SYDE 675: PROJECT

PAPER: SEA ICE FORECASTING USING ATTENTION-BASED ENSEMBLE LSTM

2022-04-18

Sea Ice Forecasting using Attention-based Ensemble LSTM

Sahara Ali [1] Yiyi Huang [1] Xin Huang [1] Jianwu Wang [1]

Abstract

Accurately forecasting Arctic sea ice from sub-seasonal to seasonal scales has been a major scientific effort with fundamental challenges at play. In addition to physics-based earth system models, researchers have been applying multiple statistical and machine learning models for sea ice forecasting. Looking at the potential of data-driven sea ice forecasting, we propose an attention-based Long Short Term Memory (LSTM) ensemble method to predict monthly sea ice extent up to 1 month ahead. Using daily and monthly satellite retrieved

trends, the Arctic ocean could be sea ice free by 2050 (Notz & Stroeve, 2018). Such rapid changes has profound local and global impacts on transporting routes, resource development, coastal erosion, military and civilian infrastructure, Arctic coastal communities (hunting and transportation by indigenous populations), wildlife (e.g., polar bear access to food sources).

Though most of the scientists agree that this rapid Arctic warming is a sign of human-caused climate change, studying the causes of Arctic amplification and forecasting sea ice has become one of the most hyped questions in the Earth Science research (Holland et al., 2019). Current operational

Zacharie Gousseau & Nishad Rajmalwar
Systems Design Engineering

UNIVERSITY OF WATERLOO | FACULTY OF ENGINEERING

Presentation outline

1. Overview of the original paper
2. Compare with ML methods & statistical baselines
3. Investigating improvements

OVERVIEW: ALI ET AL.

Sea ice extent prediction using attention-based LSTMs

Problem statement

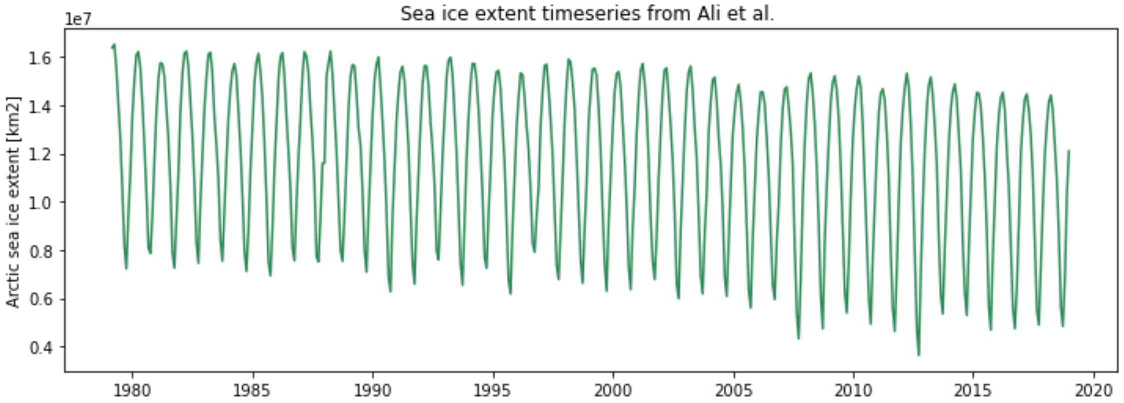
- Goal: Predict arctic sea ice extent (SIE) with a 1-month lead time
  - i.e., the average SIE for the next month
- Input variables sourced from the ERA5 global reanalysis product, and averaged over everything North of 25°N
- Target variable (SIE) calculated from NSIDC's sea ice concentration product (from Nimbus-7 passive microwave data)

Input variables

VARIABLE	RANGE	UNIT
SURFACE PRESSURE	[400,1100]	HPA
WIND VELOCITY	[0,40]	M/S
SPECIFIC HUMIDITY	[0,0.1]	KG/KG
AIR TEMPERATURE	[200,350]	K
SHORTWAVE RADIATION	[0,1500]	W/m^2
LONGWAVE RADIATION	[0,700]	W/m^2
RAIN RATE	[0,800]	MM/DAY
SNOWFALL RATE	[0,200]	MM/DAY
SEA SURFACE TEMPERATURE	[200,350]	K
SEA SURFACE SALINITY	[0,50]	PSU
SEA ICE CONCENTRATION	[0, 100]	%

