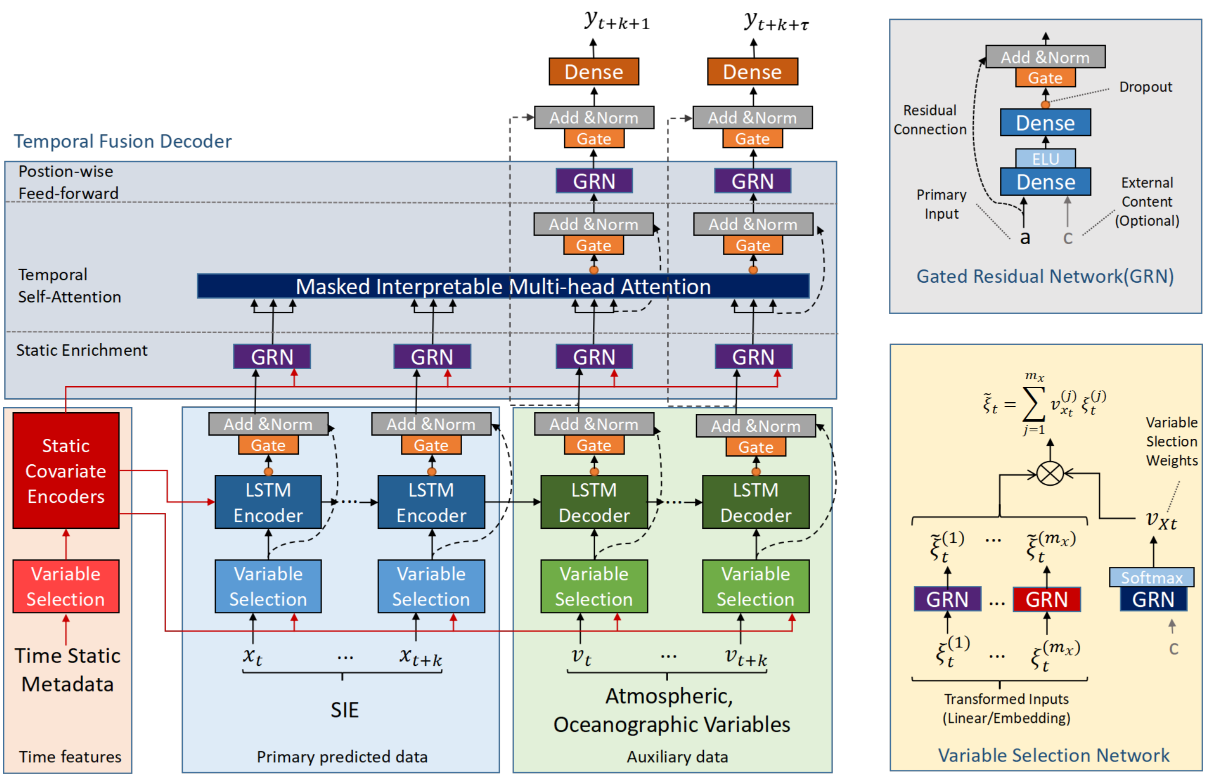


**Figure 2.** The IceTFT architecture is adapted on the basis of original TFT(Lim et al., 2021). The static time metadata, historical SIE data and other atmospheric and oceanographic variable are all inputs to the IceTFT. The Variable Selection Network (VSN) is used to select the most useful features. The Gated Residual Network (GRN) enables efficient information flow by skipping connections and gating layers.

SIE predictions. This facilitates the analysis of the physical mechanisms underlying sea ice development and make the IceTFT structure model more interpretable.

From Fig.2, the first input type is time static metadata, which is calculated by counting the days from the beginning of time. The original TFT is designed to use static covariate encorder to integrate static features from that, and use GRN to generate different context vectors that are linked to the different locations. In the IceTFT model, we apply this design to provide time information, so static covariate encorder condition temporal dynamics through these context vectors, and static enrichment layer enhance these temporal features. The second input is SIE which is the primary data for predicting in the IceTFT structure model. And the other inputs are several physical variables that are used to provide atmospheric, oceanographic features. IceTFT uses an LSTM encoder-decoder to enhance locality information of these time series. That is the advantage of capturing anomalies and cycling with them. In addition, the IceTFT uses an interpretable multi-head self-attentive mechanism to learn long-term features at different time steps. Each head can learn different temporal features and attend to a common set of input. Finally, in order to skip over extra features, the outputs are processed by GRN in position wise feed-forward layer. In this paper, we did not employ quantile forecast, which is the loss function in original TFT. Since SIE has decreased over the

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



past years, there is some mutagenicity in the summer. So the original design is not appropriate for SIE forecast, we used mean square error (MSE) as loss function to replace it.

## 3 Selected Predictors

As the subject of this research is the prediction of SIE, the historical data itself can provide data features for SIE prediction. SIE is defined as the total area covered by grid cells with SIC >15%, which is a common metric used in sea ice analysis. The first dataset of the model is the monthly SIE, which is provided from the NSIDC Sea Ice Index Version 3 (Fetterer and Windnagel., 2017). It contains daily and monthly SIE data in ASCII text files from November 1978 to the present. The area of this dataset is a region of the Arctic Ocean(39.23°-90°N, 180°W-180°E), and the monthly SIE is derived from the daily SIE for each month.

Since the development of Arctic sea ice is influenced by a variety of physical factors such as the atmosphere and the ocean, we choose a number of variables to aid the proposed model for SIE prediction in order to help it learn more physical mechanisms and improve forecasting skills. Numerous studies have analyzed the causal relationship between sea ice and physical variables, due to the fact that fluctuations in sea ice can be generated by different dynamical and thermodynamical processes and other factors. Huang et al. (2021) summarized recent studies and well-known atmospheric processes which connected with sea ice, and presented the causality graph as Fig.3.

From the study of Huang et al. (2021), the arrows b and c indicate that the increase of cloudiness and water vapor in the Arctic basin is due to local evaporation or enhanced water vapor transport, resulting in an increase in downward longwave radiation flux (DLWRF)(Luo et al., 2017). And DLWRF dominates surface warming and enhances sea ice melting in winter and spring(Kapsch et al., 2016, 2013). The melting of sea ice increases the air temperature, which in turn increases the DLWRF at the surface (Kapsch et al., 2013). At the same time, once the surface albedo is significantly reduced by sea ice melting, solar radiation may be absorbed by the ocean, which will further accelerate the sea ice melt in late spring and summer(Choi et al., 2014; Kapsch et al., 2016). Kapsch et al. (2016) studied the impact of realistic anomalies of DLWRF and downward shortwave radiation flux (DSWRF) on sea ice by applying a simplified forcing in a coupled climate model (the arrows e and f). Additionally, Liu and Liu (2012) conducted numerical experiments on the MITgcm model using reanalysis dataset to demonstrate that changes in surface air temperature and DLWRF have played a significant role in Arctic sea ice decline in recent years, and that changes in surface air specific humidity (SHUM) can regulate interannual variability in sea ice area. Therefore, in order to make our proposed model learn the atmospheric process, we select 2m air temperature (AT), DSWRF, DLWRF, SHUM these variables.

Moreover, the snow layer can regulate the growth rate of sea ice due to its highly insulating properties, and the accumulation of precipitation on the sea ice pack significantly influences the depth of the snow layer(Sturm and Matthew, 2002). And rain can melt, compact, and densify the snow layer, thus reducing the surface albedo and promoting sea ice melting(Perovich and D., 2002). The loss of snow on ice leads to a significant decrease in surface albedo over the Arctic Ocean, resulting in additional surface ice melt by absorbing more solar radiation(Screen and Simmonds, 2012). Higher precipitation and snowfall may lead to a thicker snowpack, which has implications for sea ice changing (Bintanja and Selten, 2014). Some researchers have studied

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



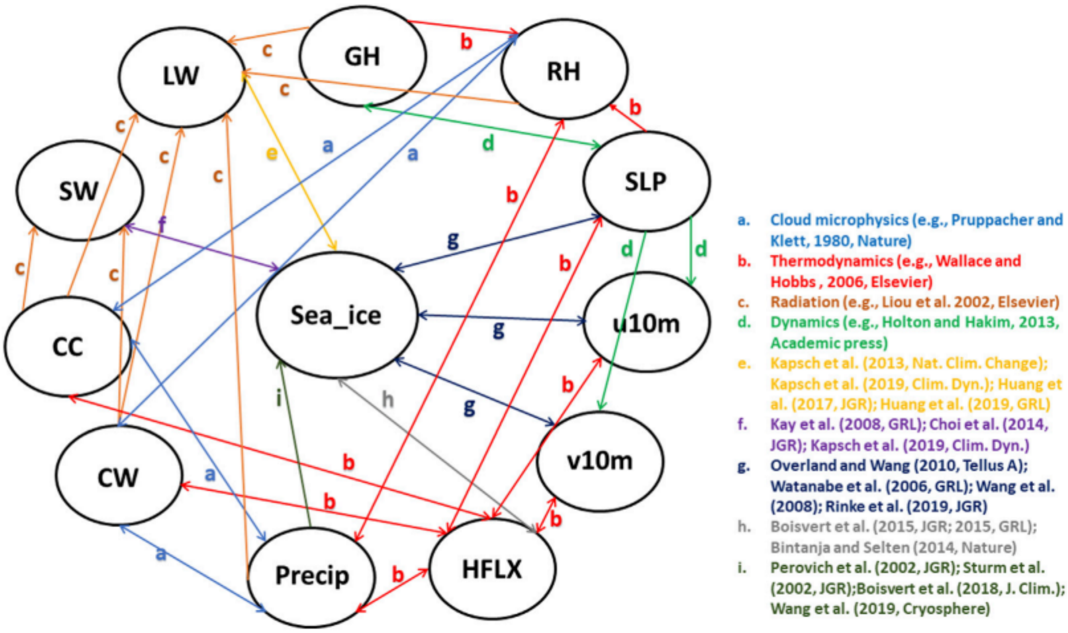


**Figure 3.** The causality graph , which is from the study(Huang et al., 2021) between important atmospheric variables and sea ice over the Arctic. Note that the processes a-d are well-known atmospheric processes, which may be outlined in multiple textbooks. The processes e-i are summarized from recent peer-reviewed publications, and they are ongoing research. The sea ice here represents sea ice coverage and / or sea ice thickness; GH is the geopotential height; RH is relative humidity; SLP means sea level pressure; u10m and v10m represent meridional and zonal wind at 10 m, respectively; HFLX is the sensible plus latent heat flux; Precip is the total precipitation; CW is the total cloud water path; CC is the total cloud cover; SW and LW represent net shortwave and longwave flux at the surface, respectively.

