MT-IceNet - A Spatial and Multi-Temporal DeepLearning Model for Arctic Sea Ice ForecastingSahara AliUniversity of Maryland Baltimore CountyBaltimore, MD, USAsali9@umbc.eduJianwu WangUniversity of Maryland Baltimore CountyBaltimore, MD, USAjianwu@umbc.eduAbstract—Arctic amplification has altered the climate patternsboth regionally and globally, resulting in more frequent andmore intense extreme weather events in the past few decades.The essential part of Arctic amplification is the unprecedentedsea ice loss as demonstrated by satellite observations. Accuratelyforecasting Arctic sea ice from sub-seasonal to seasonal scaleshas been a major research question with fundamental challengesat play. In addition to physics-based Earth system models,researchers have been applying multiple statistical and machinelearning models for sea ice forecasting. Looking at the potentialof data-driven approaches to study sea ice variations, we proposeMT-IceNet – a UNet-based spatial and multi-temporal (MT)deep learning model for forecasting Arctic sea ice concentration(SIC). The model uses an encoder-decoder architecture withskip connections and processes multi-temporal input streams toregenerate spatial maps at future timesteps. Using bi-monthly andmonthly satellite retrieved sea ice data from NSIDC as well asatmospheric and oceanic variables from ERA5 reanalysis productduring 1979-2021, we show that our proposed model providespromising predictive performance for per-pixel SIC forecastingwith up to 60% decrease in prediction error for a lead time of6 months as compared to its state-of-the-art counterparts.Index Terms—spatiotemporal data mining, neural networks,UNet, sea ice forecasting, climate changeI. INTRODUCTIONThe Arctic is a region with unique climate features. Forinstance, in the Arctic, the Sun never rises over the horizonbecause of which the seasonal variations in polar day andnight are extreme. The enormous areas of Arctic ice andsnow are responsible for reflecting sunlight back to spacewhich keeps the planet cool and regulates global and regionalweather patterns. However, the Arctic sea ice has seen acontinuous decline since 1979 and is left half of which itwas in 1970. Therefore understanding Arctic Amplificationand forecasting sea ice is a key research topic of climatescience. It is important to predict fluctuations in the Arcticsea ice by modeling the weather patterns as it can improve ourunderstanding of potential changes facing the global climate.To study climate change, environmentalists and domainexperts rely greatly on dynamic forecasting systems [12] thatare mainly based on coupled Earth System Models. However,over the last few years, researchers have shifted their focusto data driven Artificial Intelligence (AI) approaches likemachine learning and deep learning. Since the climate datapresents high spatiotemporal correlations, machine learningmodels have shown promising results in spatiotemporal datamining leading to short and long-term weather forecasting.Machine Learning (ML) can provide valuable tools to tackleclimate change. For example, ML approaches can be usedto forecast El-Nino events, hurricanes, and ocean eddies andunderstand the role of greenhouse gases and aerosols onclimate trends and events.Recent works on climate analytics include Convolutionaland Recurrent Neural Network [6] based models and somehybrid modeling approaches like Convolutional LSTM [16],[20] and GraphCNN [4]. However, due to the unique natureof the problem of forecasting Arctic sea ice, there are severallimitations to the existing solutions and multiple challenges.Enlisted below are some of the prevailing challenges inforecasting Arctic sea ice:• Performance versus lead-time trade-off while predictingper-pixel sea ice variations from sub-seasonal (two weeksto three months) to seasonal (three months to two years)scales.• Inability to capture the annual minimum and maximumpeak values of sea-ice in the non-stationary time-seriesdatasets.• Small data problem owing to the availability of only fewdecades worth of observational data.In this paper, we propose a modeling framework, MT-IceNet, to tackle the aforementioned challenges with promis-ing results. Our implementation code can be accessed at theBig Data Analytics Lab GitHub repository1.A. Problem DefinitionIn all the research works conducted (details in Section II),there has not been a one size fits all solution proposed to tacklethe problem of simultaneously detecting, monitoring and pre-dicting sea ice variations. Therefore, in this paper, we proposeMT-IceNet – a fast converging UNet-based [19] spatial andmulti-temporal (MT) regression model for forecasting Arcticsea ice concentration (SIC) at sub-seasonal to seasonal scales.More formally, given:Input1github.com/big-data-lab-umbc/sea-ice-prediction/tree/main/mt-icenetarXiv:2308.04511v1 [physics.ao-ph] 8 Aug 2023– X1i, monthly observational and reanalysis data wherei = [1, 5] with a rolling window of 12 monthly recordsequivalent to one year.