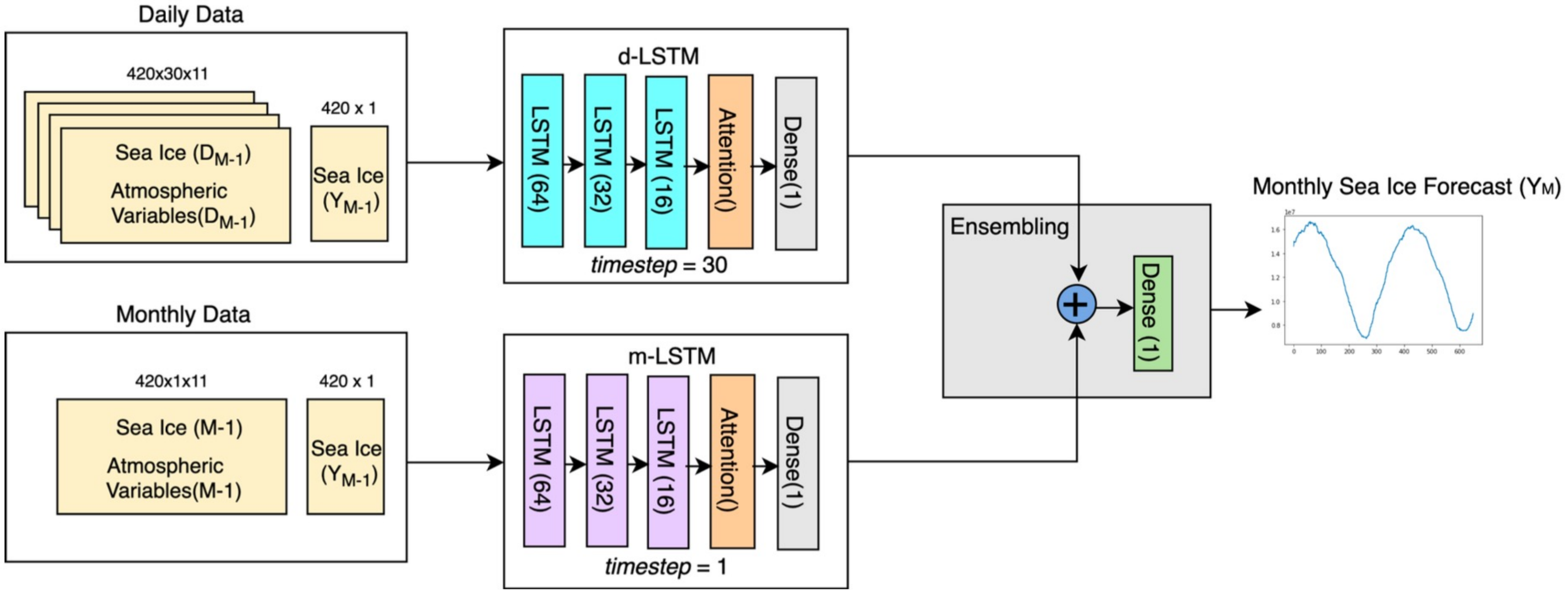
Target variable

VARIABLE	RANGE	UNIT
SEA ICE EXTENT	[0, 1e7]	km^2



Innovations

1. Combine daily and monthly observations as an ensemble model
2. Implement attention to learn to focus on specific features of the hidden states



UNIVERSITY OF WATERLOO | FACULTY OF ENGINEERING

Metrics used

- R^2 score
  - $R^2 = 1 - \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - u)^2}}$
- Percent Root Mean Squared Error (%RMSE) using mean
  - $RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$
  - $\% RMSE = \frac{RMSE}{mean(\hat{y})}$

Replicating their results

- Summary: Could not replicate their results, despite having the original code and data available
  - The models do not produce consistent results between random model initializations, likely due to poor regularization.
  - We suspect their results are the product of a single model run, rather than an average of multiple trials.