the correlation between river runoff and sea ice and found that river runoff has some influence on sea ice melting(He-Ping et al., 2000; Tong et al., 2014). Precipitation at high latitudes would also increase arctic river discharge, and the river may have the positive effect of maintaining thicker ice(Weatherly and Walsh, 1996)(the arrows i). Therefore, we also select precipitation (PRECIP), snow fall (SF) and river runoff (RUNOFF) to make the proposed model learn these processes.

In addition to atmospheric processes, to improve the interpretability of the model, we also make it learn some characteristics of ocean. Previous studies demonstrated the effect of sea surface temperature (SST) on Arctic sea ice. Bushuk et al. (2017) found that the SST provides the essential source of memory for melt-to-growth reemergence. Xi et al. (2019) supported that the additional assimilation of SST improves the predicted accuracy of SIE and SIT in the marginal sea ice zone. Therefore, we also selected SST variables to provide the oceanographic features for the model.

The mean values of each of the above eight variables over the global region were used as input data for the model. And we calculated the correlation coefficient between these and SIE. The results are provides in Table 1. The variables with the highest correlation coefficient with SIE, as shown in the table, are SST, AT, RUNOFF, and DLWRF, which are bolded. And these variables are all connected to surface evaporation and surface heat in the Arctic hydrological cycle. To make the model learn

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



more physical mechanisms, we selected clear sky downward longwave flux (CSDLF), clear sky downward solar flux (CSDSF) and upward solar radiation flux (USWRF) these radiative variables. The total of eleven physical variables are shown in Table 1.

**Table 1.** The specifications of eleven physical variables are used as the input of TFT model.

| Types | Variables | $R^2$ | Source | Unit |
|---|---|---|---|---|
| Atmospheric | 2m air temperature (AT) | **-0.7060** | NCEP-NCAR Reanalysis 1 | degK |
| | 2m specific humidity (SHUM) | -0.5965 | NCEP-NCAR Reanalysis 1 | grams/kg |
| | downward shortwave radiation flux (DSWRF) | 0.4080 | NCEP-NCAR Reanalysis 1 | $W/m^2$ |
| | downward longwave radiation flux (DLWRF) | **0.6638** | NCEP-NCAR Reanalysis 1 | $W/m^2$ |
| | clear sky downward solar flux (CSDSF) | 0.3684 | NCEP-NCAR Reanalysis 1 | $W/m^2$ |
| | clear sky downward longwave flux (CSDLF) | -0.6555 | NCEP-NCAR Reanalysis 1 | $W/m^2$ |
| | upward solar radiation flux (USWRF) | 0.4898 | NCEP-NCAR Reanalysis 1 | $W/m^2$ |
| Oceanographic | sea surface temperature (SST) | **-0.8756** | NOAA Optimum Interpolation SST V2 | degC |
| Other | precipitation (PRECIP) | -0.3809 | NCEP-NCAR Reanalysis 1 | $kg/m^2/s$ |
| | river runoff (RUNOFF) | **0.7590** | NCEP-NCAR Reanalysis 1 | $kg/m^2$ |
| | snow fall (SF) | 0.4452 | Boulder Monthly Means:Snowfall | inches |

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023

Geoscientific Model Development Discussions

## 4 Experiment Scheme

### 4.1 Settings of Comparison Experiment

Due to Arctic SIE decline has accelerated in recent years with the sparse dataset, many researchers (Chi and Kim, 2017; Kim et al., 2020; Chi et al., 2021) suggest that the recent time period have more useful features than early period for recent forecast, so they divided more data from the overall dataset to train. Therefore, we specified different periods for the training model. In addition, the slicing method of the data also has an impact on the results of the model, so we conducted comparison experiments for different slicing methods.

For traditional machine learning, datasets are often divided into training, valid, and testing data, and too many training sets may lead to overfitting. Consequently, we set January 1982 to December 2011 as training data, 2012 to 2015 as validation data, and others as testing data; this experiment is described as IceTFT-general. And 36 years of data (1982-2016) are used to train the IceTFT-2019 model for predicting SIE in 2019, and only from 2017 to 2018 are used for validation. Table 2 shows all the different periods of training, validation, and testing. In addition, considering the periodicity of SIE is too long, it is also a challenge for deep learning to capture the whole features. SIE always archives its maximum in March and minimum in September, and it consists of melting and freezing these two processes. Therefore, we halved forecast duration and set another experiment to compare the performance between 12-month forecast and 6-month forecast.

**Table 2.** The settings of different experiments

| Experiment Name | Training (Number) | Validation (Number) | Testing (Number) | Training Length | Output Length | Slicing Method |
|---|---|---|---|---|---|---|
| IceTFT-align | 1982.01~2012.12 (29) | 2012.01~2017.12 (4) | 2017.01~2021.12 (3) | 12 | 12 | align |
| IceTFT-general | 1982.01~2012.12 (349) | 2012.01~2017.12 (49) | 2017.01~2021.12 (37) | 12 | 12 | rolling |
| IceTFT-Pred2019 | 1982.01~2017.12 (409) | 2017.01~2018.12 (1) | 2018.01~2021.12 (25) | 12 | 12 | rolling |
| IceTFT-Pred2019(6) | 1982.01~2017.12 (409) | 2017.01~2018.12 (1) | 2018.01~2021.12 (25) | 12 | 6 | rolling |
| IceTFT-Pred2020 | 1982.01~2018.12 (421) | 2018.01~2019.12 (1) | 2019.01~2021.12 (13) | 12 | 12 | rolling |
| IceTFT-Pred2020(6) | 1982.01~2018.12 (421) | 2018.01~2019.12 (1) | 2019.01~2021.12 (13) | 12 | 6 | rolling |
| IceTFT-Pred2021 | 1982.01~2019.12 (432) | 2019.01~2020.12 (1) | 2020.01~2021.12 (1) | 12 | 12 | rolling |
| IceTFT-Pred2021(6) | 1982.01~2019.12 (432) | 2019.01~2020.12 (1) | 2020.01~2021.12 (1) | 12 | 6 | rolling |

Our experiments include both rolling and alignment slicing methods. The above experiments used a rolling method to generate input. Fig. 4 (a) shows the process of rolling. A slice of data consists of 12 time-step inputs and 12 time-step labels, and the whole length is 24. Using the rolling method to move the sliding window one time step, we can obtain next 24 time-step

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023

Geoscientific Model Development Discussions

slice data. While the IceTFT-align experiment uses the align method which is shown in Fig. 4 (b). Align inputs requires that the first time-step data is January in each slice of data. With rolling method, the model can only learn location information but lose temporal features due to the moving time series during training. So the design of static component is valuable for the IceTFT to condition temporal dynamics, and we set up a comparison experiment to validate the contribution of static information. These designs are also shown in Table 2.

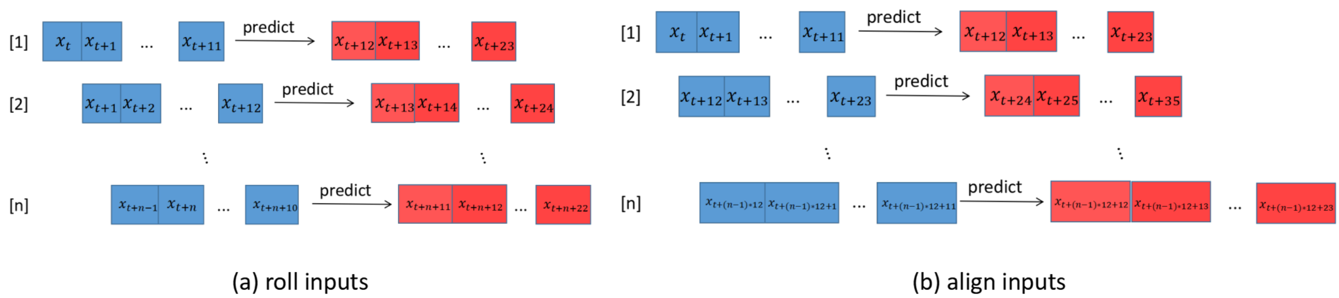


**Figure 4.** Two methods for generating inputs

### 4.2 Evaluation

There are three metrics used to evaluate the model performance: mean absolute error (MAE), root mean square error (RMSE) and root mean square deviation (RMSD), and the equations are as follows. Specially, RMSD can be used to further investigate the possible reasons for the discrepancy between the observed and predicted values of the SIE. In the formulas of metrics, the range of $n$ is from 1 to 12. Where $y$ and $\tilde{y}$ means the SIE observation and in prediction, and subscript $i$ represents $i$th month ordinal of a year. The RMSD is defined as the average distance between simulations and observations. It includes "bias" and "variance" these two components (Zheng et al., 2021). The first component is the mean bias of standard deviations, and the second can be viewed as the mean variation in the square of the difference between the standard deviations of the simulation and the observation. Where the R denotes the correlation coefficient between $y$ and $\tilde{y}$. For above all the metrics, the smaller value means the model has better performance.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \tilde{y}_i| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2} \tag{2}$$

$$\begin{aligned} bais &= (std(y) - std(\tilde{y}))^2 \\ variance &= \frac{1}{n}(y^2 + \tilde{y}^2 - 2y\tilde{y}R) \\ RMSD^2 &= bais + variance \end{aligned} \tag{3}$$

