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 predict, National Marine Environmental predict Center, Beijing, China
[*]These authors contributed equally to this work.

Correspondence: Shijin Yuan (yuanshijin2003@163.com)

Abstract.

Annual reductions in Due to global warming, Arctic sea ice extent (SIE) due to global warmingis rapidly decreasing each year. According to the International Panel on Climate Change (IPCC) climate model projections, the summer Arctic will be nearly sea ice 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 which will have a great impact on global climate change. As a result, accurate predictions of Arctic sea ice are of significant interest. In most current studies, the majority of deep learning-based SIE prediction models focus on single-step one-step prediction, and they not only have short lead times but also have limited forecasting skills. In additionlimited predicting skills. Moreover, these models often lack interpretability. In this studypaper, we construct the Ice Temporal Fusion Transformer (IceTFT) model, which consists mainly mainly consists of the variable selection network (VSN), the long short-term memory (LSTM) encoder, and multi-headed attention mechanism. Then we We select 11 predictors for the IceTFT model, including SIE, atmospheric , and ocean and oceanic variables according to the physical mechanisms influencing affecting sea ice development. And the VSN in IceTFT can automatically adjust the weights of predictors and filter spuriously correlated variables. We also The IceTFT model can provide the 12-month SIE directly according to the inputs of the last 12 months. We 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 hindcasting experiment for 2019-2021 and prediction for 2022. For the hindcasting of 2019-2021, the average monthly prediction errors of are less than 0.21 10^6 km^2. And it predicts the September SIE nine months in advance with prediction error of , and the September prediction errors are less than 0.1 10^6 km^2, which is superior to the models from Sea Ice Outlook (SIO). Furthermore, For September prediction of 2022, we submitted the prediction to the SIO in 2022 June and the IceTFT still has higher prediction skills. Furthermore, the VSN in IceTFT can automatically adjust the weights of predictors and filter spuriously correlated variables. Based on this, we analyze the sensitivity of the selected predictors to the SIEprediction. It verifies predict SIE. This confirms that the IceTFT model has some a 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 temperature rise in the Arctic has accelerated. This phenomenon, known as Arctic amplification, has hastened accelerated the melting of Arctic sea ice, which may have a potential influence impact on weather patterns and the climate of the Northern Hemisphere(Liu et al., 2013; Cohen et al., 2014). According to the 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 impact on global climate change. Therefore, it is important to predict the development of Arctic sea ice, and it can provide a strong which can be an important reference for studying and predicting global climate change trends.

However, in recent years, the Arctic sea ice Over the past few decades, the Arctic Ocean has been warming (Polyakova et al., 2006) and Arctic sea ice is melting rapidly. The September Arctic sea ice extent (SIE) has been observed to be decreasing rapidly. From 1979 to 2013, the average September SIE decreased by around declined on average by about 14% per decade from 1979 to 2013, and by approximately about 50% by 2020 (Johannessen et al., 2020; Ramsayer, 2020). It The SIE in September 2020 is the second lowest in the satellite record (1979–2020), according to data value from the National Snow and Ice Data Center (NSIDC) Sea Ice Index V3 (SII). The rapid (SII) (SII) data (1979-2020). 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 difficult.

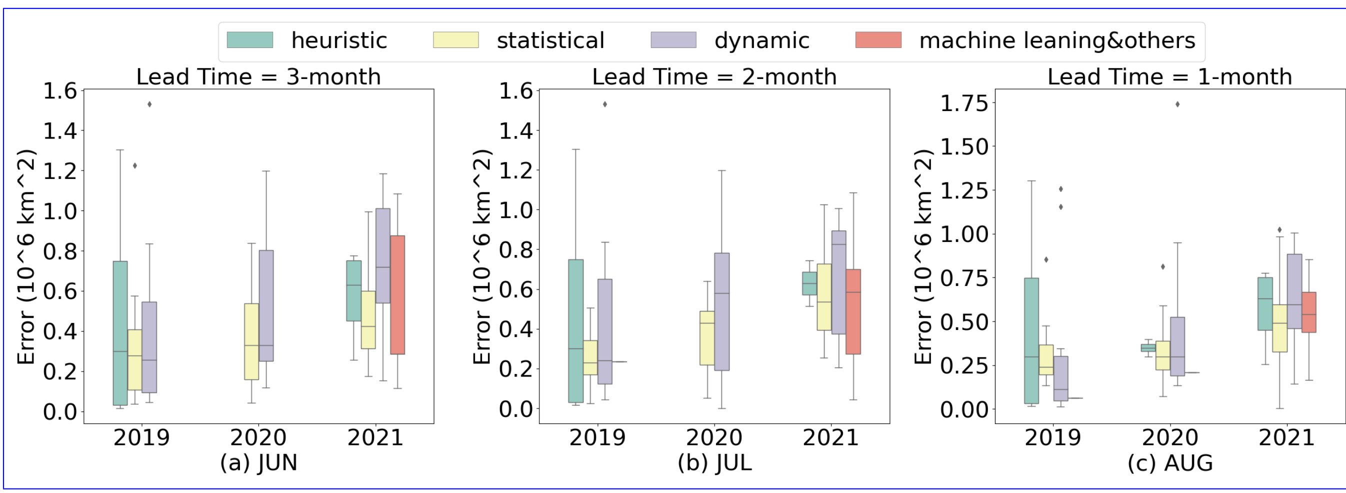

Figure 1. The September SIE prediction errors with a lead time of 3 (2,1) months from June (a) (July (b),August (c)) for 2019-2021, which are published in the Sea Ice Outlook (SIO) by Sea Ice Prediction Network (SIPN)

SIE has extremely cyclical, always reaching the maximum in March and the minimum in September(Kwok and Untersteiner, 2011). It is very difficult to predict the September minimum due to the influence of multiple physical factors. Consequently, the accurate forecast for SIE has become a challenging focus.

Since 2008,the annual Figure 1 a (b, c) shows that the September SIE prediction errors with a lead time of 3 (2,1) months of 2019 to 2021, which are published in the Sea Ice Outlook (SIO) has been provided by the by 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 . Since 2008, the SIPN collects annually September predictions with a lead time of 1-3 months from global research institutions. And it represents the current forecast predict level and community knowledge of the state and evolution of Arctic sea ice on the sub-seasonal-to-seasonal (S2S) timescale . For example, the average SIE in September 2020 (Wei et al., 2021). From Fig.1, it can be seen that there 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 predictions and observations. Surprisingly, prediction skill did not improve significantly as predict lead time was reduced, it is consistent with other study (Stroeve et al., 2014). According to the SIO, 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. 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.

Fig.1 also shows the forecast skills in SIO over 2019–submissions for machine learning until 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 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. As a result, deep learning can learn the features of the non-linear development of sea ice, which is extremely promising for sea ice prediction.

Up to nowIn recent years, deep learning (DL) methods have been progressively employed to predict Arctic increasingly used to predict sea ice. Such as Chi and Kim (2017) firstly applied long short-term memory (LSTM) to Chi and Kim (2017) first applied LSTM to a one-month Arctic SIC prediction model , and only used observed sea ice data by remote sensing sensors as an input of that. Then , forecast model for the sea ice concentration (SIC) prediction. 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 SIC prediction model using convolutional neural networks (CNNs) by incorporating that incorporate SIC, atmospheric, and oceanic variables. Due to the CNN cannot capture the time series time-series dependence, they trained 12 models to provide each month prediction of the year produce predictions for each month 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 these models. Andersson et al. (2021) proposed the IceNet model that learns 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 observation data. And they also trained multiple models to provide months of 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 used the iteration method to get obtain a weekly SIC forecastprediction. In these studies, they only focus on one-step models. To provide produce long-term predictions, they used recursive approach a recursive approach, which can result in increasing errors, or they trained more models, which can lead to a growing may increase the cost and time of calculation. Therefore, these studies make it very evident that there computation. These studies highlight that long-term prediction has been less research on long-term prediction than on researched than short-term forecasting. That prediction. This ignores the periodicity of SIE, which reaches its maximum in March and minimum in September(Kwok and Untersteiner, 2011). In addition, little attention has been paid to the explainability of the deep learning modelhas 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 . Different variables were used as the input inputs of the proposed model for different channels, which does not shed light on provide insight into how the model utilizes the full channels of the input. These channels data may have an incomprehensible impacteffect. Although the model in Andersson et al. (2021) has been was pre-trained on climate simulation, the impact of that with climate simulations, the effect on prediction is also unexplained. Compared to dynamic models, DL deep learning models are considered as "black box" due to a "black box" due to the lack of physical mechanisms.

To Our research team constructs deep learning models with interpretable and high prediction skills based on the physical mechanisms of various weather and climate phenomena, which include ENSO, NAO, etc (Mu et al., 2019, 2020, 2021, 2022). In this paper, to improve the long-term forecast prediction skill for SIE and analyze the impact of different effects of various factors on SIEforecasting, we introduce a new SIE prediction model , IceTFT, which is an interpretable model based on the Temporal Fusion Transformer (TFT) (Lim et al., 2021), IceTFT, an interpretable model with high prediction skill. The IceTFT model directly predicts 12 months of averaged Arctic can directly predict 12-month 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 prediction. We select 11 predictors based on the physical mechanisms and correlation analysis of Arctic sea icedevelopment, we select several atmospheres, ocean variables as input to improve the forecast skill. And the design of the , which include SIE, atmospheres, and ocean variables. The variable selection network (VSN) design 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 modelsin SIO. On this basis, we can conduct sensitivity analysis experiments to quantify the role of predictors on the SIE prediction. The physical mechanisms affecting sea ice development can also be identified, which can provide a reference for selecting assimilation variables for dynamical models. 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 developmentwe submitted the September prediction of 2022 to SIO in June 2022. The prediction skill of IceTFT with lead time of 9 months outperforms most other models.

The contributions of this paper are as follows:

1) The IceTFT model directly provides uses LSTM encoders to summarize past inputs and generate context vectors, so it can directly provide a 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 . And it can predict September SIE 9 months in advancewith a small error. 3, which is longer than other studies with lead time of 1-3 months. IceTFT has the lowest prediction errors for hind-cast experiments from 2019 to 2021 and actual prediction of 2022, which compared with SIO.