– X2i, bi-monthly observational and reanalysis data wherei = [1, 5] with a rolling window of 24 bi-monthly recordsequivalent to one year.MT-IceNet learns from past values of atmospheric and oceansvariables (details in Table I), along with past SIC spatial mapsto forecast:Output– Y , monthly per-pixel Sea Ice Concentration (SIC) valuesat lead time of N months, where N = [1, 6].Here, lead time represents future forecasts of SIC valueswith a lag of one to six months between the input predictorsX and outcome predictand Y.B. ContributionsIn light of the aforementioned background information, themain goal of this research is to develop a spatiotemporaldeep learning model that forecasts Arctic sea ice concentration(SIC) at future months, given spatial data at multiple sub-seasonal scales i.e. bi-monthly (15 days) and monthly levels.Our major contributions are:• We combine reanalysis and observational meteorolog-ical data from multiple sources into two self-curatedspatiotemporal datasets of uniform geographic grid andmulti-temporal resolutions.• We propose MT-IceNet - a spatial and multi-temporaldeep learning model that incorporates a multi-streamlearning approach for multi-temporal data and forecastsea-ice on a monthly seasonal scale of up to 6 months.• We perform a thorough comparative analysis betweenMT-IceNet, baseline models and recently proposed SICprediction models for forecasting Arctic Sea ice at sea-sonal scale.The rest of the paper is organized as follows. Some ofthe important related work is reported in Section II. SectionIII describes the details of our dataset. Our proposed modelis presented in Section IV. Section V provides results andanalysis of our experimental study and comparative analysis.Finally, we conclude our paper and share future directions inSection VI.II. RELATED WORKMajority of the recent works on climate analytics eitherinclude Convolutional and Recurrent Neural Network basedmodels or some hybrid modeling approaches like Convolu-tional LSTM [20] and GraphCNN [4]. [6] proposed a fullydata-driven Long Short-Term Memory (LSTM) model basedapproach for Arctic Sea-ice forecasting and compared it with atraditional statistical model; they found that the LSTM showedgood performance for 1-month sea ice concentration (SIC)prediction, with less than 9 × 106 km2 of average monthlyRoot Mean Square Error (RMSE) and around 11×106 km2 ofmean absolute error during the melting season. [15] developeda 2D-CNN model that takes as input 8 atmospheric predictorsto predict SIC with 1 month's lead time. They compared theperformance with Random Forest baseline model, achievingRMSE of 5.76 × 106 km2. [16] worked on daily predictionof the Arctic Sea Ice Concentration using reanalysis databased on a Convolutional LSTM Network. They proposeda ConvLSTM model to predict SIC for T timestep givenT −1 and T −2 25 km resolution observational data fromNational Snow and Ice Data Center (NSIDC) (2008-2018).They compared their model with a 2DCNN model that takesin a spatial map with pixel grids from T −1 timestep. Theirmodel achieved an RMSE of 11.2 × 106 km2 as compared tothe 2DCNN with RMSE of 13.7 × 106 km2.Ensembling is another hybrid modeling approach whereoutputs from multiple models are combined to improve per-formance, whereas it also reduces variance and generalizationerrors. [14] worked on an MLR + DNN ensemble model usingBayesian Model Averaging to predict sea-ice concentrationsfor the next 10-20 years. They evaluated their model usingcorrelation co-efficient (R2 score) and achieved normalizedRMSE of 0.8. [1] proposed an attention-based LSTM ensem-ble that takes in multi-temporal, daily and monthly, data andpredicts sea ice extent (SIE) for T + 1 timestep, achievingan RMSE of 4.9 × 106 km2. To explore the potential ofprobabilistic modeling approaches for forecasting sea ice andto aid uncertainty quantification, [2] performed a thoroughcomparative analysis of four probabilistic and two baselinemachine learning and deep learning models and publishedbenchmarking results for sea ice forecasting for multiple leadtimes on these models. They evaluated these models perfor-mance using RMSE error and R2 scores and reported GaussianProcess Regression (GPR) to achieve the most competent re-sults. Our work takes inspiration from IceNet proposed by [3].IceNet is a U-Net [19] based probabilistic model for seasonalsea-ice forecasting. Their model takes in images as input andforecasts as output Sea Ice Probabilities (SIP) for three classes(open-water region SIC < 15%, ice-edge region 15% < SIC< 80%, and confident ice region SIC > 