## 5 Results and Discussions

### 5.1 Impacts of Slicing Method of Inputs on Predictions

Table. 3 shows the prediction results for IceTFT model with different slicing methods. Compared to IceTFT-align and IceTFT-general, IceTFT-align run had a slight advantage in 2019 (IceTFT-align: RMSE of 34.04%, RMSD of 0.7129 ; IceTFT-general: RMSE of 34.87%, RMSD of 0.8206), but it had higher error than IceTFT-general run overall. This may be due to the fact that IceTFT-align run did not contain a sufficient number of samples for training. The model cannot learn enough features to predict. It demonstrates that the rolling method is effective in improving forecasting skills. And it is difficult to predict with high confidence for a model with too little training data.

**Table 3.** The three metrics (MAE, RMSE, RMSD) among the experiments with different slicing method on SIE prediction during 2019-2021.

| Year | Experiment Name | MAE | RMSE | RMSD |
|---|---|---|---|---|
| 2019 | IceTFT-align | 29.66% | 34.03% | 71.29% |
| | IceTFT-general | 26.87%(-2.97%) | 34.87%(+0.84%) | 82.06%(+10.77%) |
| 2020 | IceTFT-align | 43.30% | 49.15% | 90.40% |
| | IceTFT-general | 32.04%(-11.26%) | 41.62%(-7.53%) | 90.23%(-0.17%) |
| 2021 | IceTFT-align | 50.59% | 60.58% | 131.04% |
| | IceTFT-general | 44.95%(-5.64%) | 54.65%(-5.93%) | 121.25%(-9.79%) |

### 5.2 Impacts of Input Length on Predictions

In this study, the IceTFT model provides 12-month predictions. As the long-term prediction is a challenge for the model, we shortened the input length in half for a comparison experiment. The 6-month data as input of the IceTFT model, and the model generates the next 6-month predictions. Using 2019 as an example, the results of the monthly error are set out in Fig.5. Surprisingly, the errors of model with a short input length are rather larger than that with 12-month. Probably because the period of SIE is 12-month, it is more reasonable to set the time step to 12-step. Moreover, the 6-step time window is too

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023

Geoscientific Model Development Discussions

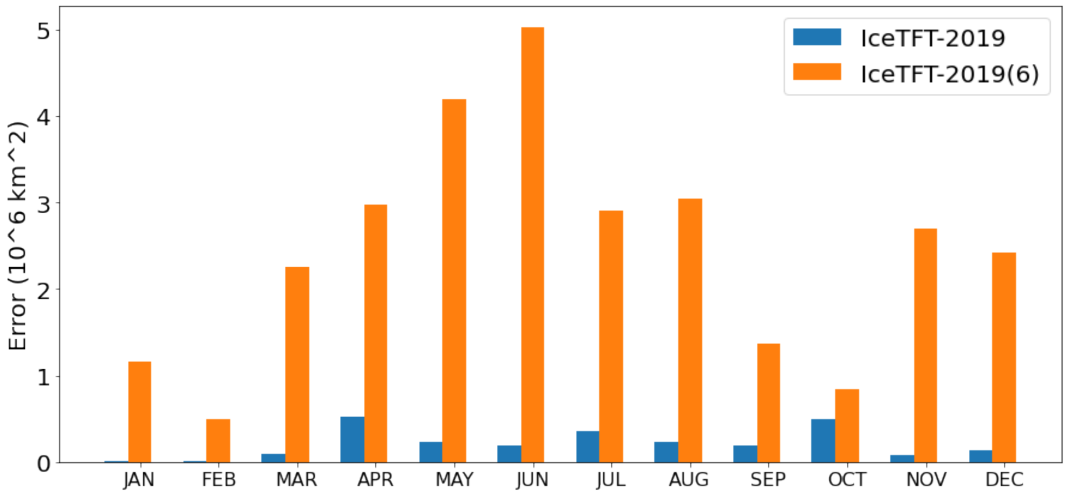


**Figure 5.** The error between SIE observations and predictions of the IceTFT with different input length in 2019

short to include both the March maximum and September minimum in each epoch. This may affect the model learning for the properties of extremes, hence increasing the inaccuracy of the extremes. Therefore, it is not the case that a shorter forecast time will result in a smaller predicted error, and the input length should be set according to the period of the data.

### 5.3 Impacts of Training Setting on Predictions

For IceTFT-Pred2019, IceTFT-Pred2020, and IceTFT-Pred2021 run, the training samples are more than IceTFT-general, and they have smaller predicted errors. It implied that adding more samples can improve the performance. From MAE and RMSE, the IceTFT-Pred2020 run has not shown as much improvement as IceTFT-Pred2019 and IceTFT-Pred2021 due to the anomalies. But it also has a significant reduction in RMSD. These results show that the training data that is close to prediction time is more useful than others that are far away. Due to the uncertainty in the model, we trained the model 20 times for each of these runs. Then we recorded the best predicted results and the mean predicted results. The mean predicted results represent the forecast skill of IceTFT model, while the best predicted results represent the performance of IceTFT model to capture the features of SIE. For each setting, the model can obtain the predicted results with low error through multiple training. Even though the mean predicted results have a slightly larger error than the best, the average predicted error of the model for each month is within 0.3 $10^6 km^2$. In 2019 and 2021, the difference between the best and mean prediction is not significant from the RMSE, which is not more than 0.04 $10^6 km^2$. Compared to the results of these two years, the errors of the mean predicted results increased in 2020. This is because there is a second record-low SIE in September 2020. Moreover, due to the predicted period is too long relatively, evaluating the forecast skill of the IceTFT model using MAE as the loss function is difficult. A low MAE does not mean that the model can predict all 12-step with low errors. The IceTFT model focuses on different physical factors during several training, and generates predictions with different trends. The model is hard to predict this minimum value accurately in each training, so the errors of mean prediction are much higher than best one.

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



**Table 4.** The three metrics (MAE, RMSE, RMSD) among the experiments with different training settings on SIE predictions during 2019-2021.

| Year | Experiment Name | MAE | RMSE | RMSD |
|---|---|---|---|---|
| 2019 | IceTFT-general | 26.87% | 34.87% | 82.06% |
| | IceTFT-Pred2019(mean) | 21.26%(-5.61%) | 26.68%(-8.19%) | 47.56%(-34.5%) |
| | IceTFT-Pred2019(best) | 16.49%(-10.38%) | 19.42%(-15.45%) | 45.54%(-36.52%) |
| 2020 | IceTFT-general | 32.04% | 41.62% | 90.23% |
| | IceTFT-Pred2019(mean) | 30.16%(-1.88%) | 38.08%(-3.54%) | 61.82%(-28.41%) |
| | IceTFT-Pred2020(mean) | 28.47%(-3.57%) | 37.47%(-4.15%) | 58.94%(-61.29%) |
| | IceTFT-Pred2020(best) | 20.07%(-11.97%) | 24.78%(-16.84%) | 48.90%(-41.33%) |
| 2021 | IceTFT-general | 44.95% | 54.65% | 121.25% |
| | IceTFT-Pred2019(mean) | 19.90%(-25.05%) | 24.75%(-29.9%) | 57.82%(-63.43%) |
| | IceTFT-Pred2020(mean) | 25.45%(-19.5%) | 33.45%(-21.2%) | 77.59%(-43.66%) |
| | IceTFT-Pred2021(mean) | 25.77%(-19.18%) | 30.18%(-24.47%) | 70.71%(-50.54%) |
| | IceTFT-Pred2021(best) | 16.84%(-28.11%) | 26.77%(-27.88%) | 66.89%(-54.36%) |

Compared the results between IceTFT-general run and IceTFT-Pred2019 run in 2019, IceTFT-Pred2019 run had lower error relatively. For the best run of IceTFT-Pred2019, the MAE reduces 10.38%, and RMSE and RMSD reduce 15.45% and 36.52% respectively. From the prediction results for 2020 and 2021, we can also get similar consequences. For example, IceTFT-Pred2020 run had more one year training data than IceTFT-Pred2019 run, and that caused the model to learn some features of recent time. Consequently, IceTFT-Pred2020 run has lower error than IceTFT-Pred2019 run for 2020 prediction, and IceTFT-Pred2021 run has lower error than TFT-Pred2020 run for 2021. Interestingly, IceTFT-Pred2019 run had higher accuracy for 2021 prediction. They both resulted in good performance with a small difference in three metrics. It may be that the trends of SIE between these two years are similar (We discussed the reason in Sect. 5.5). According to the results in Table. 4, RMSE is slightly higher than MAE for these experiments. RMSE is more susceptible to outlier influence than MAE. This illustrates that the model with optimal experiment settings produces 25% mean error monthly at most from MAE, but generate higher error in some months from RMSE.