2) The IceTFT model is interpretable. It can automatically filter out dummy correlation variables , avoiding spuriously correlated variables and adjust the weight of inputs through VSN, reducing noise 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 it can also explore the contribution of different input variables to SIE forecasts and uncover predictions and reveal the physical mechanisms of sea ice development.

The remaining part remainder 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 we selected. Sect. 4 introduces the evaluation metrics used. Sect. 5 presents the experiment designs and evaluation , and the experiment results and analyzes affecting SIE prediction are introduced optimal setting of IceTFT model. The results of hindcasting experiments from 2019 to 2021 and of prediction for 2022 are presented in Sect. 5. 6 and Sect. 7. Sect. 8 discusses the contribution of the inputs to the SIE predictions and the analysis of the physical mechanisms by which they affect sea ice.

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 Deep learning has good performance in this area, however, time series prediction, but previous research mostly used CNNwhich 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 , ConvLSTM, which still have high prediction errors. The transformer model makes the attention mechanism in Transformer model fully capture the time series temporal dependence, and it has higher performance than other traditional recurrent neural networksperforms better than the traditional Recurrent Neural Network (RNN) models (Vaswani et al., 2017). Based on Transformer model, Temporal Fusion Transformer the transformer model, temporal fusion transformer (TFT) was proposed for multi-step forecasting prediction (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 at the local level, but also uses a multi-head attention block to capture long-term dependencies.

Due to the pronounced periodicity of sea iceThe VSN design in the TFT model species adjusts the weights of the variables and makes it interpretable. The TFT has been verified that it has small prediction errors in several areas.

The sea ice dataset is a time series with pronounced periodicity, which has one peak and one a peak and a 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 prediction. And sea ice is affected by multiple 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.

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.

Therefore, we construct the SIE yearly forecast structure model named IceTFT, which is modified in the design of input and loss function making its prediction more difficult. We propose the IceTFT model for SIE prediction based on TFT. , as follows:

(1) The original design of TFT used known future data to support in the prediction of the primary time series data. Since sea ice melting can be affected by various physical factors, and the various mechanisms responsible for sea ice variability have not yet been elucidated. To help the model learn the physical mechanics underlying the mechanisms underlying SIE, we modified this that part to use atmospheric and oceanographic variables with the same moment as the SIE. In addition, TFT is SIE.

(2) The TFT relies on positional encoding to capture temporal features. When time series data is rolled into the model, it may lose the temporal information of input datathe input data may be lost. To solve this problem, we set time static 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

(3) The original TFT uses quantile prediction as a loss function. Since SIE has decreased in recent years, there is some mutagenicity in summer. Therefore, the original design is not appropriate for predicting SIE. We used the mean square error (MSE) as the loss function to replace it.

The IceTFT architecture is shown in Fig.2. Three types of dataset datasets are the inputs of the IceTFT, and to IceTFT, which include time-static metadata, SIE and other physical variables. And each type is selected by a separate variable selection network to filter unnecessary noises. Through the use of VSN to filter out unnecessary noise. The structures of VSN and Gated Residual Network (GRN) are shown in Fig. 3. By using the GRN, the VSN calculates the weight of each variable of predictioncontributioncontributing to the prediction, allowing the model to focus on the most significant features rather than overfitting irrelevant features. And it can filter spuriously The VSN can also filter out spurious correlated variables to improve the accuracy of SIE predictions. This

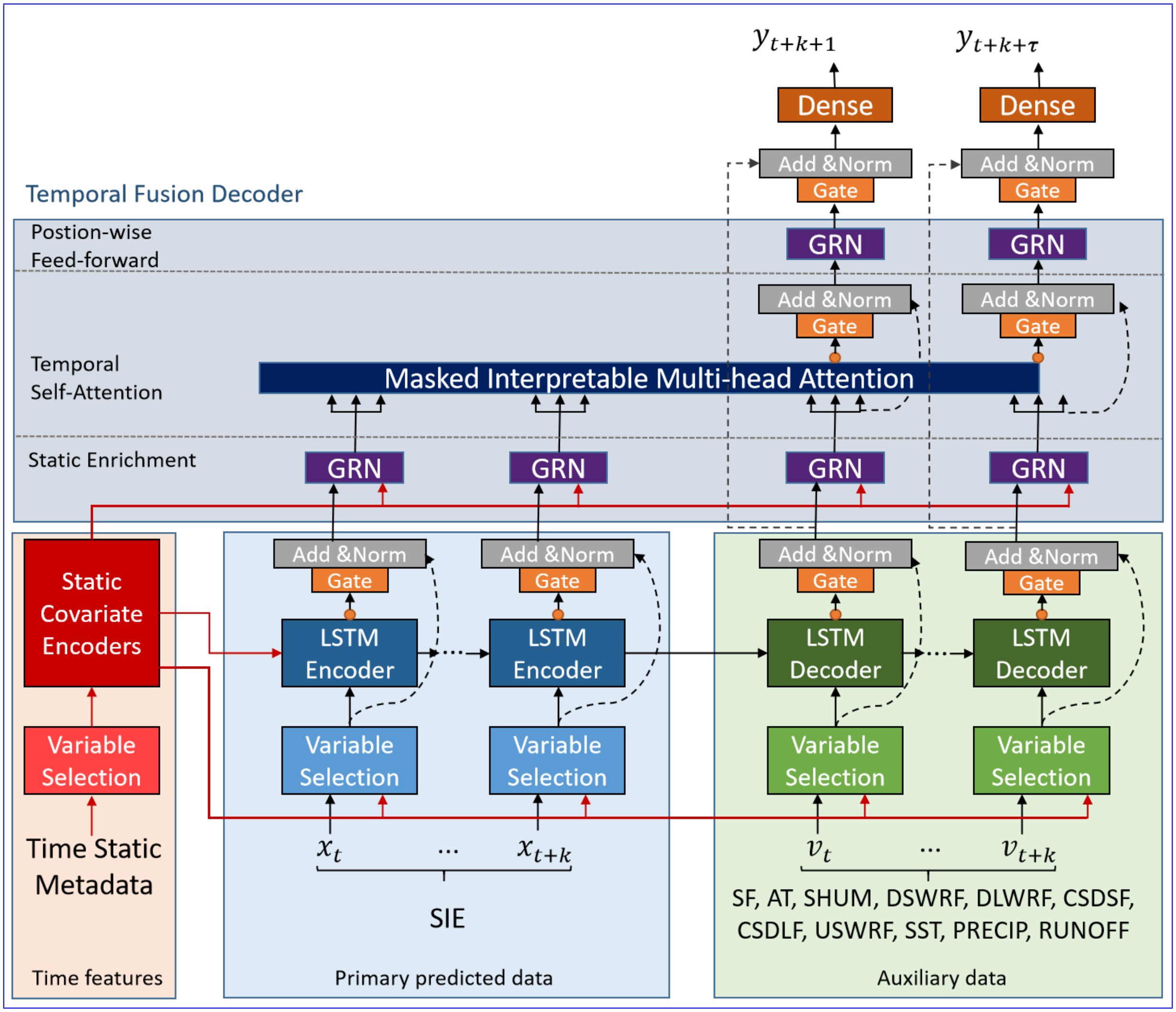

Figure 2. The IceTFT architecture is adapted on the basis of original TFT(Lim et al., 2021). The static time metadata, historical SIE data and other atmospheric and oceanographic variables are all inputs to the IceTFT. The auxiliary data include snowfall (SF),x 2m air temperature (AT), 2m surface air specific humidity (SHUM), downward shortwave radiative flux (DSWRF), downward longwave radiation flux (DLWRF), clear sky downward longwave flux (CSDLF), clear sky downward solar flux (CSDSF), upward solar radiation flux (USWRF), sea surface temperature (SST), precipitation (PRECIP) and river runoff (RUNOFF).

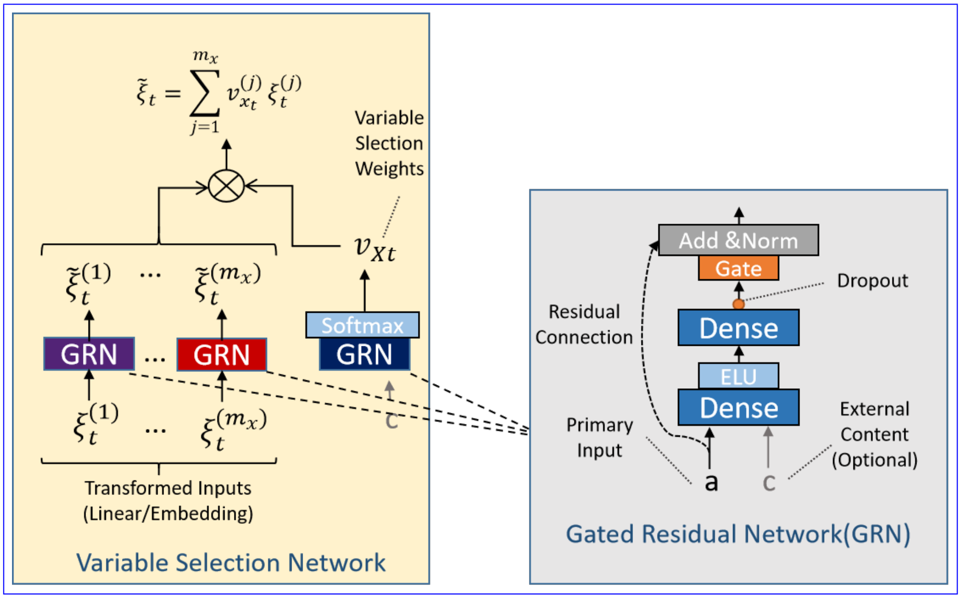

Figure 3. The components in the IceTFT. The Variable Selection Network (VSN) is used to select the most useful features. The Gated Residual Network (GRN) enables efficient information flow by skipping connections and gating layers.