80%) for next 6months. Through probabilistic deep learning, they showedtheir forecasted values to be competent with the physics-based ECMWF seasonal forecast system SEAS5 [12]. IceNetis pretrained using Coupled Model Intercomparison Project(CMIP6) 2,220 years (1800-2011) simulation data and is fine-tuned on NSIDC's observational data from 1979 to 2011. Theyevaluated their model performance on observational data from2012-2017 using integrated ice edge error (IIEE) and binaryaccuracy (BACC). Following IceNet, [18] proposed SICNet,based on a Temporal Spatial Attention Module (TSAM) thatcaptures SIC variations for a lead time of 7 to 28 days. Theyevaluated their work using Mean Absolute Error (MAE), MeanAbsolute Percentage Error (MAPE) and BACC. However,there are two major differences in our proposed MT-IceNet,IceNet and SICNet. One, MT-IceNet produces spatial patternsthrough per-pixel prediction for SIC values contrary to SIPclassification of IceNet. Second, MT-IceNet shows promisingresults in the prediction of SIC on greater lead times i.e. 1TABLE IVARIABLES INCLUDED IN THE DATASETVariableSourceUnitsRangeDimensionsFrequencySea Ice ConcentrationNSIDC% per pixel0-100448 x 304dailyLongwave RadiationERA5W/m20-30066 x 360hourlyRain RateERA5mm/day0-80066 x 360dailySnow RateERA5mm/day0-20066 x 360dailySea Surface TemperatureERA5K200-35066 x 360dailyto 6 months whereas SICNet predicts SIC on a weekly i.e.subseasonal scale.III. DATASETFor this study, we use observational sea-ice and reanalysisatmospheric and meteorological data which is available from1979 till the present. The reanalysis data is available with openaccess and can be obtained from European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-5 global reanalysisproduct [8]. Whereas the sea-ice concentration (SIC) valuesare obtained from Nimbus-7 SSMR and DMSP SSM/I-SSMISpassive microwave data version 1 [5] provided by the NationalSnow and Ice Data Center (NSIDC). These variables alongwith their spatiotemporal resolution details are enlisted inTable I.For the Arctic region, the SIC observational dataset containsan uncertainty of about +-15% during the summer season dueto a high number of melt ponds that can skew the data [5].During the winter months, this uncertainty decreases to about+-5% as the sea ice tends to reach its peak in concentrationlevels. However, for modeling purposes, this concentrationdata can be considered as the ground truth.The inclusion of these variables is based on their causallinks with sea ice variations [11] and also based on theirphysical impact on weather trends in the Arctic. For instance,sea surface temperature provides information on oceanic heat.Similarly, earlier rainfalls during spring trigger earlier Arcticice and snow melt [7], [17]. Further, as highlighted by [9],[10], regional differences in atmospheric pressure cause anincrease in Arctic humidity, which in turn enables higher levelsof longwave radiation to reach the sea surface. Consequently,this can lead to earlier melting of sea ice. In short, each ofthe chosen predictors impacts Arctic sea ice through complexoceanic and atmospheric physical interactions.A. Data PreprocessingEach of the chosen data variables come in a different spatialand temporal resolution. For instance, NSIDC provides dailysea-ice concentrations in 25km resolution that is 448×304,whereas the reanalysis data is available in 1 ∘ resolution, i.e.360×180, as hourly and daily records. For our proposedmodel, we required 5 dimensional inputs of shape (samples,timesteps, height, width, features). To achieve this, the firststep after downloading raw data was to regrid individual 1 ∘ ERA5 variables corresponding to the Arctic geolocation of90N, 60N, 180E, 180W into the NSIDC polar projections,that is the 25km spatial (448 × 304) resolution of the sea iceconcentration (SIC). To begin with, an empty array with thelatitude and longitude geolocation index was created. Next, thevalues from previous dimensions are interpolated to the newdimensions and stored in the empty grid with new lat × londimensions using the XESMF Python API. The variablesacquired in hourly data were aggregated to the daily timescale.After the spatial and temporal rescaling was performed onindividual variables, they were combined into a single h5 filewith D x H x W x F dimensions. Here D is the total number ofdays, H is the height of images corresponding to 448 latitude,W represents the width, which corresponds to 304 longitudeand F is the number of features which is 5.1) Monthly Data: To generate the monthly dataset for ourmodel, we averaged the daily 30, 31 or 28 values correspond-ing to the different months. Special care was taken for leapyears, e.g. in case of a leap year, we averaged 29 entriesfor February. This gave us 504 monthly records. Then wesequentially divided the data into training and testing sets of408 and 96 months. To reshape the data, a stateless rollingwindow was applied to the training and testing data, creating384 samples of 12 months each. Sample one contained months1-12, sample two contained months 2-13, and the last samplecontained months 372-384. Finally we got our training data inthe shape M × T × H × W × F, where M = 384 samples,T = 12 months, H × W =