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



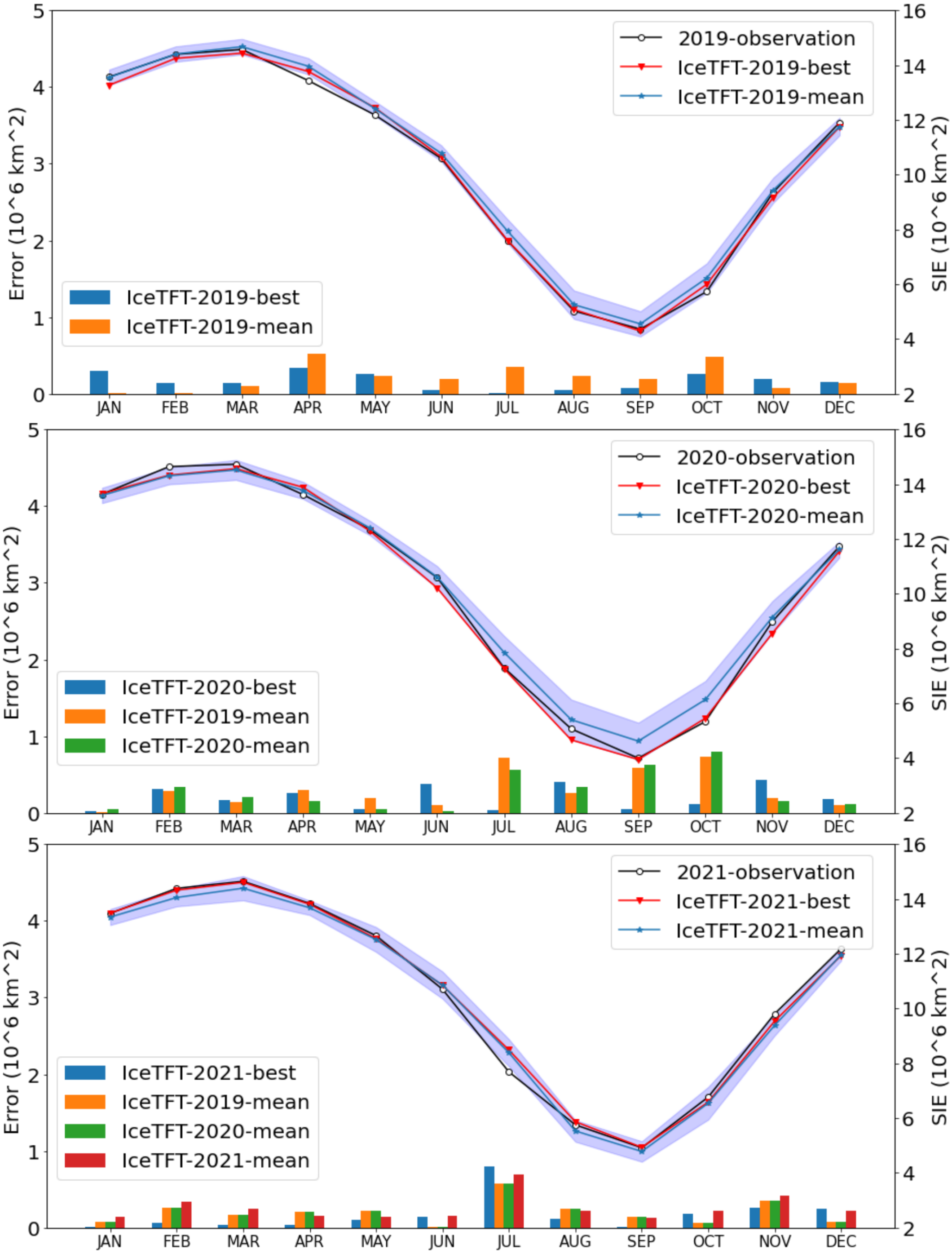


**Figure 6.** The SIE predictions, and the monthly error between observations and those with different experiment designs in 2019-2021

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023

Geoscientific Model Development
Open Access
Discussions

### 5.4 Performance of IceTFT for 12-month SIE predictions

In Fig.6, there is a clear trend of predictions for different years, and it also shows the monthly error between observations and predictions. As can be seen, the predictions of multiple training form a forecast period in which the vast majority of observations fall within the range. Except for September 2020, the mean predicted results have the same trend as the observations. In terms of the monthly error of the model with different settings, all the experiment runs had high errors in October or November. In addition, they had another high error in July, except for 2019. Due to global warming, it is a challenge to predict SIE in summer. In the melt seasons, which is from June to September, the SIE continued decline with steep slope. The line passing through the observed value of SIE in June and July has the steepest slope. It demonstrates that the SIE reduced significantly from June to July. Thus, it is difficult to predict downturn. And as a result, the July prediction is higher than observation with higher error. The SIE archive minimum in September, and sea ice becomes frozen after that time. Similarly, as temperature anomaly or other climate effect, the October or November prediction is on the high side. For 2021 predictions in Fig. 6 (c), the errors of IceTFT-Pred2019 run are smaller than IceTFT-Pred2021 run in the winter, but higher in the summer. Though the IceTFT-Pred2019 run has more accuracy than IceTFT-Pred2021 run from three metrics, it produces more error in September. As a results, the metric is not merely a performance benchmark for prediction. In addition, the monthly errors did not show a monotonically increasing trend, and it did not raise with time step increased. The model used direct predicted method to avoid the superimposition of errors in recursive approach, and it improved the accuracy of predictions. The same disadvantage exists for dynamic models, in that the predicted error increases with increasing forecast period. This issue was resolved by the IceTFT model, which generated longer-period predictions with smaller errors.

To further explore the potential causes for the inaccuracy between the SIE observations and predictions, the error between the detrended quarterly SIE observations and predictions over the period 2019-2021 are shown in Fig. 7. The RMSD ranges from 0.076 to 0.918 million km2 in Fig.7 (a), and the findings from the three years show a wide spread in RMSD on quarter. Figure 7 (b) displays a histogram of the temporal variation in the squared RMSD, consisting of the “bias” and “variance” following Eq. (4). Except for a few months, the magnitude of the bias is substantially larger than the variation in each year, indicating that the change in bias is the primary factor driving the increase in RMSD. The correlation coefficients in Fig.7 (c) also display an obvious reduction in the spring of 2020, which is consistent with the variations of variance in Fig.7 (b). This result indicates that the significant lower correlation coefficients are partially responsible for the RMSD peak. In addition, Fig.7(d) shows the standard deviations of the predictions of IceTFT model and observations, and the annual standard deviation represents the amplitude of the seasonal cycle of SIE. The results show that the difference between these two standard deviations obviously increases, which contributes to the sharp increase in bias during the same period. Thus, we can improve the predicted model by focusing on the seasonal variability in the predictions to reduce the RMSD.

### 5.5 Comparisons with SIO

In order to evaluate the performance of IceTFT model for SIE minimum prediction, we collected contributions that were provided to SIO over the past three years. Figure 8 presents the error between the observations and predictions of SIE minimum

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



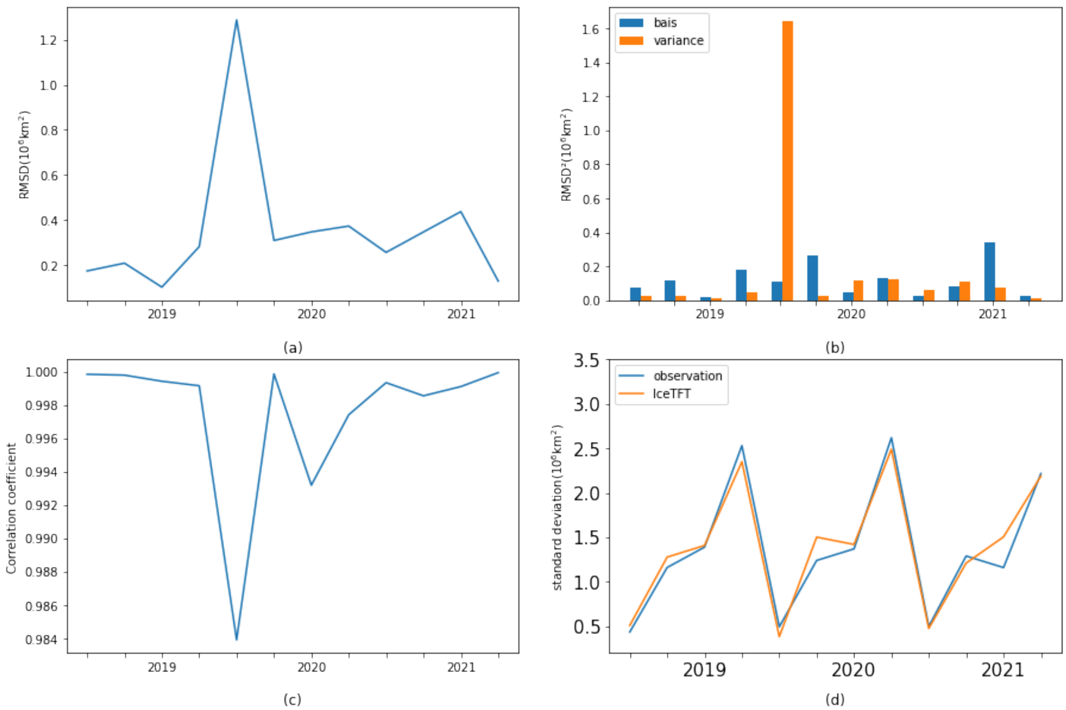


**Figure 7.** Time series of the RMSD between the detrended quarterly SIE from observations and predictions based on the IceTFT-model over the period 2019-2021: (a) RMSD; (b) squared RMSD (histogram), consisting of "bias" and "variance"; (c) correlation coefficient between predictions and observations; (d) standard deviation of predictions (orange line) and observations (blue line).