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

We can define the SIE prediction with IceTFT as a multivariate spatial–temporal sequence prediction problem as illustrated in Eq. 1,

$$\hat{\mathrm{SIE}} = \mathrm{F}_\theta(\mathrm{X}), \{\mathrm{TIME}_{\mathrm{static}}, \mathrm{SIE}, \mathrm{VAR}_{\mathrm{physical}}\} \subseteq \mathrm{X} \tag{1}$$

where F_θ represents the IceTFT model (θ denotes the trainable parameters in the system). We have experimentally determined the optimal hyperparameters, the size of hidden layers is 160, the bach size is 128, the number of multi-head self-attentive is 4, the dropout rate is 0.1, the max gradient norm is 0.01 and the learning rate is 0.001.$\hat{SIE} \in \mathbb{R}^{1\times N}$ is the prediction result for future N months (N=12). And $X \in \mathbb{R}^{1\times N}$ represents tree types inputs in historical N months (N=12). From Fig.2, the first input type is time static metadata , which is type of input $TIME_{static}$ is time-static metadata calculated by counting the days from the beginning of time. The original TFT IceTFT model is designed to use static covariate encorder encoder to integrate static featuresfrom 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 information so that the static covariate encoder conditions the temporal dynamics through these context vectors , and static enrichment layer enhance and the static enhancement layer enhances these temporal features. The second input is SIE, which is the primary data for predicting prediction in the IceTFT structure model. And the other inputs are several physical variables that are The other inputs $VAR_{physical}$ are various physical variables used to provide atmospheric , and oceanographic features. IceTFT uses an LSTM encoder-decoder to enhance the locality information of these time series. That is This has 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 inputinputs. Finally, in order to skip over extra to skip additional features, the outputs are processed by GRN in a 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 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 and Datasets

As the subject of this research is the prediction of SIE, the historical data itself can provide data features for the 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 (Parkinson et al., 1999). The first dataset of in the model is the monthly SIE, which is provided from provided by the NSIDC Sea Ice Index Version 3 (SII). 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°-90N-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 select a number of variables to aid support the proposed model for SIE prediction in order to help it learn more physical mechanisms and improve forecasting its prediction 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 various dynamical and thermodynamic processes and other factors. Huang et al. (2021) summarized summarizes recent studies and well-known atmospheric processes which connected known atmospheric processes associated with sea ice , and presented and presents the causality graph as Fig.4.

From the study of Huang et al. (2021), the arrows b and c indicate show 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 the downward longwave radiation flux (DLWRF) (Luo et al., 2017). And the 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, solar radiation may be absorbed by the ocean once the surface albedo is significantly reduced by sea ice melting, solar radiation may be absorbed by the ocean, which will further accelerate the further accelerating sea ice melt in late spring and summer (Choi et al., 2014; Kapsch et al., 2016). Kapsch et al. (2016) studied the impact effects of realistic anomalies of in DLWRF and downward shortwave radiation radiative flux (DSWRF) on sea ice by applying a simplified forcing in a coupled climate model (the arrows e and f). AdditionallyIn addition, Liu and Liu (2012) conducted numerical experiments on with the MITgcm model using a reanalysis dataset to demonstrate that changes in surface air temperature and DLWRF have played a significant role in the decline of 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 For

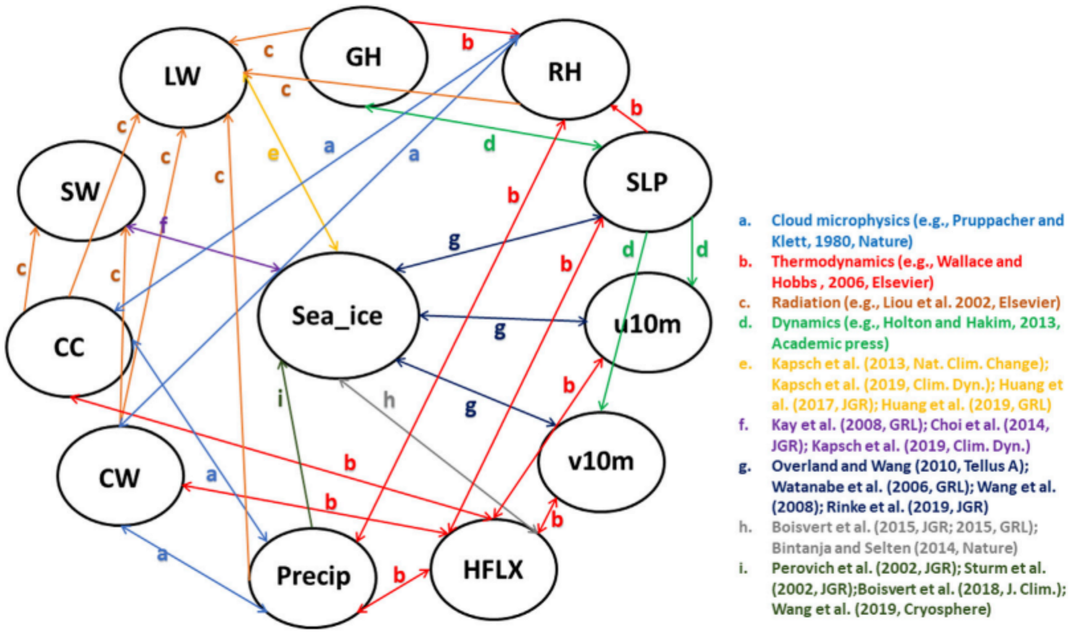

Figure 4. The causality graph , which is derived from the study (Huang et al., 2021) between important atmospheric variables and sea ice over the Arctic. Note that the processes a-d a-d are well-known atmospheric processes , which may that can be outlined in multiple several textbooks. The processes e-i Processes e-i are summarized summaries from recent peer-reviewed publications, and they are the subject of ongoing research. The sea Sea ice here represents sea ice coverage cover and/or sea ice thickness; GH is the geopotential height; RH is relative humidity; SLP means represents sea level pressure; u10m and v10m represent meridional and zonal wind at 10 m, respectively; HFLX is the sensible plus latent heat flux; Precip precip is the total precipitation; CW is the total cloud water path; CC is the total cloud cover; SW and LW represent the net shortwave and longwave flux at the surface, respectively.

our proposed model to learn the atmospheric process, we select these variables: 2m air temperature (AT), DSWRF, DLWRF, and SHUM these variables.

MoreoverIn addition, the snow layer can regulate the growth rate of sea ice due to because of its highly insulating properties, and the accumulation of precipitation on the sea ice pack significantly influences affects the depth of the snow layer (Sturm and Matthew, 2002). And rain can melt, compact, and densify the snow layer, thus reducing the reducing surface albedo and promoting sea ice melting (Perovich and D., 2002). The loss of snow on the ice leads to a significant decrease reduction in surface albedo over the Arctic Ocean, resulting in additional surface ice melt by absorbing at the surface as more solar radiation is absorbed (Screen and Simmonds, 2012). Higher precipitation and snowfall may can lead to a thicker snowpack, which has implications for sea ice changing affects sea ice change (Bintanja and Selten, 2014). Some researchers have studied 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 Arctic river discharge, and the river may river flow could have the positive effect of maintaining thicker ice (Weatherly and Walsh, 1996) (the arrows i). Therefore, we also select precipitation (PRECIP), snow fall snowfall (SF), and river runoff (RUNOFF) to make so that the proposed model can learn these processes.

In addition to atmospheric processes, to To improve the interpretability of the model, we also make it learn some characteristics of ocean ocean features in addition to atmospheric processes. Previous studies demonstrated the effect effects 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 the resurfacing of melt to growth reemergence. Xi et al. (2019) supported that the additional assimilation of SST improves the predicted predictive accuracy of SIE and SIT in the marginal sea ice zonezone of sea ice. 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 in the global region were used as input data for the model. And we calculated the correlation coefficient between these them and SIE. The results are provides shown in Table 1. The variables with the highest correlation coefficient with SIE, as shown in the tableTable, are SST, AT, RUNOFF , and DLWRF, which are boldedin bold. And these variables are all connected to surface evaporation and surface heat in the Arctic hydrological cycle. To make the model learn 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 listed in Table 1.

Table 1. The specifications names, types, sources, units of eleven the 11 physical variables are used as and their correlation coefficients with the input of TFT model.SIE

4 Experiment Scheme

3.1 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 3 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 comparethe performance between 12-month forecast and 6-month forecast.

The settings of different experiments Experiment Training Validation Testing Training Output Slicing Name (Number) (Number) (Number) Length Length MethodIceTFT-align 1982.01~2012.12 (29) 2012.01~2017.12 (4) 2017.01~2021.12 (3) 1212alignIceTFT-general 1982.01~2012.12 (349) 2012.01~2017.12 (49) 2017.01~2021.12 (37) 1212rolling

IceTFT-Pred2019 1982.01~2017.12 (409) 2017.01~2018.12 (All the sources of dataset used in the IceTFT are listed in Table. 1. Except for SST and SF, other data are from National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP-NCAR) Reanalysis 1) 2018.01~2021.12 (25) 1212rolling

(r)2-7IceTFT-Pred2019(6) 1982.01~2017.12 (409) 2017.01~2018.12 ((Kalnay et al., 1996). To explore whether the model depends on the dataset, we also used another reanalysis dataset to compare. We replaced the data from the NCEP-NCAR Reanalysis 1) 2018.01~2021.12 (25) 126rollingIceTFT-Pred2020 1982.01~2018.12 (421) 2018.01~2019.12 (1) 2019.01~2021.12 (13) 1212rolling

(r)2-7IceTFT-Pred2020(6) 1982.01~2018.12 (421) 2018.01~2019.12 (1) 2019.01~2021.12 (13) 126rolling

IceTFT-Pred2021 1982.01~2019.12 (432) 2019.01~2020.12 (1) 2020.01~2021.12 (1) 1212rolling

(r)2-7IceTFT-Pred2021(6) 1982.01~2019.12 (432) 2019.01~2020.12 (1) 2020.01~2021.12 (1) 126rolling

Our experiments include both rolling and alignment slicing methods. The above experiments used a rolling method to generate input. Fig. 5 (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 slice data. While the IceTFT-align experiment uses the align method which is shown in Fig. 5 (b) with Japanese 55-year Reanalysis (JRA-55) (Japan Meteorological Agency, Japan, 2013). The results of correlation coefficients with the JRA55 dataset are similar and we omit to show them in Table. 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 Table3. Two methods for generating inputs1.