448×304 pixel images andF = 5 features. Similarly, the final shape of the test set was84 × 12 × 448 × 304 × 5.2) Bi-Monthly Data: Deep learning models require a largevolume of diverse training data to generalize well on theunseen test data. However, our monthly dataset is comprisedof only 504 records. To counter this small data problem, wegenerated the second temporal resolution of semi-monthly orbi-monthly data that not only increases the dataset size butalso helps our model focus on sub-seasonal patterns. From ourprevious work [1], we observed high frequency data capturessub-seasonal fluctuations better and in turn helps the modellearn the seasonal patterns. For this, similar to aggregating 30or 31 daily values, we aggregated samples of 15, 16 or 14daily records depending on the annual months of the year.For example, for January the two bi-monthly records werecalculated by taking the average of 15 days and 16 daysrespectively. Similar to the rolling window applied to monthlydata, we applied a 24 timestep rolling window to bi-monthlydata to correspond to the same annual cycle as its monthlycounterpart. The final dimensions for bi-monthly training andtest sets were 384×24×448×304×5 and 84×24×448×304×5Fig. 1. End-to-end pipeline of our predictive model.respectively.IV. METHOD: MT-ICENETOur proposed MT-IceNet model is a UNet-based spatial andmulti-temporal (MT) deep learning model for forecasting per-pixel Arctic sea ice concentrations (SIC). The model uses anencoder-decoder architecture with skip connections and multi-temporal data to regenerate spatial maps at future timesteps.As shown in Figure II, we started off by downloading the rawdata from multiple sources mentioned in Section III. Next,we preprocessed the data to bring it into uniform spatialand temporal resolutions. We then reshaped the data into 5dimension and sequentially split it into training and testingsets. Sequential splitting is performed to retain the seasonalitypatterns in the data. We then built our baseline and proposedmodel MT-IceNet and finally evaluated the performance of allmodels using the Root Mean Squared Error (RMSE), MeanAbsolute Error (MAE) and R-squared (R2) score. The detailsof our baseline models and constituent blocks of our proposedmodel are as follows.A. Baseline ModelsTo design our baseline models, we utilized two widely usedspatiotemporal deep learning techniques that are, the Convolu-tional Neural Network (CNN) and Convolutional Long ShortTerm Memory (ConvLSTM) models. One reason for choosingthese two models is that they have been used in previoussolutions proposed for this problem. Another reason is thatthey also work as the constituent blocks of our proposed modelMT-IceNet.1) Convolutional Neural Network (CNN): We designed asimple CNN model with three 2D convolutional layers usingthe Keras API. Each of these layers was followed by a 2DMax Pooling layer for dimensionality reduction. We flattenedthe output from the third CNN layer and appended two fullyconnected (Dense) layers for regression. Finally, we reshapedthe output from the final Dense layer into 448 × 304 to retrievethe predicted spatial maps. The input to this model were mini-batches of 3D tensors of shape H × W × F correspondingto 448 × 304 × 5 dimensional images whereas the outputwas monthly SIC percentage values in the shape of 448 ×304 images for multiple lead times. To incorporate the leadtime in monthly forward predictions, a lag (offset) of 1 to 6months was created in input features and target SIC values byremoving the first 1 to 6 rows from the SIC column and last 1to 6 rows from the 5 input features. So that a January 1979 datasample would correspond to February 1979 SIC values for alag of 1 month, a January 1979 data sample would correspondto March 1979 SIC values for a lag of 2 months, and so on.The model was trained six times, each time for a different lagvalue.2) Convolutional Long Short Term Memory (ConvLSTM):We further designed a ConvLSTM model by appending oneConvLSTM layer at the beginning of our baseline CNNmodel. Since ConvLSTM model requires 4D input tensors,we used the same input train and test datasets generated forour proposed MT-IceNet model for training this ConvLSTMbaseline model, that is, T ×H ×W ×F. This model was alsodeveloped using the Keras API. We trained four ConvLSTMmodels using the monthly dataset