315 in different models. Compared to other models, the IceTFT model has the smallest error in prediction over the three years. For hindcasting experiments, machine learning always results in less error after several training. Therefore, as can be seen, the best results of IceTFT model has the smallest error, but the error increased a lot in mean results. The mean prediction indicates that the forecast skill of the model is relatively stable. In addition to 2020, the mean prediction of IceTFT model is superior to other models. In Sect. 5.2, we supported the mean prediction with a larger error for the anomalies in September to be meaningless. Because, in the mean prediction, in order to make small error for an anomalous minimum, the model must have a lower bound on its predictions during multiple training process. This is challenging to achieve as the model is limited by historical SIE data. Furthermore, the errors of all models are smaller relatively in 2019 (Green histogram in Fig.8). The SIE minimum observation reached its second-lowest value in 2020, and the anomalous caused the errors increased (Orange histogram are longer than blue). While the extremely low anomalies continue to influence the 2021 predictions, the predicted error of most models has increased (see the blue histogram in Fig.8). Even the predicted error is greater than that of 2020. However, the IceTFT model is not influenced by anomalies from the previous year, and only focuses on the physical factors influencing the development of sea ice in that year.

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023




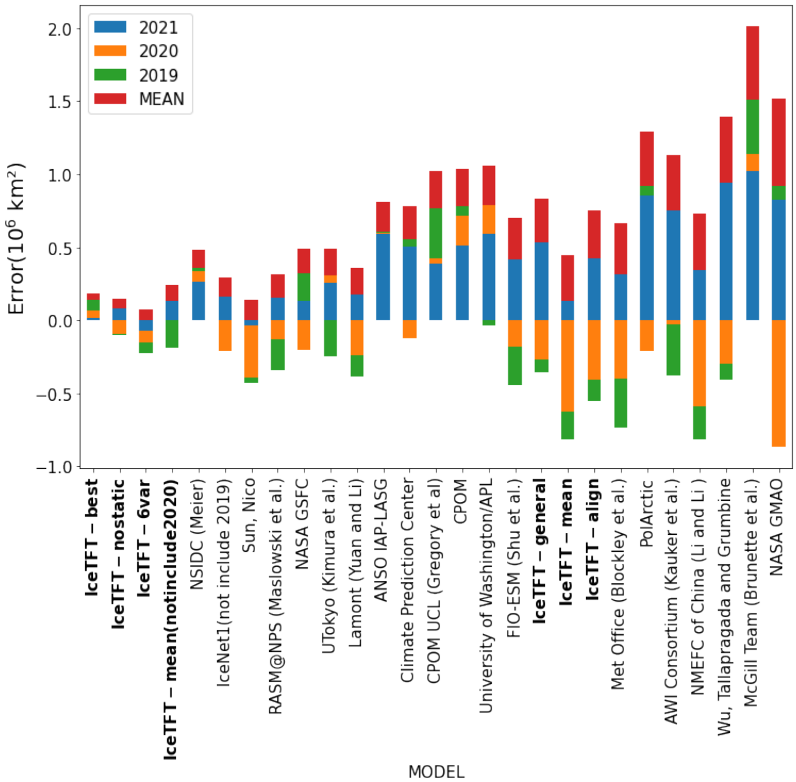


**Figure 8.** The error between predictions and observations of different models in three year

Specially, in June, we submitted to SIO the forecast result of the 2022 SIE minimum. According to the conclusions in Sect. 5.2, the closer the training set is to the forecast time, the higher skill the model has in forecasting. So we utilized the result of the IceTFT-Pred2021-best run as the forecast value for 2022. As we only use 12 months of data for 2021 to forecast 2022, and the forecast was nine months ahead for September, therefore, we did not submit new forecasts in July, August, and September additionally. Figure 9 shows the 2022 SIE predictions of different contributions of SIO in different lead time. It can be seen that the prediction results of the different models for 2022 are similar to the findings for the previous three years, and their forecasting skills do not improve with the reduction in lead time. Specially, the lead time of the contributions in SIO is up to three months, but our proposed model has a long lead time of up to nine months. Compared with the 2022 observed SIE which is 4.869 $10^6$ $km^2$, the closer predictions are from the IceTFT, LPHYS2268-CDDF and Kondrashov Dmitri (UCLA). Interestingly, all three of these models are based on statistical models or machine learning methods where they use SIE to predict SIE directly, instead of SIC. This suggests that using SIE to predict SIE has a smaller error than using SIC and can provide a favourable reference for SIE minimum prediction. For those contributions based on dynamical models, some of them have larger errors, and their predictions are erratic. For example, the model of Sun Nico performs relatively well compared to forecasts from other dynamical models, but they only submit predictions with small errors in June and September, with larger

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023




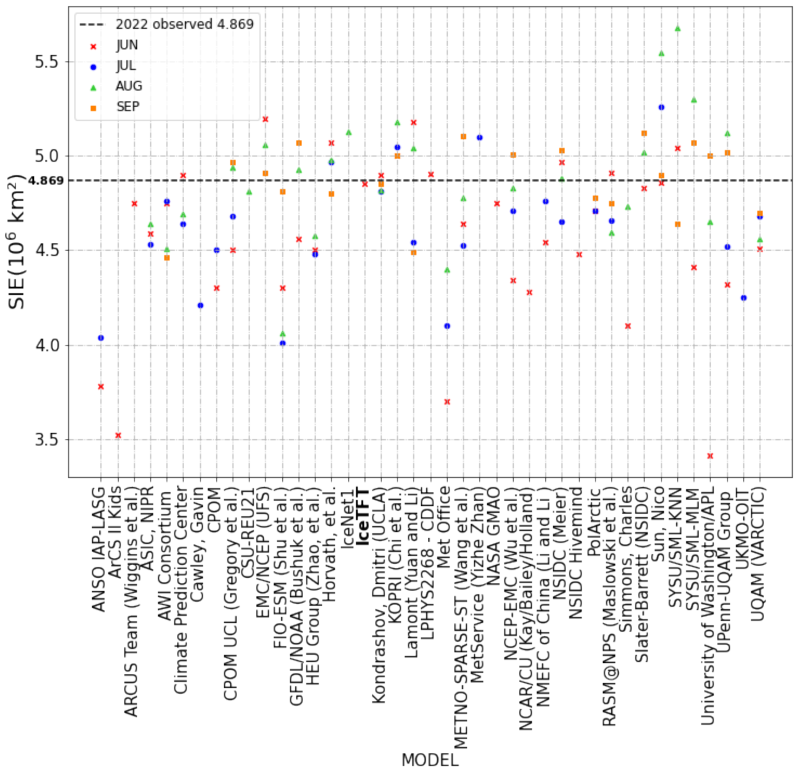


**Figure 9.** Contributions from SIO for June, July, August and September estimates of September 2022 pan-Arctic sea-ice extent.

errors in other months. This indicates that the dynamical model has the ability to predict SIE, but there is too much uncertainty, leading to unstable predictions. As can be seen from Fig. 8 and Fig. 9, our proposed model has higher forecasting skills than the other models in both hindcast experiments and real forecasts, and it obtains smaller prediction errors with longer lead times.

## 5.6 Validity of Static Time Metadata

We set the compared experiments to calculate the contributions of static metadata on the IceTFT model. We calculated the deviations between the predicted error of the IceTFT without static metadata and the IceTFT model. The results are presented in Table. 5. The blue areas indicate that the model without static metadata has less error than the model with static metadata. Conversely, red areas indicate the model without static metadata has higher error, and it represents the improvement due to static metadata. In IceTFT-Pred2019 run from Table. 5 (a), static metadata did not cause much improvement in 2019 predictions (more blue areas than red). And it caused more errors in some months, especially in April and October (blue areas). Fortunately, it had no apparent influence in September. And for 2020 and 2021 predictions, there was a significant positive effect from the results. Static metadata enables the model to learn more features with time dependence and predict unexpected values more accurately. Although the IceTFT model with static metadata had some negative effects on first year predictions, it made significant improvements in distance year predictions. That improves forecast skills in future.