3.1 Evaluation

4 Evaluation metrics

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 observation and prediction 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 ith month ordinal of a year. The RMSD is defined as the average distance between simulations predictions 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 predictions 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.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \tilde{y}_i| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2} \tag{3}$$

$$
\begin{aligned}
bais &= (std(y) - std(\tilde{y}))^2 \\
variance &= \frac{1}{n}(y^2 + \tilde{y}^2 - 2y\tilde{y}R) \\
RMSD^2 &= bais + variance
\end{aligned}
\tag{4}
$$

5 The Optimal IceTFT Model

6 Results and Discussions

5.1 Impacts of The Slicing Method of Inputson Predictions

To explore the optimal slicing method of inputs, we used rolling and alignment slicing methods for comparison. Figure 5 (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 the next 24 time-step slice data. The experiment with rolling method named IceTFT-rolling, while the IceTFT-align experiment uses the align method which is shown in Fig. 5 (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.

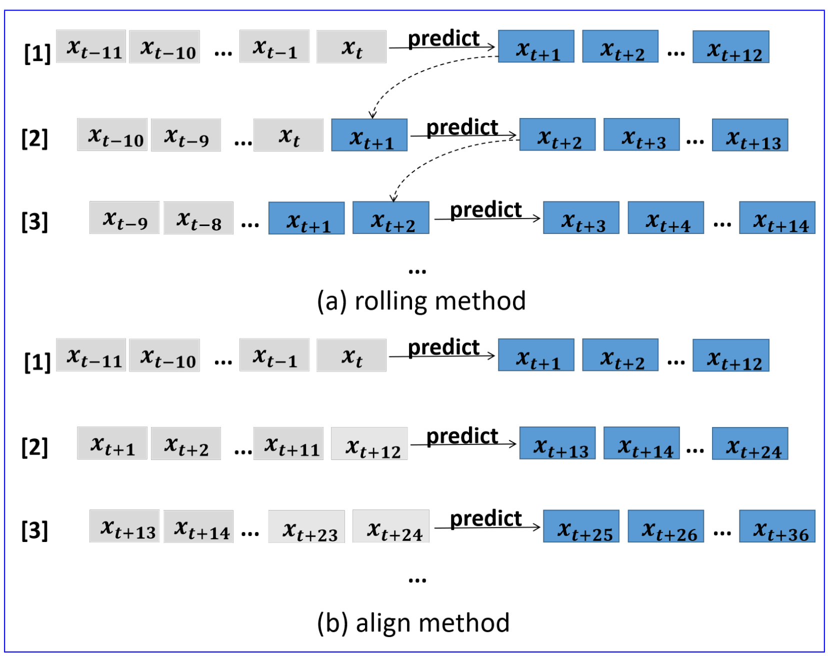

Figure 5. The two slicing methods of inputs

Table. 2 shows the prediction results for IceTFT model with different slicing methods. Compared to IceTFT-align and IceTFT-general, IceTFT-rolling, the IceTFT-align run model 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), from RMSE and RMSD of 2019, but it had a higher error than IceTFT-general run IceTFT-rolling model overall. This may be due to the fact that IceTFT-align run model 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 prediction skills. And it is difficult to predict with high confidence for a model with too little training data. Therefore, the optimal slicing method of inputs is rolling method.

Table 2. The three metrics (MAE, RMSE, RMSD) among the experiments two models with different slicing method methods on SIE prediction predictions during 2019-2021.

Predictive Year	Experiment	2019			2020	
Model Name	MAE	RMSE	RMSD*2019	IceTFT-alignMAE	29.66% RMSE	34.03% RMSD
*2020IceTFT-align	43.30% 0.2966	49.15% **0.3403**	90.40% (r)2-5 **0.7129**	IceTFT-general 0.4330	32.04%(-11.26%) 0.4915	41.62%(-7.53%)0.9040
*2021IceTFT-rolling	IceTFT-align**0.2687**	50.59% 0.3487	60.58% 0.8206	131.04% (r)2-5 **0.3204**	IceTFT-general **0.4162**	44.95%(-5.64%) **0.9023**

5.2 Impacts of The Input Length on Predictions

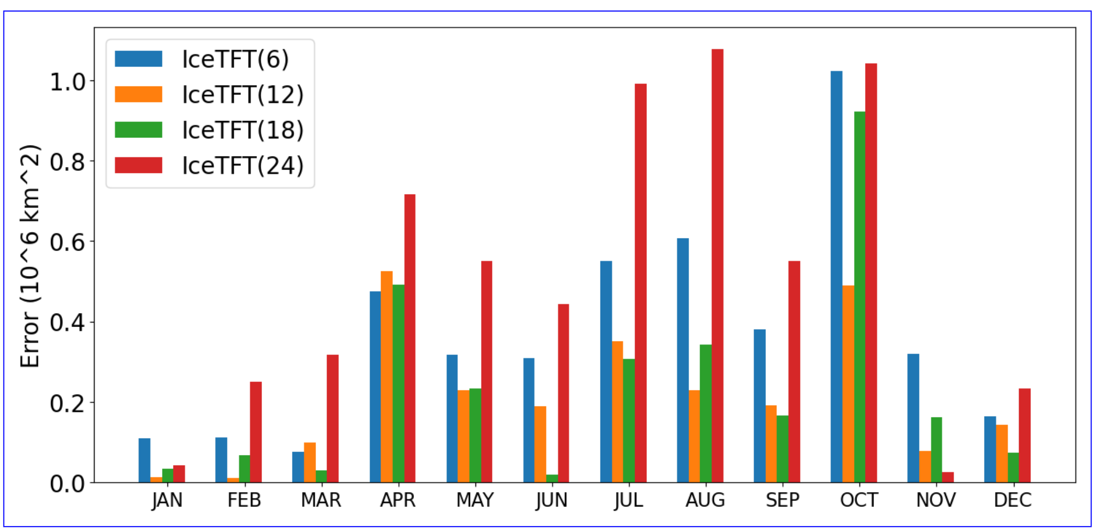

Figure 6. The error between SIE observations and predictions prediction errors of the IceTFT with different input length in lengths for 2019

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. To investigate the effect of input length on the prediction skill, we chose to set up four sets of comparison experiments with input length of 6,12,18 and 24. Using 2019 prediction as an example, the results of the monthly error are set out errors are shown in Fig.6. Surprisingly, the errors of model with a short input length are rather larger than that with 12-monthThe results of 2022-2021 are similar and we omit to show them. As a whole, the prediction errors for the models with the input lengths of 6 and 24 are significantly higher than the results for models with other lengths. 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 time window of 6 is too short to include both the March maximum and the September minimum in each epoch. This may affect the model learning for the properties of extremes, hence features of the extremes, increasing the inaccuracy of the extremes. Therefore, it is not the case that a shorter forecast time will result in a smaller predicted error, However, if the input lengths are too long, the correlation between the recent historical SIE sequence and the future SIE sequence is weakened, increasing the prediction error. In addition, the errors of a model with 18-month are comparable to that with the 12-month, but for the difficult prediction of 2019, i.e., October, which has a large slope, the error of a model with 18-month is significantly higher than that with the 12-month. Therefore, for the monthly prediction of SIE, a reasonable choice for the input length is 12-month, it probably is because the period of SIE is 12-month.

6 The Hindcasting Experiment Results for 2019-2021

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 has more useful features than the early period for recent prediction, so they divided more data from the overall dataset to train. We use all the data before the prediction year for training and testing. For example, the IceTFT-2018 model used to predict SIE from 2019 to 2021, which set from 1982 to 2016 as training data, and from 2017 to 2018 as validation data. And IceTFT-2019 and the input length should be set according to the period of the data. IceTFT-2020 are similar settings, the detail settings are shown on Table. 3.

Table 3. The sets of training, validation and prediction in three models

Model Name	Training	Validation	Predictive Year
IceTFT-2018	1982~2016	2017~2018	2019~2021
IceTFT-2019	1982~2017	2018~2019	2020~2021
IceTFT-2020	1982~2018	2019~2020	2021

Based on the results in Sect. 5, our later models all use a rolling method to slice the inputs and the length of the input is chosen to be 12. In this study, we evaluated the prediction skill of the IceTFT by analyzing the results of the hind-cast experiment results from 2019 to 2021. 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 prediction skill of IceTFT model, while the best predicted results represent the performance of IceTFT model to capture the features of SIE. The results are shown on Table. 4.

6.1 Impacts Performance of Training Setting on PredictionsIceTFT for 12-month SIE predictions

Table 4. The three metrics (MAE, RMSE, RMSD) among three models on SIE predictions for 2019-2021.

Predictive Year		2019			2020			2021		
Model Name		MAE	RMSE	RMSD	MAE	RMSE	RMSD	MAE	RMSE	RMSD
IceTFT-2018	best	**0.1649**	**0.1942**	**0.4554**	-	-	-	-	-	-
	mean	0.2126	0.2668	0.4756	0.3016	0.3808	0.6182	0.1990	**0.2475**	0.5782
IceTFT-2019	best	-	-	-	**0.2007**	**0.2478**	**0.4890**	-	-	-
	mean	-	-	-	0.2847	0.3747	0.5894	0.2545	0.3345	0.7759
IceTFT-2020	best	-	-	-	-	-	-	**0.1684**	0.2677	**0.6689**
	mean	-	-	-	-	-	-	0.2577	0.3018	0.7071

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 Form Table. 4, the models 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 10^6km^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 10^6km^2. Compared to the results of these two years, the errors of the mean predicted results increased in of 2020. This is because there is a second record-low SIE in of September 2020. Moreover, due to the predicted period is being too long relatively, evaluating the forecast prediction 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 12-month 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 the best one.