for lead times of 1 to 6months. To incorporate the lag, an offset was added to trainand test sets in a similar manner as done for our CNN baselinemodel.B. MT-IceNet ModelOur proposed Multitemporal (MT)-IceNet is a U-Net basedmodel that comprises two paths of neural network layers,the contractive path that we represent as encoder, and theexpansive path represented as the decoder. The architecturediagram is shown in Figure 2.1) U-Net: A U-Net based model was first introduced forimage segmentation for bio-medical imagery [19]. It comprisesthree constituent blocks; the encoder, the decoder and thebottleneck block that acts as a bridge between both the encoderand decoder. What distinguishes a U-Net architecture froma transformer based model is the use of skip connectionsbetween different layers of encoder and decoder. These skipconnections provide upsampling layers important featuresFig. 2. Model Architecture of the Multi-temporal (MT) IceNet Model.from the downsampling layers that are lost due to the depthof the network.2) Encoder: Our encoder comprises two downsamplingblocks. The first block consists of a ConvLSTM layer thattakes in monthly data as input in the shape T × H × W × F,here T = 12. The output of ConvLSTM layers is passed totwo 2D convolution layers, followed by a batch normalizationlayer and a 2D max pooling layer. The second block followsthe same architecture with the difference of input shape. Here,the ConvLSTM layer if given bi-monthly data of the shape24 × H × W × F. In every successive layer of the encoder,we increment the output channels by a multiplicative factor of2, as shown in Figure 2. All CNN layers use the same 3 × 3kernel size filters whereas the ConvLSTM layer uses a 5 × 5kernel. The activation function ReLU is used in all the encoderlayers. The encoder part of our model helps learn low-levelspatiotemporal dependencies in the data and identifies patternsneeded for predicting SIC spatial maps.3) Decoder: The purpose of the decoder block is to up-sample the low-level features learnt from the data and helpreconstruct the spatial map in the same dimension as the inputbut at a future timestep. Similar to the encoder, the decodercomprises two upsampling blocks. Every block comprises a2×2 upsampling layer using the nearest interpolation methodand a 2 × 2 kernel size filter. The skip connection is builtby concatenating the output of each upsampling layer withthe output from a corresponding downsampled feature mapgenerated by the encoder, as shown in the Figure 2. Oncethe outputs are concatenated, they are passed through two2D convolutional layers. The output channel size of everyCNN layer is reduced by a factor of 2 in order to regain theinitial input dimension. Finally, a 1×1 convolution with linearactivation is applied to the decoder's output to generate thepredicted spatial map.C. PostprocessingWe performed two post-processing steps on the predictionsgenerated by our model. Since our predictions correspondto sea ice concentration values that are basically percentagevalues between 0 to 100, we rescale the values predicted bythe model to [0,100] by clipping all predictions less than 0 to 0and all predictions greater than 100 to 100. This helps interpretthe regression results. Further, to help visualize the predictions,we multiplied the predicted spatial maps with a binary landmask. Since we are not interested in land-area predictions,this multiplicative step discards the land area predictions byassigning them zero-weightage while all ocean and water bodypredictions are retained. This also helps in evaluating themodel using the evaluation metrics discussed in Section V.V. EVALUATIONWe first present the experimental setup for our researchwork in Section V.A. We then move forward to compare ourperformance with the baseline CNN and ConvLSTM models.We also compared our work with two recently proposedsolutions to SIC forecasting using 1) multitask-ConvLSTMmethod [13] and 2) IceNet [3]. We present the results of thiscomparative analysis in Section V.B. Finally we perform thequalitative analysis of our MT-IceNet predictions in SectionV.C.A. Experimental SetupAll our experiments are performed using the Amazon WebServices (AWS) cloud-based Elastic Compute Cloud (EC2)accelerated computing instances with high frequency 2.5 GHz(base) Intel Xeon Scalable Processor, 32 vCPUs and 64 GBsof GPU memory. The total storage space required for ourexperiments is around 600 GBs which includes our trainand test datasets, model computations and storage and visualillustrations of our results. Our MT-IceNet model is trainedusing Keras Functional API with a Tensorflow backend andhas around 148,000 trainable parameters. This is 99% lessthan the 44 million trainable weights of the IceNet [3] model.Through a less complex