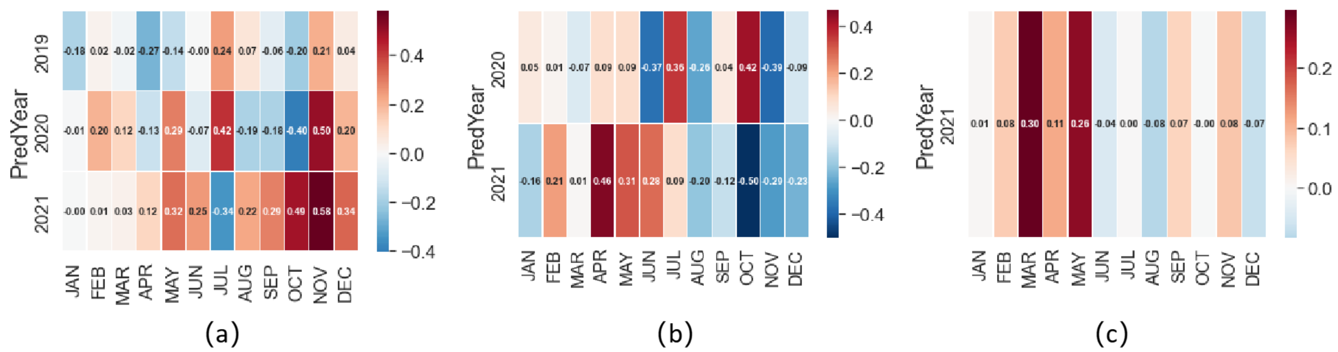


**Table 5.** The deviation accuracy between the IceTFT model without static metadata and IceTFT($IceTFTnostatic_{error} - IceTFT_{error}$), with the heatmap values shown within each grid cell. (a) IceTFT-Pred2019-nostatic run deviation accuracy relative to that of IceTFT-Pred2019 in three years. (b)IceTFT-Pred2020-nostatic run deviation accuracy relative to that of IceTFT-Pred2020 in two years. (c)IceTFT-Pred2021-nostatic run deviation accuracy relative to that of IceTFT-Pred2021 in 2021.

Similarly, the model with static metadata produced more errors in some months for first year predictions (2020) in Table. 5 (b) and for 2021 predictions in Table. 5 (c). And for distance year predictions (2021) Table. 5 (b), the positive effect is more obvious than 2020 (more red areas than 2020). It should be noted that the static metadata gives large errors for all months after July. This may be due to the fact that 2020 is a special year and the static metadata gives the model a strong temporal signature, influencing the forecast results based on data from previous years.

### 5.7 Interpretability Analysis

To investigate the contribution of different variables to SIE prediction in the model, we examined the variable sensitivity on different prediction time, which is from 2019 to 2021. Kim et al. (2020) added random Gaussian noises to inputs and calculated the change in RMSE to evaluate variable sensitivity. In this study, we applied this method to compare the contributions of variables. We added random Gaussian noises with a zero mean and one standard deviation to each variable in turn. Then we calculated the new RMSE of model with new inputs and compared the changes of RMSE (Eq.4).

$$Sensitivity(Var_x) = \frac{Changed \quad RMSE \quad with \quad variable \quad x \quad containing \quad noises}{Original \quad RMSE} \tag{4}$$

As shown in the Table. 6, we have bolded the values with higher sensitivities. Generally speaking, the sensitivity is greater than 1, which means that the variable plus noise increases the predicted error. However, when the sensitivity is less than 1, it indicates that the change in variable enhances the accuracy of the predictions. This may be because there is uncertainty in the original data, and the extra noise corrects the data in a beneficial direction for forecasting. The particular cases can give us new

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



ideas for improving forecast skills. To maintain the same range for all the data, the values less than 1 are taken as the inverse of them and marked with a negative sign.

**Table 6.** The eleven variable sensitivity of variables for the IceTFT model in INITIAL experiment

| Model | PredYear | SST | AT | DLWRF | DSWRF | PRATE | RUNOFF | CSDLF | CSDSF | USWRF | SH | SF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| IceTFT-Pred2019 | 2019 | **1.193** | **1.208** | **1.485** | **1.319** | **1.157** | -1.002 | 1.079 | **2.416** | **1.219** | 1.084 | 1.091 |
| IceTFT-Pred2020 | 2020 | 1.046 | 1.001 | 1.073 | **1.207** | -1.001 | **2.097** | **1.354** | **1.285** | **1.700** | -1.010 | -1.011 |
| IceTFT-Pred2021 | 2021 | **2.432** | **2.677** | 1.016 | **3.548** | 1.083 | **1.350** | 1.077 | **2.748** | **2.836** | 1.322 | 1.065 |

**Table 7.** The six variable sensitivity of variables for the IceTFT model in CTRL experiment

| Model | PredYear | SST | AT | DSWRF | RUNOFF | CSDSF | USWRF |
|---|---|---|---|---|---|---|---|
| IceTFT-Pred2019 | 2019 | **1.115** | 1.078 | -1.048 | **1.217** | **1.723** | **1.369** |
| IceTFT-Pred2020 | 2020 | **1.285** | 1.019 | 1.098 | **1.746** | -1.002 | **1.237** |
| IceTFT-Pred2021 | 2021 | **2.226** | 1.186 | **6.214** | 1.301 | **3.612** | **3.594** |

A higher sensitivity value indicates that the variable makes significant contributions to predictions. As for the predictions in three years, the variables with a high sensitivity to the predictions are SST, AT, DSWRF, RUNOFF, CSDSF and USWRF. Most of these variables are radiation-related, and shortwave radiation has a greater impact than longwave radiation. This finding is consistent with that in Fig.3, where surface air temperature and radiation fluxes influence sea surface temperature and thus sea ice melting. However, PRATE and SF do not contribute much to SIE predictions. Most likely because they are influenced by a variety of climatic factors and cannot be directly used in SIE predictions. Interestingly, 2020 is a more exceptional year than 2019 and 2021, which reached the second-lowest value for September on record. SST and AT have lower sensitivity to 2020 in our experiments. This provides a new idea for investigating the factors that influence the 2020 anomaly. And it may be because these 11 variables, which we select, are not the main factor of the unusually small values produced by 2020. Another reason is that these variables were treated as monthly mean estimates of global in the experiments and may have lost their relevance for the Arctic, leading to some impact on the forecast.

Multi-variate input of model can increase the training time and the uncertainty. We screened six variables of the highest contributors and redo the same experiments to further investigate the impact of these physical variables on sea ice forecasts. These variables include SST, AT, DSWRF, RUNOFF, CSDSF and USWRF. The results of the experiment are shown in the Table.7. The experiment with 11 variables is noted as INITIAL, while that with only 6 variables is noted as CTRL. Moreover, to analyze the prediction results after reducing the model inputs, we calculated the difference in predicted errors between the

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



two experiments. And we plotted the heat map as shown in Table.8. Negative values are shown in blue, indicating the CTRL experiment has a smaller error than the INITIAL. Conversely, positive values are indicated in red, representing lower errors for INITIAL experiment with 11 physical factors.

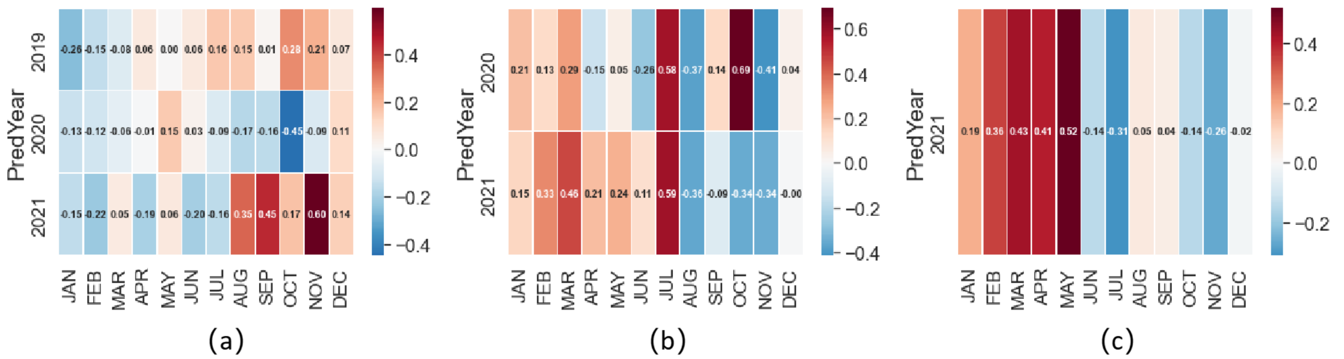


**Table 8.** The deviation accuracy between the IceTFT model six variables (CTRL) and the IceTFT with eleven variables (INITIAL) ($CTRL_{error} - INITIAL_{error}$), with the heatmap values shown within each grid cell. (a) IceTFT-Pred2019 with CTRL run deviation accuracy in three years compared to INITIAL. (b) IceTFT-Pred2020 with CTRL-run deviation accuracy relative to that with INITIAL in two years. (c) IceTFT-Pred2021 with CTRL run deviation accuracy compared to INITIAL.

From Table.6 and Table.7, it can be seen that all six variables had a high sensitivity in INITIAL experiment, but the sensitivity has changed in CTRL experiment. For 2019 prediction in IceTFT-Pred2019 run, only two of the six variables were relatively sensitive. This indicates that in CTRL experiment, the variable selection networks of IceTFT-Pred2019 run set a greater weight on the CSDSF, USWRF rather than SST and AT. These changes cause the more errors in summer, autumn, and winter, but fewer errors in spring, as can be seen from the first row of Table.8 (a). The most likely explanation is that SST, AT, and other variables from Table.6 have a greater impact on summer and autumn predictions for 2019. And as can be seen from the second row of Table.8 (a), the CTRL experiment has fewer errors almost monthly for 2020 predictions. It is also because that SST and AT have lower sensitivity in CTRL experiment, and this conclusion is consistent with the INITIAL experiment, which suggests that SST and AT may not be the main factors affecting the 2020 predictions. In addition, the impact for 2021 predictions is similar to that for 2019 in the last row of Table.8 (a), and the red areas are darker in autumn and winter. It demonstrates that the factors affecting the 2019 predictions are similar to those for 2021, and SST and AT have a greater impact on 2021 than 2019. This also validates the experiment results in Table.4, explaining that the IceTFT-Pred2019 run had higher accuracy for 2021 predictions.