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 The IceTFT-2019 model had more one year training data than IceTFT-Pred2019 runthe IceTFT-2018 model, and that caused the model to learn some features of recent time. Consequently, IceTFT-Pred2020 run IceTFT-2019 model has lower error than IceTFT-Pred2019 run IceTFT-2018 model for 2020 prediction, and IceTFT-Pred2021 run IceTFT-2020 model has lower error than TFT-Pred2020 run IceTFT-2019 model for 2021. Interestingly, IceTFT-Pred2019 run These results show that the training data that is close to prediction time is more useful than others that are far away. Interestingly, the IceTFT-2018 model also 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 we discussed the reason in Sect. 5.58). 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 generates higher error in some months from RMSE.

6.2 Performance of IceTFT for 12-month SIE predictions

In From the bar graphs in Fig. 7, there is a clear trend of predictions for different years, and it also shows the monthly error between observations and predictionserrors. As can be seen, the predictions of multiple training form a forecast predict 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 to 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 the 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. 7 (c), the errors of IceTFT-Pred2019 run IceTFT-2018 model are smaller than IceTFT-Pred2021 run IceTFT-2020 model in the winter, but higher in the summer. Though the IceTFT-Pred2019 run IceTFT-2018 model has more accuracy than IceTFT-Pred2021 run the IceTFT-2020 model from three metrics, it produces more error in September. As a resultsresult, 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 the 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 predict 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 inaccuracybetween the SIE observations and predictions, the error , we calculated the RMSD between the detrended quarterly SIE observations and predictions over the period the predictions for the 2019-2021 period. The results are shown in Fig. 8. The RMSD ranges from 0.076 to 0.918 million km2 in Fig.8 (a), and the findings from the three years show a wide spread in RMSD on quarter. Figure 8 (b) displays a histogram of the temporal variation in the of squared RMSD, consisting of the “bias” and “variance” following according to 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 It can be seen that there is a very large

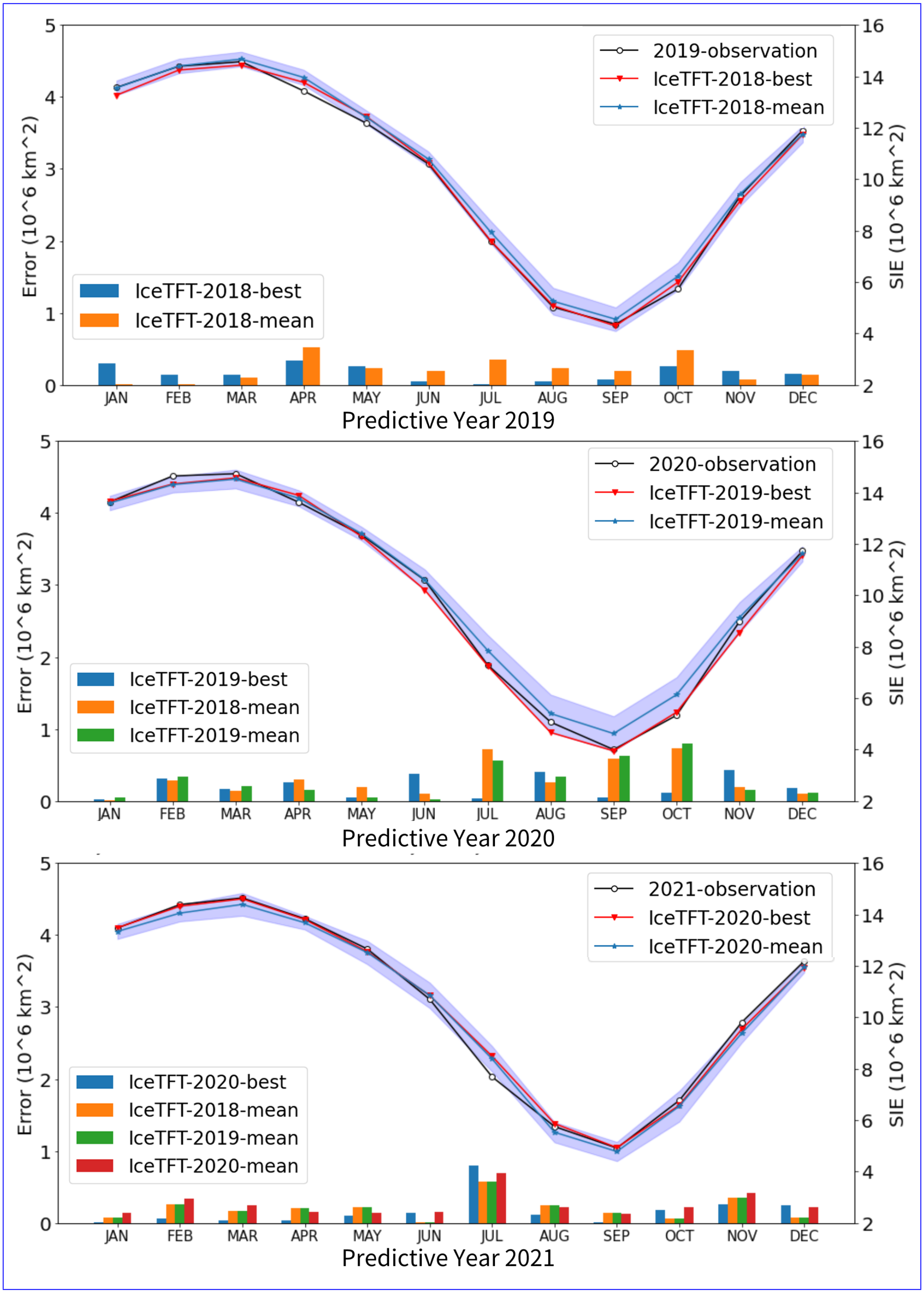

Figure 7. The SIE predictions, observations, and the monthly error between observations and those with different experiment designs in errors for 2019-2021. The line graph represents the observations and SIE predictions, corresponding to the y-axis on the right; the bar graph represents the errors, corresponding to the y-axis on the left.

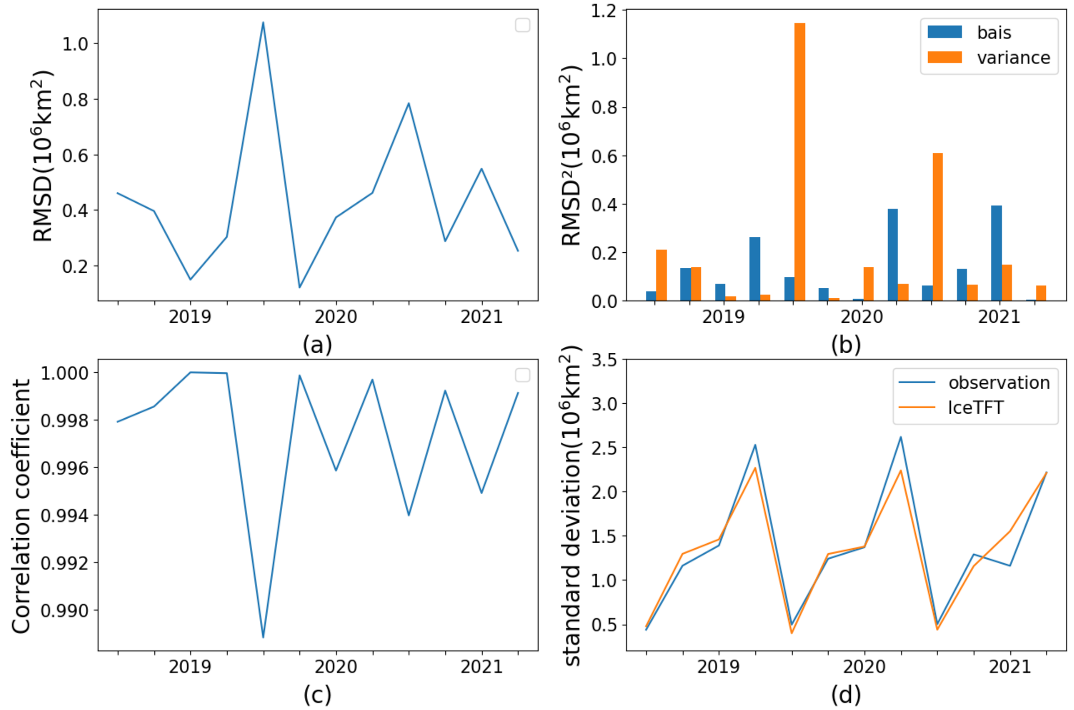

Figure 8. Time series of the RMSD between the detrended quarterly SIE from observations and predictions based on the IceTFT-model over the period 2019-2021: (a) RMSD; (b) squared RMSD (histogram), consisting of “bias” and “variance”; (c) correlation coefficient between predictions and observations; (d) standard deviation of predictions (orange line) and observations (blue line).

variance in the spring (JFM) of 2020 and 2021, which is responsible for the high RMSD in this season. The correlation coefficients in Fig.8 (c) also display an obvious reduction in the spring of spring 2020, which is consistent with the variations of variance variance variations in Fig.8 (b). This result indicates that the significant lower correlation coefficients are partially responsible for the RMSD peak. In addition, Moreover, except for a few months, the magnitude of the bias is substantially larger than the variation in Fig.8 (b), indicating that the change in bias is the main factor for the increase in RMSD. Figure 8(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 increasesis obviously increasing, which contributes to the sharp lager increase in bias during over the same period. Furthermore, this is consistent with the finding in Fig. 7. The IceTFT with large model errors when the SIE trend is more volatile, i.e. when the slope is larger, such as in July and October. The biases are larger for the season containing these two months. This suggests that IceTFT does not fully capture the signals from the historical data and does not reflect the seasonal variability in the SIE. Thus, we can improve the predicted predictive model by focusing on the seasonal variability in the predictions to reduce the RMSD.