architecture, we also show how simpleapproaches can generate better results.We trained our model using Adam optimizer, Root MeanSquared Error (RMSE) loss and trained it on 100 epochs usingthe Early stopping criteria with a learning rate of 0.0001. Dueto the high dimensionality of input data and a limited RAM,we could only process mini-batches of 4 samples each.Evaluation Metrics: We report the RMSE, MAE and R2performance evaluation scores for our model in Tables II, IIIIV. Since it is a spatiotemporal 3D dataset corresponding tolatitude and longitude values, we customized the RMSE andMAE metrics for our models evaluation using the followingformula:RMSESIC =vuutΣIΣJY [i, j] −ˆY [i, j]2N(1)MAESIC =ΣIΣJY [i, j] −ˆY [i, j]N(2)Here, Y represents ground truth while ˆY represents predictedSIC values. i corresponds to 448 latitude, j corresponds 304longitude values and N represents the total number of testsamples. While both, RMSE and MAE, are metrics to calculateerror, RMSE gives relatively higher weightage to large errorand can help in capturing the variance in error magnitudes.R2 = 1 −RSSTSS(3)We further evaluated our model using the R2 score. As shownin Eq. 3, RSS represents the sum of squares of residuals andTSS represents the total sum of squares. Higher R2 scorerepresents better performance.B. Comparative AnalysisFor the comparative analysis, we first trained the baselinemodels CNN and ConvLSTM to predict SIC values for alead time of 1 to 6 months. We then trained the Multitask-ConvLSTM and IceNet on our dataset to predict SIC valuesfor the same lead times. We are grateful to the authorsfor providing open-access to their codes on github. Since,IceNet is a computationally expensive classification model thattakes in 50 input features and requires 1TB of memory, wecustomized their models to a light-weight version comprisingof only 2 downscaling and 2 upscaling blocks. We furthertweaked their output layer by removing the classification layerand replacing it with a regression layer to generate 448 × 304spatial maps at multiple lead times. We refer to this modifiedversion as IceNet†. The multitask-ConvLSTM was proposedto jointly predict per-pixel SIC values and the total sea-iceextent corresponding to the entire spatial region. All thesemodels were trained and evaluated using the same train andtest split to have a fair comparison of their performance. Inour comparison with other models, we only took into accountthe SIC prediction and ignored the sea-ice extent values. Wefirst analyzed the performance of baseline models for SICprediction and compared it with our MT-IceNet predictions.1) Quantitative Analysis: In Tables II and IV, it is evidentby increasing values of RMSE and MAE scores that bothCNN and ConvLSTM have poor predictive performance ascompared to MT-IceNet, where our model reportedly de-creased the RMSE error by 51% as compared to CNN, by40% as compared to ConvLSTM and by 58% as comparedto IceNet†, for all lead times. The same trend was observedin MAE error where MT-IceNet reduced the MAE error bymore than 60% as compared to its Multi-task ConvLSTM andIceNet counterparts. We further noticed that MT-IceNet has asignificantly better R2 score with a notable lead of around 10%for all lead times as compared to the CNN and ConvLSTMmodels. The best results have been highlighted in bold in allour result tables.As seen in Table II, both multitask-ConvLSTM and IceNet†take a sharp increase in RMSE score after a lead time time ofone month, whereas MT-IceNet still shows a trivial increasein the RMSE scores with only a 2 point increase in RMSEand 1 point increase in MAE from lead time of 1 month tothe sixth month. To our surprise, the highest reported errorsare from IceNet† model where the RMSE and MAE errorshave significantly higher and R2 scores have very low valuesfor the IceNet† predictions. It is evident from all three metricresults that MT-IceNet outperforms all baseline and recentlyproposed models for SIC forecasting by showing a promisingand persistent predictive performance on greater lead times.The second best performance is achieved by the baselineConvLSTM model. An interesting observation here is that allFig. 3. NSIDC Observed Sea Ice Concentration (%) vs MT-IceNet Predictions for Summer 2020.Fig. 4. MT-IceNet Summer Prediction difference plots for multiple lead times.Fig. 5. NSIDC Observed Sea Ice Concentration (%) vs MT-IceNet Predictions for Winter 2020.Fig. 6. MT-IceNet Winter Prediction difference plots for multiple lead times.Fig. 7. Time-series for derived Sea Ice Extent from MT-IceNet predictions at multiple lead times.TABLE IIRMSE SCORES (IN %) FOR SIC PREDICTION FROM MULTIPLE MODELSModelOne month lagTwo months lagThree months lagFour months lagFive months lagSix months