For 2020 prediction in IceTFT-Pred2020 run, some of the selected variables were not sensitive to their predictions in INITIAL experiment, the high sensitive variables were not fully included. Consequently, their prediction errors have changed significantly in CTRL experiment according to Table.8 (b). From Table.7, the sensitivity of SST became higher, but that of RUNOFF, DSWRF and CSDSF decreased in CTRL experiment. According to the first row of Table.8 (b), these changes cause more errors than gains, especially higher errors in spring, July, and October. This may be due to the fact that the three variables

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023



with reduced sensitivity have a greater impact on the 2020 predictions. But the more sensitive SST than INITIAL experiment makes model in CTRL experiment improve the forecasting skills for 2021 predictions significantly in autumn and winter. Compared with those high sensitivity variables of IceTFT-Pred2021 run in INITIAL experiment from Table.6, AT and other radiation-related variables are not highly weighted of IceTFT-Pred2020 run in CTRL experiment, increasing 2021 predicted errors in spring and summer from the second row of Table.8 (b). It indicates that AT and radiation-related variables may have an important impact on 2021 predictions in spring and summer.

Similarly, from the results of Table.6 and Table.7, the variables that significantly contributed to the 2021 predictions in the INITIAL experiment were still selected in the CTRL experiment, but the weights of these sensitive variables have also changed. In particular, the sensitivity of DSWRF becomes the highest, much higher than the other radiation variables, while the sensitivity of AT decreases. The model with IceTFT-Pred2021 run also generated more accurate predictions after June and more errors in spring and early summer, as shown in Table.8 (c). The results are similar to the 2021 predictions in IceTFT-Pred2020 which are shown in Table.8 (b). However, DSWRF, CSDSF and USWRF these radiation-related variables all had high sensitivities for IceTFT-Pred2021 run in CTRL experiment, DSWRF is much more sensitive than the other two variables. In comparison to the INITIAL experiment, these variables had comparable sensitivities. The imbalanced weights led to an increase in predicted errors in CTRL experiment. This suggests that there is some link between these radiation-related variables that collectively affect forecasting skills.

Previous research has demonstrated that sea ice melting is influenced by a complex set of radiative feedback mechanisms(Goosse et al., 2018). Warming Arctic air temperatures cause sea ice to melt, exposing large amounts of sea surface and thus reducing albedo. The absorption of solar shortwave radiation by the ocean raises sea surface temperatures, which triggers an Arctic amplification effect and creates a positive feedback mechanism that exacerbates the melting of sea ice (Perovich et al., 2007; Screen and Simmonds, 2010). It can be seen that among these processes, AT and SST are the direct factors that influence the melting of sea ice, while longwave radiation and shortwave radiation play an indirect role in this positive feedback mechanism. Consequently, during the melt season, a relatively small area of sea ice cover exposes a large area of sea surface, and warming seawater affects sea ice melt. Since our model cannot simulate the process of radiation absorption by the ocean, SST can provide the IceTFT model with a direct factor affecting sea ice melt. However, for the freezing season, when the sea ice cover is large and the exposed sea surface area is small, the effect of SST on sea ice melt is relatively small. Rather, heat fluxes and warming air temperatures from water vapor, cloud cover and radiation mechanisms have a greater effect on sea ice melt(Kapsch et al., 2013; Boisvert and Stroeve, 2015). Thus validating the conclusions of our experiments that SST is an important factor influencing prediction from August to October, while radiation-related variables and AT are from January to May. In addition, we found in our experiment that 2020, as the anomalous year with the second lowest SIE on record, had a different main factor affecting its SIE prediction than the other years. Other researches have shown that the influence on 2020 SIE is primarily caused by the relaxation of the Arctic dipole(Liang et al., 2022). Whereas it can be seen from our experimental results, neither SST nor AT is a major factor affecting the 2020 SIE prediction, we will continue to investigate the reasons affecting the SIE anomaly in the future.

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023

Geoscientific Model Development Discussions

## 6 Conclusions

In this study, an interpretable long-term forecast model for the annual forecast of SIE, IceTFT, is developed. It employs the LSTM encoder, multi-headed attention mechanism, and static time metadata to enhance the learning of the temporal dependence of the SIE, and the variable selection network to filter unnecessary noise from the input data. In contrast to other models, the IceTFT model uses a total of 11 variables, including atmospheric, oceanic, and sea ice, as inputs to provide relevant mechanisms about the sea ice development process. Moreover, the IceTFT model can directly predict the next year SIE based on the previous year data, avoiding the need to train multiple models or the error accumulation of iterative forecasting. In our experiments, we analyze the impact on the forecasts of time periods of the training set, slicing methods of input data, and the length of input. The results show that the closer the data is to the forecast time, the greater the contribution to the forecast. And the rolling method to slice data increases the number of dataset, improving the accuracy of prediction. Furthermore, the 12-step input includes the entire SIE cycle, and its forecast skill is higher than that of the 6-step input model. We employ the MAE, RMSE and RMSD metrics to evaluate the accuracy of the predictions in the IceTFT model for three years, from 2019 to 2021. The experimental results show that the monthly average forecast error of IceTFT model is less than 0.21 $10^6$ $km^2$. For the SIE minimum forecast, compared to other models in SIO that only provide three-month advance forecasts, IceTFT model not only has the smallest forecast error, with a three-year average SIE minimum forecast error of less than 0.05 $10^6$ $km^2$, but it also provides nine-month advance forecasts. Moreover, the IceTFT model still has similar higher forecasting skill in the actual forecast of the 2022 SIE minimum. Finally, we conducted a sensitivity analysis of the variables in the IceTFT model to investigate the physical factors affecting SIE forecasts. The results indicate that the factors affecting the 2020 SIE forecast are different from those of other years. Except for 2020, for the melt season, SST has a greater influence on SIE forecasts, while for the freeze season, AT and radiation-related variables have a greater influence than SST. These sensitivities can help researchers study the mechanisms of sea ice development, and they also provide useful references for the selection of variables in data assimilation or the input of deep learning models.

*Code and data availability.* The source code of the IceTFT is available at https://doi.org/10.5281/zenodo.7409157(Luo,2022). Thanks to National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Laboratory(PSL), Boulder Climate and Weather Information for providing the NCEP-NCAR Reanalysis 1 data (https://psl.noaa.gov/, Rayner et al., 1996) and Boulder Monthly Means: Snowfall(https://psl.noaa.gov/); NOAA National Centers for Environmental Information (NCEI) for providing Optimum Interpolation SST V2 data (https://www.ncei.noaa.gov/data/sea-surface-temperature-optimum-interpolation/, Reynolds et al., 2007;Reynolds et al., 2007;Huang et al., 2020); National Snow and Ice Data Center, a part of CIRES at the University of Colorado Boulder for providing Sea Ice Index, Version 3 data(https://nsidc.org/data/g02135, Fetterer et al., 2017).

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023


Geoscientific Model Development Discussions

*Author contributions.* All authors designed the experiments and carried them out. Xiaodan Luo developed the model code and performed the experiments. Bin Mu and Xi Liang reviewed and optimized the code and experiments. Xiaodan Luo and Shijin Yuan prepared the manuscript with contributions from all co-authors.

*Competing interests.* The contact author has declared that neither they nor their co-authors have any competing interests.

*Acknowledgements.* This study was supported in part by the National Natural Science Foundation of China (grant no. U2142211), in part by the National Key Research and Development Program of China (grant no. 2020YFA0608002), in part by the National Natural Science Foundation of China (grant no. 42075141), in part by the Key Project Fund of Shanghai 2020 “Science a Technology Innovation Action Plan” for Social Development (grant no.20dz1200702).

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023


Geoscientific Model Development
Discussions

## References

Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D. C., Wilkinson, J., Phillips, T., Byrne, J., Tietsche, S., Sarojini, B. B., Blanchard-Wrigglesworth, E., Aksenov, Y., Downie, R., and Shuckburgh, E.: Seasonal Arctic sea ice forecasting with probabilistic deep learning, Nature communications, 12, 5124, 2021.

Bintanja, R. and Selten, F. M.: Future increases in Arctic precipitation linked to local evaporation and sea-ice retreat, Nature, 509, 479–482, 2014.

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

Boisvert, L. N. and Stroeve, J. C.: The Arctic is becoming warmer and wetter as revealed by the Atmospheric Infrared Sounder, Geophysical Research Letters, 42, 4439 – 4446, 2015.

Bushuk, Mitchell, Giannakis, and Dimitrios.: The Seasonality and Interannual Variability of Arctic Sea Ice Reemergence., Journal of Climate, 2017.

Chi, J. and Kim, H. C.: Prediction of Arctic Sea Ice Concentration Using a Fully Data Driven Deep Neural Network, Remote Sensing, 9, 1305–, 2017.

Chi, J., Bae, J., and Kwon, Y.-J.: Two-Stream Convolutional Long- and Short-Term Memory Model Using Perceptual Loss for Sequence-to-Sequence Arctic Sea Ice Prediction, Remote Sensing, 13, https://doi.org/10.3390/rs13173413, 2021.

Choi, Y. S., Kim, B. M., Hur, S. K., Kim, S. J., Kim, J. H., and Ho, C. H.: Connecting early summer cloud-controlled sunlight and late summer sea ice in the Arctic, Journal of Geophysical Research, D. Atmospheres: JGR, 2014.