6.2 Comparisons with SIO

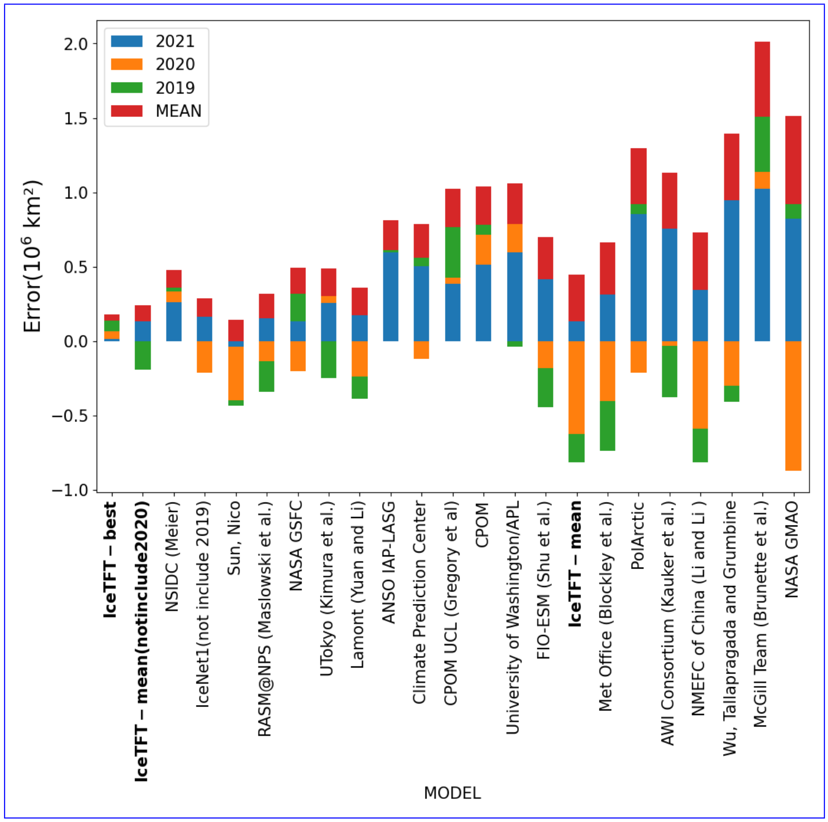

Figure 9. The error between predictions and observations prediction errors of different models in three yearSIO for 2019-2021

In order to evaluate We evaluated the performance of the IceTFT model for SIE minimum prediction , we collected contributions that were provided to SIO over the past three years. Figure September prediction in terms of hindcasting experiments and actual prediction experiments, and collected contributions submitted to SIO in recent years. For hindcasting experiments, Fig. 9 presents the error between the observations and predictions of SIE minimum errors of September in different models from 2019 to 2021. The types and the data used in the SIO models are listed in the Appendix A. Compared to the other models, the IceTFT model has the smallest lowest error in prediction over the three years. For hindcasting experiments, machine learning always results in less error after several training . Therefore, as Machine learning always leads to a lower error after repeated training in hindcasting experiments. As can be seen, the best results of IceTFT model has the IceTFT model have the smallest error, but the error increased a lot in mean results. The mean prediction indicates that the forecast prediction skill of the model is relatively stable. In addition to 2020, the mean prediction of IceTFT model is superior to the 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 To make small error for an anomalous minimum in the mean prediction, the model must have a lower bound on its predictions during the multiple training process. This is challenging to achieve as the model is limited by the historical SIE data. Furthermore, the errors of all models are smaller relatively in to 2019 (Green histogram in Fig.9). The SIE minimum September observation reached its second-lowest value in 2020, and the anomalous caused the errors increased (Orange histogram histograms are longer than bluethe blue ones). While the extremely low anomalies continue to influence the 2021 predictions, the predicted prediction error of most models has increased (see the blue histogram in Fig.9). Even the predicted prediction error is greater than that of 2020. However, 2020, but the IceTFT model is not influenced by anomalies from the previous year, and only focuses focuses only on the physical factors influencing that influence the development of sea ice in that year.

6.3 Impacts of Datasets on Predictions

Table 5. The three metrics (MAE, RMSE, RMSD) among three models with reanalysis datasets of JRA-55 on SIE predictions for 2019-2021. Except for SST and SF, other inputs were replaced with JRA-55.

Predictive Year		2019			2020			2021		
Model Name		MAE	RMSE	RMSD	MAE	RMSE	RMSD	MAE	RMSE	RMSD
IceTFT-2018	best	**0.1681**	**0.2214**	**0.4936**	-	-	-	-	-	-
	mean	0.2891	0.3659	0.7165	0.3616	0.4858	0.8166	0.2255	0.2959	0.6131
IceTFT-2019	best	-	-	-	**0.2676**	**0.3360**	**0.6406**	-	-	-
	mean	-	-	-	0.4434	0.6585	1.0836	0.2130	0.2479	0.5458
IceTFT-2020	best	-	-	-	-	-	-	**0.1428**	**0.1801**	**0.4272**
	mean	-	-	-	-	-	-	0.1951	0.2203	0.4966

To investigate whether the prediction results of IceTFT are affected by the source of input data, we replaced the data from the NCEP-NCAR Reanalysis 1 in Table. 1 with JRA-55. The same experiments were conducted. Different data sources may be associated with different observation errors, but the physical trends embedded in these data are similar. IceTFT model can automatically adjust the weights of the input data during the training process by adaptively learning the features according to the forecast errors. The label data with different errors can affect the prediction error calculated by the IceTFT model and thus have a large impact on the prediction skill. Theoretically speaking, the prediction skill of the IceTFT model is limited by the source of the label data and does not depend on the source of the input data.

However, the results are shown in Table. 5. It can be seen that the best results of the three models are relative to the original results which are from Table. 4, but the mean predictions are higher. This indicates that the models can always get the optimal predictions after several training epochs in the hind-cast experiments and are not limited to the datasets. However, the existence of different observation errors in different datasets makes the bias trends of the predictions different, and therefore makes the mean predictions different. Since the prediction errors using NCEP-NCAR Reanalysis 1 are a little smaller, because in this paper we still use the original dataset for the experimental analysis.

7 The Actual Prediction Results for 2022

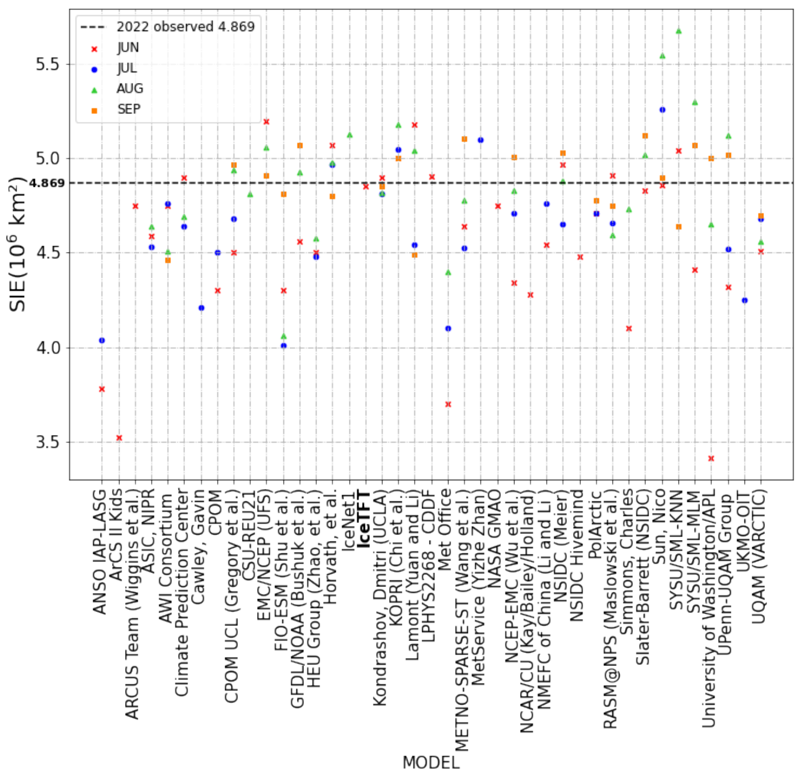

Figure 10. Contributions The predictions of models from SIO for June, July, August and September estimates of September 2022 pan-Arctic sea-ice extentSIE.

Specially, in June, For actual prediction, we submitted to SIO the forecast SIPN the prediction result of the 2022 SIE minimum. September prediction in June 2022. According to the conclusions in Sect. 5.26.1, the closer the training set is to the forecast prediction time, the higher skill the model has in forecastingpredict. So we utilized the result of the IceTFT-Pred2021-best run as the forecast value for train the IceTFT-2021 model to predict 2022. As we only use 12 months of data for 2021to forecast 2022, and the forecast prediction was nine months ahead for September, therefore, we did not submit new forecasts a new prediction in July, August, and September additionally. Figure 10 shows the 2022 SIE predictions of different contributions models of SIO in different lead timetimes. 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 prediction 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 September 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 predictions from other dynamical models, but they only submit predictions with small errors in June and September, with larger 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. 9 and Fig. 10, our proposed model has higher forecasting prediction skills than the other models in both hindcast experiments and real forecastspredictions, and it obtains smaller prediction errors with longer lead times.

7.1 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. **??**. 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. **??** (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. 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. **??** (b) and for 2021 predictions in Table. **??** (c). And for distance year predictions (2021) Table. **??** (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.

8 Interpretability Analysis

8.1 Interpretability AnalysisSensitivity experiments

To investigate the contribution of different variables to SIE prediction in the model, we examined the variable sensitivity on different prediction timetimes, 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 apply this method to compare the contributions of variables. We added The equation of IceTFT can be expanded and simply expressed as Eq. 5, where x_i represents the input variables, w_i is the weight of the corresponding variable, and $\hat{\theta}$ are represents trainable parameter than the weights. We add random Gaussian noises with a zero mean and one standard deviation to each variable in turn. Then we calculated , which can make some change in prediction (Eq. 6). Then we calculate the new RMSE of the model with new inputsand compared , and compare the changes of RMSE (Eq.4).