lagCNN11.3411.7512.5112.7113.0312.72ConvLSTM9.129.4511.0011.0710.269.89Multi-task ConvLSTM

[13]11.7312.0212.6115.7913.1212.86IceNet†13.0717.4217.1418.4321.3718.20MT-IceNet5.506.737.617.967.877.77TABLE IIIR2 SCORES FOR SIC PREDICTION FROM MULTIPLE MODELSModelOne month lagTwo months lagThree months lagFour months lagFive months lagSix months

lagCNN0.8380.8260.8020.7960.7860.796ConvLSTM0.8920.8850.8390.8400.8600.870Multi-task ConvLSTM

[13]0.8230.8130.7930.6730.7730.781IceNet†0.7770.6040.6160.5560.4030.567MT-IceNet0.9580.9290.9190.9110.9130.915TABLE IVMAE SCORES (IN %) FOR SIC PREDICTION FROM MULTIPLE MODELSModelOne month lagTwo months lagThree months

lagFour months lagFive months lagSix months lagCNN3.1633.3573.4703.7033.6554.142ConvLSTM2.5692.8383.1783.2413.0413.219Multi-task ConvLSTM [13]3.6653.6223.9325.5914.3054.267IceNet†8.5648.7498.4489.53612.2779.157MT-IceNet1.3131.9831.9622.1642.1541.967models show an improvement in performance after the leadtime of 4 months, as evident by the RMSE, R2 and MAEscores. Though this is an interesting finding, the actual causeof this performance improvement is yet to be known.2) Qualitative Analysis: To evaluate the quality of our per-pixel predictions, we plot the spatial maps generated by theMT-IceNet model over the Arctic region using Python's car-topy API for geospatial projections. Figures 3 and 5 show theforecasted spatial plots where every pixel value lies between[0,100]. Here 100 represents 100% ice concentration whereas0 represents the absence of ice in that specific pixel. Sinceeach pixel corresponds to a 25 × 25 km2 land area, anyvalue ranging in between 0 to 100 represents the percentagearea covered with ice in that region. Looking at Figure 3,it is observed that MT-IceNet overpredicts September seaice at greater lead times which is the trickiest to predict.Nonetheless, our model shows great performance throughoutMarch predictions which is the peak Winter time in the Arctic,as shown in Figure 5.To have a clear identification of regions with incorrect pre-dictions, we plot the differences in the actual SIC observationsand the predicted values for multiple lead times, both forSummer and Winter peak months, i.e., September and March,as shown in Figures 4 and 6. Upon inspecting Figure 4, wesee that model performs poorly only near the coastal areas ofGreenland. For March, we notice that the model underpredictsthe sea ice over the coastal areas. This can be considered aminor performance flaw as edge predictions are usually thetrickiest for spatiotemporal models. For September predictions,as shown in Figure 4, we notice how model overpredictsSummer sea ice at greater lead times. This is due to the conceptof seasonal barrier, according to which the seasonality patternsare hard to identify from a distance of more than 3 months.Using the SIC values, we calculated the overall sea ice extentfor the entire region by calculating the area-weighted sum ofthe Arctic region using the per-pixel area map provided byNSIDC. We plotted these sea ice extent values as a time-series plot for multiple lead times, as shown in Figure 7.We noticed how our model overpredicts summer sea ice andunderpredicts winter sea ice at greater lead times. We alsonoticed the performance improvement in lead times 5 (red) and6 (lime) where the model predictions once again come closerto the actual observations (blue). Overall, we did not find anysharp increase or decrease in the SIC model predictions as thelead time increases. This means our model can overcome theperformance versus lead-time tradeoff that is faced by most ofthe models proposed for seasonal predictions.VI. CONCLUSIONS & FUTURE WORKIn this paper, we presented our work on a spatiotemporaldeep learning model that jointly learns from multi-temporalinputs to forecast Arctic sea ice at lead times of 1 to 6 months.Through experiment and ablation study, we showed how ourmodel outperforms the baseline and recent state-of-the-artapproaches using a U-Net based architecture by overcomingthe small data problem and seasonality barrier challenge. OurMT-IceNet not only outperforms the baseline and other recentwork but also shows a consistency in forecasting SIC valuesat greater lead times. We believe our proposed model cansubstantially improve our ability in predicting the future Arcticsea ice changes at sub-seasonal to seasonal scales, which isfundamental for forecasting transportation routes, length ofopen water, resource development, coastal erosion, and threatsto Arctic coastal communities and wildlife.In the future, we plan to extend our work to multi-scalespatiotemporal modeling in order to jointly process fine andcoarse resolutions of geolocation information that can be vitalin solving similar Earth Science problems. We further plan toincorporate the attention mechanism in our model to identifyimportant contributing factors to the prediction. Lastly, we planto work on data-driven causal discovery to study variations inArctic sea ice using spatiotemporal deep learning models.ACKNOWLEDGEMENTThis work is supported by NSF grants: CAREER: Big DataClimate Causality (OAC-1942714) and HDR Institute: HARP- Harnessing Data and Model Revolution in the Polar Regions(OAC-2118285). We thank Dr. Yiyi Huang (NASA LangleyResearch Lab) for her assistance in introducing the dataset.REFERENCES[1] S. Ali, Y. Huang, X. Huang, and J. Wang. Sea Ice Forecasting usingAttention-based Ensemble LSTM. Tackling Climate Change with Ma-chine Learning Workshop at ICML. arXiv preprint arXiv:2108.00853,2021.