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., and Overland, J. a.: Recent Arctic amplification and extreme mid-latitude weather, Nature Geoscience, 7, 627–637, 2014.

Fetterer, F., K. K. W. N. M. M. S. and Windnagel., A. K.: Sea Ice Index, Version 3, https://doi.org/10.7265/N5K072F8, 2017.

Fritzner, S., Graversen, R., Christensen, K. H., Rostosky, P., and Wang, K.: Impact of assimilating sea ice concentration, sea ice thickness and snow depth in a coupled ocean–sea ice modelling system, The Cryosphere, 13, 491–509, 2019.

Goosse, H., Kay, J. E., Armour, K. C., Bodas-Salcedo, A., Chepfer, H., Docquier, D., Jonko, A. K., Kushner, P. J., Lecomte, O., Massonnet, F., Park, H., Pithan, F., Svensson, G., and Vancoppenolle, M.: Quantifying climate feedbacks in polar regions, Nature Communications, 9, 2018.

He-Ping, L. I., You-Ming, X. U., and Rao, S. Q.: Analysis on Influence of Sea Ice in North Pole Area on Runoff in the Upper Yellow River during Flood Seas on, ADVANCES IN WATER SCIENCE, 2000.

Huang, Y., Kleindessner, M., Munishkin, A., Varshney, D., Guo, P., and Wang, J.: Benchmarking of Data-Driven Causality Discovery Approaches in the Interactions of Arctic Sea Ice and Atmosphere, Frontiers in Big Data, 4, https://doi.org/10.3389/fdata.2021.642182, 2021.

Johannessen, O. M., Bobylev, L. P., Shalina, E. V., and Sandven, S.: Sea ice in the Arctic: past, present and future, Springer, 2020.

Kapsch, M. L., Graversen, R. G., and Tjernstr?M, M.: Springtime atmospheric energy transport and the control of Arctic summer sea-ice extent, Nature Climate Change, 3, 744–748, 2013.

Kapsch, M.-L., Graversen, R. G., Tjernström, M., and Bintanja, R.: The Effect of Downwelling Longwave and Shortwave Radiation on Arctic Summer Sea Ice, Journal of Climate, 29, 1143 – 1159, https://doi.org/10.1175/JCLI-D-15-0238.1, 2016.

Kim, Y. J., Kim, H. C., Han, D., Lee, S., and Im, J.: Prediction of monthly Arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks, The Cryosphere, 14, 1083–1104, 2020.

Kwok, R. and Untersteiner, N.: The thinning of Arctic sea ice, Phys. Today, 64, 36–41, 2011.

Liang, X., Li, X., Bi, H., Losch, M., Gao, Y., Zhao, F., Tian, Z., and Liu, C.: A Comparison of Factors That Led to the Extreme Sea Ice Minima in the Twenty-First Century in the Arctic Ocean, Journal of Climate, 35, 1249 – 1265, https://doi.org/10.1175/JCLI-D-21-0199.1, 2022.

Lim, B., Arık, S. Ö., Loeff, N., and Pfister, T.: Temporal fusion transformers for interpretable multi-horizon time series forecasting, International Journal of Forecasting, 37, 1748–1764, 2021.

Liu, J., Song, M., Horton, R. M., and Hu, Y.: Reducing spread in climate model projections of a September ice-free Arctic, Proceedings of the National Academy of Sciences of the United States of America, 110, 12 571–12 576, 2013.

Liu, Q., Zhang, R., Wang, Y., Yan, H., and Hong, M.: Daily Prediction of the Arctic Sea Ice Concentration Using Reanalysis Data Based on a Convolutional LSTM Network, Journal of Marine Science and Engineering, 9, 330, 2021.

Liu, X. Y. and Liu, H. L.: Investigation of influence of atmospheric variability on sea ice variation trend in recent years in the Arctic with numerical sea ice-ocean coupled model, Chinese Journal of Geophysics, 55, 2867–2875, 2012.

Luo, B., Luo, D., Wu, L., Zhong, L., and Simmonds, I.: Atmospheric circulation patterns which promote winter Arctic sea ice decline, Environmental Research Letters, 12, 054017, https://doi.org/10.1088/1748-9326/aa69d0, 2017.

Massonnet, F., Mathiot, P., Fichefet, T., Goosse, H., Beatty, C. K., Vancoppenolle, M., and Lavergne, T.: A model reconstruction of the Antarctic sea ice thickness and volume changes over 1980-2008 using data assimilation, Ocean Modelling, 64, 67–75, 2013.

Mathiot, P., Beatty, C. K., Fichefet, T., Goosse, H., Massonnet, F., and Vancoppenolle, M.: Better constraints on the sea-ice state using global sea-ice data assimilation, Geoscientific Model Development,5,6(2012-12-03), 5, 1627–1667, 2012.

Meier, W., U. S. B.-J. W. R. T. P. B. C. M. B. E. B.-W. H. E. L. C. H. M. H. E. H. T. J. J. K. J. L. F. M. J. E. O. M. S. M. S. J. S. M. W., H. V. Wiggins. Editors: Turner-Bogren, B., L. S. G., and Staudt., S.: 2020 Sea Ice Outlook Post-Season Report, Website, https://www.arcus.org/sipn/sea-ice-outlook/2020/post-season, 2021.

Overland, J. E. and Wang, M.: When will the summer Arctic be nearly sea ice free?, Geophysical Research Letters, 40, 2097–2101, 2013.

Perovich and D., K.: Seasonal evolution of the albedo of multiyear Arctic sea ice, Journal of Geophysical Research, 107, 8044, 2002.

Perovich, D. K., Light, B., Eicken, H., Jones, K. F., Runciman, K., and Nghiem, S. V.: Increasing solar heating of the Arctic Ocean and adjacent seas, 1979–2005: Attribution and role in the ice-albedo feedback, Geophysical Research Letters, 34, 2007.

Ramsayer, K.: 2020 Arctic Sea Ice Minimum at Second Lowest on Record, NASA Global Climate Change. Vital Signs of the Planet, 2020.

Ren, Y., Li, X., and Zhang, W.: A data-driven deep learning model for weekly sea ice concentration prediction of the Pan-Arctic during the melting season, IEEE Transactions on Geoscience and Remote Sensing, 2022.

Screen, J. A. and Simmonds, I.: The central role of diminishing sea ice in recent Arctic temperature amplification, Nature, 464, 1334–1337, 2010.

Screen, J. A. and Simmonds, I.: Declining summer snowfall in the Arctic: causes, impacts and feedbacks, Climate Dynamics, 38, 2243–2256, 2012.

Stroeve, J., Hamilton, L. C., Bitz, C. M., and Blanchard-Wrigglesworth, E.: Predicting September sea ice: Ensemble skill of the SEARCH Sea Ice Outlook 2008-2013, Geophysical Research Letters, 41, 2411–2418, 2014.

Stroeve, J. C., Kattsov, V., Barrett, A., Serreze, M., Pavlova, T., Holland, M., and Meier, W. N.: Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, Geophysical Research Letters, 39, 2012.

Sturm and Matthew: Winter snow cover on the sea ice of the Arctic Ocean at the Surface Heat Budget of the Arctic Ocean (SHEBA): Temporal evolution and spatial variability, Journal of Geophysical Research, 107, 8047, 2002.

https://doi.org/10.5194/gmd-2022-293
Preprint. Discussion started: 16 January 2023


Tong, J., Chen, M., Qiu, Y., Yanping, L. I., Cao, J., Sciences, O. E., University, X., and of Marine Environmental Science, S. K. L.: Contrasting patterns of river runoff and sea-ice melted water in the Canada Basin, Acta Oceanologica Sinica, 2014.

Toyoda, T., Fujii, Y., Yasuda, T., Usui, N., Ogawa, K., Kuragano, T., Tsujino, H., and Kamachi, M.: Data assimilation of sea ice concentration into a global ocean–sea ice model with corrections for atmospheric forcing and ocean temperature fields, Journal of Oceanography, 72, 235–262, 2016.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need, Advances in neural information processing systems, 30, 2017.

Voosen, P.: New feedbacks speed up the demise of Arctic sea ice, 2020.

Wayand, N. E., Bitz, C. M., and Blanchard-Wrigglesworth, E.: A Year-Round Subseasonal-To-Seasonal Sea Ice Prediction Portal, Geophysical Research Letters, 2019.

Weatherly, J. W. and Walsh, J. E.: The effects of precipitation and river runoff in a coupled ice-ocean model of the Arctic, Climate Dynamics, 12, 785–798, 1996.

Wei, K., Liu, J., Bao, Q., He, B., Ma, J., Li, M., Song, M., and Zhu, Z.: Subseasonal to seasonal Arctic sea-ice prediction: A grand challenge of climate science, Atmospheric and Oceanic Science Letters, 14, 100 052, https://doi.org/https://doi.org/10.1016/j .aosl.2021.100052, 2021.

Xi, L., Losch, M., Nerger, L., Longjiang, M., Qinghua, Y., and Chengyan, L.: Using Sea Surface Temperature Observations to Constrain Upper Ocean Properties in an Arctic Sea Ice-Ocean Data Assimilation System, Journal of Geophysical Research: Oceans, 124, 2019.

Zheng, F., Sun, Y., Yang, Q., and Longjiang, M. U.: Evaluation of Arctic Sea-ice Cover and Thickness Simulated by MITgcm, Advances in Atmospheric Sciences, 38, 20, 2021.