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

As shown in the Table. 6, we have bolded the values with higher sensitivities.7), where SIE is the observation. 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 forecastingprediction. The particular cases can give us new ideas for improving forecast prediction 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. Due to the existence of multiple variables interacting with each other, it is difficult to analyze their contribution to the prediction. Therefore, this paper only investigates the sensitivity of univariate variables.

$$\mathsf{S\hat{I}E} = \mathsf{F}_{\hat{\theta}}(\mathsf{w_1x_1},\ldots,\mathsf{w_ix_i},\ldots,\mathsf{w_nx_n}) \tag{5}$$

$$\mathsf{S\hat{I}E} + \Delta\mathsf{SIE_i} = \mathsf{F}_{\hat{\theta}}(\mathsf{w_1x_1},\ldots,\mathsf{w_ix_i} + \Delta\mathsf{x_i},\ldots,\mathsf{w_nx_n}), \Delta\mathsf{x_i} \sim \mathsf{N}(\mu,\sigma^2), \mu = 0, \sigma^2 = 1 \tag{6}$$

$$\mathsf{Sensitivity(Var_x)} = \begin{cases} \frac{RMSE(S\hat{I}E+\Delta SIE_i,SIE)}{RMSE(S\hat{I}E,SIE)}, & RMSE(S\hat{I}E+\Delta SIE_i,SIE) > RMSE(S\hat{I}E,SIE) \\ -\frac{RMSE(S\hat{I}E,SIE)}{RMSE(S\hat{I}E+\Delta SIE_i,SIE)}, & RMSE(S\hat{I}E+\Delta SIE_i,SIE) < RMSE(S\hat{I}E,SIE) \end{cases} \tag{7}$$

Table 6. The eleven variable sensitivity of 11 variables among three models for the IceTFT model 2019-2021 in INITIAL the 11var experiment

Model	PredYear	SST	AT	DLWRF	DSWRF	PRECIP	RUNOFF	CSDLF	CSDSF	USWRF	SHUM	SF
IceTFT-2018	2019	**1.193**	**1.208**	**1.485**	**1.319**	**1.157**	-1.002	1.079	**2.416**	**1.219**	1.084	1.091
IceTFT-2019	2020	1.046	1.001	1.073	**1.207**	-1.001	**2.097**	**1.354**	**1.285**	**1.700**	-1.010	-1.011
IceTFT-2020	2021	**2.432**	**2.677**	1.016	**3.548**	1.083	**1.350**	1.077	**2.748**	**2.836**	1.322	1.065

The experiment with 11 variables is noted as 11var, the results are shown in Table. 6. The values with higher sensitivity are in bold. A higher sensitivity value indicates that the variable makes significant contributions to predictions. As for the predictions in three years, a significant contribution to predictions. Multi-variate input of the model may increase training time and uncertainty. We

Table 7. The six variable sensitivity of 6 variables among three models for the IceTFT model 2019-2021 in CTRL the 6var experiment

Model	PredYear	SST	AT	DSWRF	RUNOFF	CSDSF	USWRF
IceTFT-Pred2019IceTFT-2018	2019	1.115	1.078	-1.048	**1.217**	**1.723**	**1.369**
IceTFT-Pred2020IceTFT-2019	2020	**1.285**	1.019	1.098	**1.746**	-1.002	**1.237**
IceTFT-Pred2021IceTFT-2020	2021	**2.226**	1.186	**6.214**	1.301	**3.612**	**3.594**

selected six variables with the highest contributions and redo the same experiments to further investigate the effects of these physical variables on sea ice predictions. These variables include SST, AT, DSWRF, RUNOFF, CSDSF, and USWRF. The experiment with only 6 variables is noted as 6var, and the results of the experiment are shown in Table. 7. To analyze the prediction results after reducing the model inputs, we also calculated the difference of prediction errors between the two experiments. And we plotted the heat map as shown in Table. 8. Negative values are shown in blue, indicating the 6var experiment has a lower error than the 11var. Conversely, positive values are shown in red, indicating a lower error for the 11var experiment with 11 physical factors.

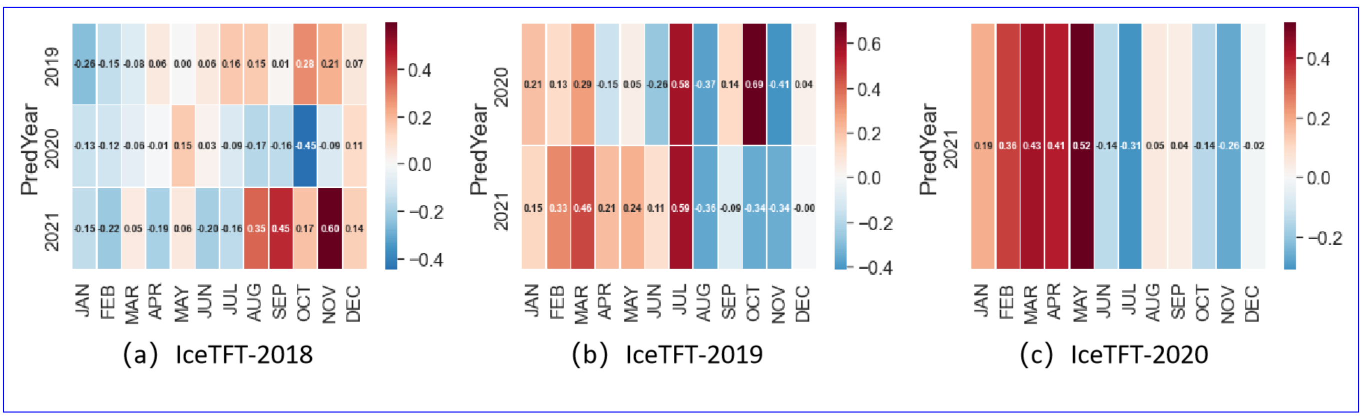

Table 8. The deviation accuracy between the IceTFT model six variables (6var) and the IceTFT with eleven variables (11var) ($6var_{error} - 11var_{error}$), with the heatmap values shown within each grid cell. (a) IceTFT-2018 in the 6var experiment deviation accuracy in three years compared to the 11var experiment. (b) IceTFT-2019 in the 6var experiment deviation accuracy relative to that in the 11var experiment in two years. (c) IceTFT-2020 in the 6var experiment deviation accuracy compared to the 11var experiment.

8.2 Analysis of the physical mechanisms on the years

From the Table. 6, we can see that the sensitivity of the predictors is not exactly the same in different years, and the VSN in the IceTFT can automatically adjust their weights to make the model produce optimal prediction. The variables with a high sensitivity to the predictions are SST, AT, DSWRF, RUNOFF, CSDSF and USWRF. Most of these variables are radiation-relatedrelated to radiation, and shortwave radiation has a greater impact than longwave radiation. This finding is consistent with that other studies in Fig. 4, where surface air temperature and radiation radiative 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. While DLWRF is highly correlated in Table. 1 but has a low sensitivity value in Table. 6, it indicates that this variable is not the cause of the sea ice change, but may be the effect due to other variables. Other studies have shown that latent heat exchange causes more water vapor and clouds to be present in the atmosphere. This enhances the atmospheric greenhouse effect and results in an increased emission of DLWRF. In addition, the increase in water vapor and clouds will lead to more PRECIP. Therefore, there is a correlation between DLWRF and PRECIP, and their sensitivity values change in agreement which both have a higher sensitivity in 2019 and are lower in other years. The positive feedback effect, along with the DLWRF, affects the development of sea ice(Kapsch et al., 2016). Since the machine learning model lacks the partial differential equations of the dynamical model, it cannot simulate the variation of clouds in positive feedback. Therefore, it is difficult to assist in SIE prediction based only on the data trends in DLWRF. While shortwave radiation is influenced by albedo to regulate the effect on sea ice development, the IceTFT model can learn the features of the albedo changes from the historical data. Therefore, the contribution of shortwave radiation in the IceTFT model is larger than that of longwave radiation.

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 are less sensitive to 2020 in our experiments. This provides a new idea for investigating the factors that influence affecting the 2020 anomaly. And it may could be because these 11 variables, which we select, are not the main factor of for the unusually small values produced by of 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 to the Arctic, leading to some impact on the forecast. prediction. Other research has 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.

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 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.

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.

8.3 Analysis of the physical mechanisms on the seasons

From Table. 6 and Table. 7, it can be seen that all six variables had a high sensitivity in INITIAL the 11var experiment, but the sensitivity has changed in CTRL changed in the 6var experiment. For 2019 prediction in IceTFT-Pred2019 runthe IceTFT-2018 model, only two of the six variables were relatively sensitive. This indicates that in CTRL the 6var experiment, the variable selection networks of IceTFT-Pred2019 run IceTFT-2018 model set a greater weight on the CSDSF, USWRF rather than SST and AT. These changes cause the more errors in summer, autumn, and winter, (JJAS), autumn(OND), but fewer errors in springwinter(JFM), 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 has 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 6var experiment has fewer errors almost monthly for 2020 predictions. It is also because that SST and AT have lower sensitivity in CTRL the 6var experiment, and this conclusion is consistent with the INITIAL 11var 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 third 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 on 2019. This also validates the experiment results in Table. 4, explaining that the IceTFT-Pred2019 run IceTFT-2018 model had higher accuracy for 2021 predictions.

For 2020 prediction in IceTFT-Pred2020 runIceTFT-2019 model, some of the selected variables were not sensitive to their predictions in INITIAL the 11var experiment, the high sensitive variables were not fully included. Consequently, their prediction errors have changed significantly in CTRL experimentthe 6var 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 the 6var experiment. According to the first row of Table. 8 (b), these changes cause more errors than gains, especially higher errors in springwinter(JFM), July, and October. This may be due to the fact that the three variables with reduced sensitivity have a greater impact on the 2020 predictions. But For 2021 predictions, the more sensitive SST than INITIAL experiment makes model in CTRL the 11var experiment makes the model in the 6var experiment improve the forecasting skills for 2021 predictions significantly in autumn and winterprediction skills significantly in summer(AS), autumn(OND). Compared with those high sensitivity variables of IceTFT-Pred2021 run in INITIAL IceTFT-2020 model in the 11var experiment from Table. 6, AT and other radiation-related variables are not highly weighted of IceTFT-Pred2020 run in CTRL IceTFT-2019 model in the 6var experiment, increasing 2021 predicted errors in springand summer (AMJ) and winter(JFM) 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 summerwinter.