[2] S. Ali, S. A. Mostafa, X. Li, S. Khanjani, J. Wang, J. Foulds, andV. Janeja.Benchmarking probabilistic machine learning models forarctic sea ice forecasting. In IGARSS 2022-2022 IEEE InternationalGeoscience and Remote Sensing Symposium, pages 4654–4657. IEEE,2022.[3] T. R. Andersson, J. S. Hosking, M. P´erez-Ortiz, B. Paige, A. Elliott,C. Russell, S. Law, D. C. Jones, J. Wilkinson, T. Phillips, et al. Seasonalarctic sea ice forecasting with probabilistic deep learning.NatureCommunications, 12(1):1–12, 2021.[4] S. R. Cachay, E. Erickson, A. F. C. Bucker, E. Pokropek, W. Potosnak,S. Osei, and B. L¨utjens. Graph Neural Networks for Improved El Ni˜noForecasting. arXiv preprint arXiv:2012.01598, 2020.[5] D. Cavalieri, C. Parkinson, P. Gloersen, and H. Zwally. Sea Ice Con-centrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS PassiveMicrowave Data, Version 1.Technical report, NASA DAAC at theNational Snow and Ice Data Center, 1996.[6] J. Chi and H.-c. Kim. Prediction of arctic sea ice concentration using afully data driven deep neural network. Remote Sensing, 9(12), 2017.[7] T. Dou, C. Xiao, J. Liu, W. Han, Z. Du, A. Mahoney, J. Jones, andH. Eicken. A key factor initiating surface ablation of arctic sea ice:Earlier and increasing liquid precipitation. The Cryosphere, 13:1233–1246, 04 2019.[8] European Centre for Medium-Range Weather Forecasts. ERA-5 globalreanalysis product.https://cds.climate.copernicus.eu/cdsapp#!/home,2021. Accessed: 2021-9-5.[9] S. Horvath, J. Stroeve, B. Rajagopalan, and A. Jahn.Arctic sea icemelt onset favored by an atmospheric pressure pattern reminiscent ofthe north american-eurasian arctic pattern. Climate Dynamics, 04 2021.[10] Y. Huang, X. Dong, B. Xi, and Y. Deng. A survey of the atmosphericphysical processes key to the onset of arctic sea ice melt in spring.Climate Dynamics, 52(7):4907–4922, 2019.[11] Y. Huang, M. Kleindessner, A. Munishkin, D. Varshney, P. Guo, andJ. Wang. Benchmarking of data-driven causality discovery approachesin the interactions of arctic sea ice and atmosphere. Frontiers in BigData, 4:72, 2021.[12] S. J. Johnson, T. N. Stockdale, L. Ferranti, M. A. Balmaseda, F. Molteni,L. Magnusson, S. Tietsche, D. Decremer, A. Weisheimer, G. Balsamo,et al. Seas5: the new ecmwf seasonal forecast system. GeoscientificModel Development, 12(3):1087–1117, 2019.[13] E. Kim, P. Kruse, S. Lama, J. Bourne, M. Hu, S. Ali, Y. Huang, andJ. Wang. Multi-task deep learning based spatiotemporal arctic sea iceforecasting. In 2021 IEEE International Conference on Big Data (BigData), pages 1847–1857. IEEE, 2021.[14] J. Kim, K. Kim, J. Cho, Y. Q. Kang, H.-J. Yoon, and Y.-W. Lee. Satellite-based prediction of arctic sea ice concentration using a deep neuralnetwork with multi-model ensemble. Remote Sensing, 11(1):19, 2019.[15] Y. J. Kim, H.-C. Kim, D. Han, S. Lee, and J. Im. Prediction of monthlyArctic sea ice concentrations using satellite and reanalysis data basedon convolutional neural networks. The Cryosphere, 14(3):1083–1104,2020.[16] Q. Liu, R. Zhang, Y. Wang, H. Yan, and M. Hong. Daily predictionof the arctic sea ice concentration using reanalysis data based on aconvolutional lstm network. Journal of Marine Science and Engineering,9(3):330, 2021.[17] A. Marcovecchio, A. Behrangi, X. Dong, B. Xi, and Y. Huang. Precip-itation influence on and response to early and late arctic sea ice meltonset during melt season. International Journal of Climatology, n/a(n/a),2021.[18] Y. Ren, X. Li, and W. Zhang. A data-driven deep learning model forweekly sea ice concentration prediction of the pan-arctic during themelting season. IEEE Transactions on Geoscience and Remote Sensing,2022.[19] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networksfor biomedical image segmentation.In International Conference onMedical image computing and computer-assisted intervention, pages234–241. Springer, 2015.[20] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c.Woo. Convolutional lstm network: A machine learning approach forprecipitation nowcasting. In Advances in neural information processingsystems, pages 802–810, 2015.