Similarly, from the results of Table. 6 and Table.7, the variables that significantly contributed to the 2021 predictions in the INITIAL 11var experiment were still selected in the CTRL 6var experiment, but the weights of these sensitive variables have also changed. In particular, the sensitivity of DSWRFbecomes 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, DSWRF, CSDSF and USWRF these radiation-related variables all had high sensitivities for IceTFT-Pred2021 run in CTRL IceTFT-2020 model in the 6var experiment, DSWRF is much more sensitive than the other two variables. In comparison to the INITIAL 11var experiment, these variables had comparable sensitivities. The imbalanced weights led to an increase in predicted errors in CTRL experiment.spring(AMJ) and winter(JFM), which are similar to the 2021 predictions in IceTFT-2019 which are shown in Table.8 (b). This suggests that there is some link between these radiation-related variables that collectively affect forecasting prediction 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 melting 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.

9 Conclusions

In this study, an interpretable long-term forecast predict model for the annual forecast predict 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 variables, and SIE, and sea ice, as inputs to provide relevant mechanisms about the sea ice development process. Moreover, the The IceTFT model can directly predict the next year SIE based on the previous year dataprovide the 12-month SIE directly according to the inputs of the last 12 months, avoiding the need to train multiple models or and the error accumulation of iterative forecastingprediction. In our experiments, we analyze the impact on the forecasts of time periods of the training set, effects on the prediction of the 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 datasets, which improves the accuracy of the prediction. Furthermore, the 12-step 12-month input includes the entire SIE cycle , and its forecast skill is higher than that of the 6-step input modelwhole cycle of SIE, and it is the optimal input length for prediction. We employ the metrics MAE, RMSEand RMSD metrics , and RMSD to evaluate the accuracy of the predictions in the IceTFT model for three years, according to hindcasting experiments from 2019 to 2021. The experimental 2021 and prediction of 2022. The IceTFT model employs the LSTM encoder, a multi-headed attention mechanism, and static time metadata to enhance the learning of the temporal dependence of SIE, so it has high prediction skill on long-term SIE prediction. In hindcasting experiments, the results show that the monthly average forecast error of prediction error of the IceTFT model is less than 0.21 10^6 km^2. For the SIE minimum forecastprediction, compared to other models in SIO that only provide three-month advance forecasts, with a lead time of 1-3 months, the IceTFT model not only has the smallest forecast prediction error, with a three-year average SIE minimum forecast prediction error of less than 0.05 10^6 km^2, but it also provides a nine-month advance forecastsprediction. Moreover, the IceTFT model still has similar higher forecasting skillin the actual forecast of the we submitted the September prediction to SIO in June 2022SIE minimum, and the IceTFT model has similar high prediction skill. Finally, we conducted a sensitivity analysis of the variables in the IceTFT model to investigate the physical factors affecting SIE forecaststhat affect the SIE predictions through the VSN design, which can adjust the weights of inputs. The results indicate that the factors affecting the 2020 SIE forecast prediction are different from those of other years. Except for 2020, for the melt season, SST has a greater influence on SIE forecastspredictions, while for the freeze season, AT and radiation-related variables have a greater influence than SST. These sensitivities can help researchers study investigate the mechanisms of sea ice development, and they also provide useful references for the selection of variables variable selection 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), National Center for Atmospheric Research (NCAR) for JRA-55: Japanese 55-year Reanalysis, Monthly Means and Variances (https://rda.ucar.edu/datasets/) 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).

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).

Appendix A: The Data used in SIO Models

Model	Type	Data
NSIDC (Meier)	Statistical	SIE
IceNet1(not include 2019) *	Machine Learning	climate simulations (CMIP6), OSI-SAF SIC and ERA5
Sun, Nico	Statistical	SIC, CryoSat-2 SIT
RASM@NPS (Maslowski et al.)	Dynamic Mode	NOAA/NCEP CFSv2, CORE2 reanalysis
NASA GSFC	Statistical	SIE, SIC
UTokyo (Kimura et al.)	Statistical	SIC
Lamont (Yuan and Li) *	Statistical	SIC, sea surface temperature (ERSST), surface air temperature,GH300, vector winds at GH300 (NCEP/NCAR reanalysis)
ANSO IAP-LASG *	Dynamic Model	wind components (U and V), Temperature (T) in atmosphere and potential temperatur
Climate Prediction Center	Dynamic Model	SIC, Climate Forecast System Reanalysis (CFSR)
CPOM UCL (Gregory et al)	Statistical	SST(ERA5 reanalysis)
CPOM	Statistical	ice area covered by melt-ponds
University of Washington/APL	Dynamic Model	SIC, CryoSat-2 SIT, SST
FIO-ESM (Shu et al.)	Dynamic Model	SST, sea level anomaly(SLA)
Met Office (Blockley et al.)	Dynamic Model	FOAM/NEMOVAR, MO-NWP/4DVar, SIC
PolArctic	Machine Learning	SIE
AWI Consortium (Kauker et al.) *	Dynamic Model	SIC,CryoSat-2 SIT, NCEP-CFSR, NCEP-CFSv2

Model	Type	Data
NMEFC of China (Li and Li)	Statistical	SIE, SIC
Wu, Tallapragada and Grumbine *	Dynamic Model	NCEP SIC Analysis for the CFSv2, NCEP GFS, GFDL MOM4,Modified GFDL SIS
McGill Team (Brunette et al.) *	Statistical	sea level pressure(SLP), area of ice exported through Fram Strait
ARCUS Team (Wiggins et al.) *	Dynamic Model	CryoSat-2 SIT, SIC and SST (MERRA-2 atmospheric reanalysis)
ASIC, NIPR	Statistical	SIC, ice thickness, ice age, mean ice divergence
ArCS II Kids	Heuristic	SIC
Cawley, Gavin	Statistical	SIE
CSU-REU21 *	Statistical	ERA5, Pan-Arctic Ice Ocean Modeling and Assimilation System(PIOMAS)
EMC/NCEP (UFS)	Dynamic Model	NSIDC NASA Team Analysis
GFDL/NOAA (Bushuk et al.) *	Dynamic Model	towards 3-D temperature, wind, and humidity data (CFSR), OISST
HEU Group (Zhao, et al.)	Statistical	SIC
Horvath, et al.	Statistical	SIE, ERA5
Kondrashov, Dmitri (UCLA)	Statistical	SIE
KOPRI (Chi et al.)	Machine Learning	SIC
LPHYS2268 - CDDF	Statistical	sea ice volume(SIV), SIT
METNO-SPARSE-ST (Wang et al.)	Statistical	SIE
MetService (Yizhe Zhan) *	Statistical	SIE, top-of-atmosphere(TOA), reflected solar radiation (RSR)
NCEP-EMC (Wu et al.)	Dynamic Model	NCEP SIC Analysis for the CFSv2
NCAR/CU (Kay/Bailey/Holland)	Heuristic	Mitch Bushuk GFDL for a synthesis project)
NSIDC Hivemind	Heuristic	SIE

Model	Type	Data
Simmons, Charles *	Statistical	Moana Loa monthly CO2 concentrations, Northern Hemisphere snow area, SIC
Slater-Barrett (NSIDC)	Statistical	SIC
SYSU/SML-KNN	Machine Learning	SIE, SIC
SYSU/SML-MLM *	Statistical	SIC, SST, surface air temperature (SAT), surface net radiation flux (NR)
UPenn-UQAM Group	Statistical	SIE, SIC
UKMO-OIT	Heuristic	-
UQAM (VARCTIC)	Statistical	SIC, SIV

References

Sea Ice Index, Version 3, https://doi.org/10.7265/N5K072F8 , publisher=National Snow and Ice Data Center, author=Fetterer, F., K. Knowles, W. N. Meier, M. Savoie, and A. K. Windnagel., year=2017.

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.

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.

Japan Meteorological Agency, Japan: JRA-55: Japanese 55-year Reanalysis, Monthly Means and Variances, https://doi.org/10.5065/D60G3H5B, 2013.

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

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437 – 472, https://doi.org/https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

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/10.1088/1748-9326/69d0, 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.

Mu, B., Li, J., Yuan, S., Luo, X., and Dai, G.: NAO Index Prediction using LSTM and ConvLSTM Networks Coupled with Discrete Wavelet Transform, in: 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, https://doi.org/10.1109/IJCNN.2019.8851968, 2019.

Mu, B., Qin, B., Yuan, S., and Qin, X.: A Climate Downscaling Deep Learning Model considering the Multiscale Spatial Correlations and Chaos of Meteorological Events, Mathematical Problems in Engineering, 2020, 1–17, https://doi.org/10.1155/2020/7897824, 2020.

Mu, B., Qin, B., and Yuan, S.: ENSO-ASC 1.0.0: ENSO deep learning forecast model with a multivariate air–sea coupler, Geoscientific Model Development, 14, 6977–6999, https://doi.org/10.5194/gmd-14-6977-2021, 2021.

Mu, B., Cui, Y., Yuan, S., and Qin, B.: Simulation, precursor analysis and targeted observation sensitive area identification for two types of ENSO using ENSO-MC v1.0, Geoscientific Model Development, 15, 4105–4127, https://doi.org/10.5194/gmd-15-4105-2022, 2022.

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

Parkinson, C. L., Cavalieri, D. J., Gloersen, P., Zwally, H. J., and Comiso, J. C.: Arctic sea ice extents, areas, and trends, 1978-1996, Journal of Geophysical Research Oceans, 104, 20 837–20 856, 1999.

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.

Polyakova, E. I., Journel, A. G., Polyakov, I. V., and Bhatt, U. S.: Changing relationship between the North Atlantic Oscillation and key North Atlantic climate parameters, Geophysical Research Letters, 33, 1–4, 2006.

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.

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, 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.