Model	R^2 Score	% RMSE
d-LSTM	0.980	4.45
m-LSTM	0.981	4.21
EA-LSTM	**0.982**	**4.11**

Table 1. Original results by Ali et al.

Model	R^2 Score		% RMSE	
	Mean	Std. Dev.	Mean	Std. Dev.
d-LSTM	**0.983**	**2.72e-3**	**4.15**	**3.30e-3**
m-LSTM	0.974	7.56e-3	4.86	7.15e-3
EA-LSTM	0.964	6.56e-3	5.78	5.19e-3

Table X. Results obtained in this work using n=20 model runs.

ML METHODS & BASELINES

How do simple algorithms compare?

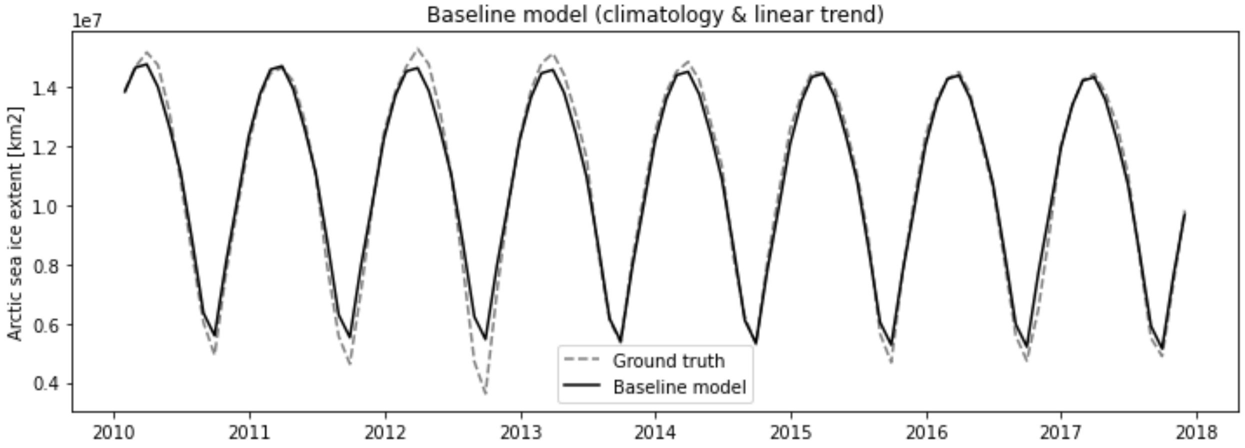
Comparison of ML model performance

Model	R^2 Score		% RMSE	
	Mean	Std. Dev.	Mean	Std. Dev.
Linear Regression	0.975	1.14e-16	4.85	2.14e-15
Ridge Regression	0.975	2.28e-16	4.87	2.14e-15
Lasso Regression	-0.12	0.00	32.4	5.69e-15
Decision Tree	0.967	4.41e-3	5.57	3.82e-3
Random Forest	**0.981**	**5.45e-4**	**4.25**	**6.01e-4**
Gradient Boosting	0.975	2.18e-4	4.85	2.11e-4
XGBoost	0.979	2.28e-16	4.45	7.12e-16
Polynomial Regression (deg=2)	0.927	1.14e-16	8.31	1.42e-15
SVM Regression	0.969	1.14e-16	5.43	0.00

Table 4. Results from traditional machine learning methods on Sea Ice prediction with a lead time of 1 month calculated over 20 experiments

Comparing to a baseline

- To properly assess the skill of their model, we compare their results to a simple statistical baseline: climatology with a linear trend
  - Climatology: Monthly averages of the entire training set
  - Linear trend: Linear trend of the residuals after removing climatology (to account for the steady decrease in SIE)

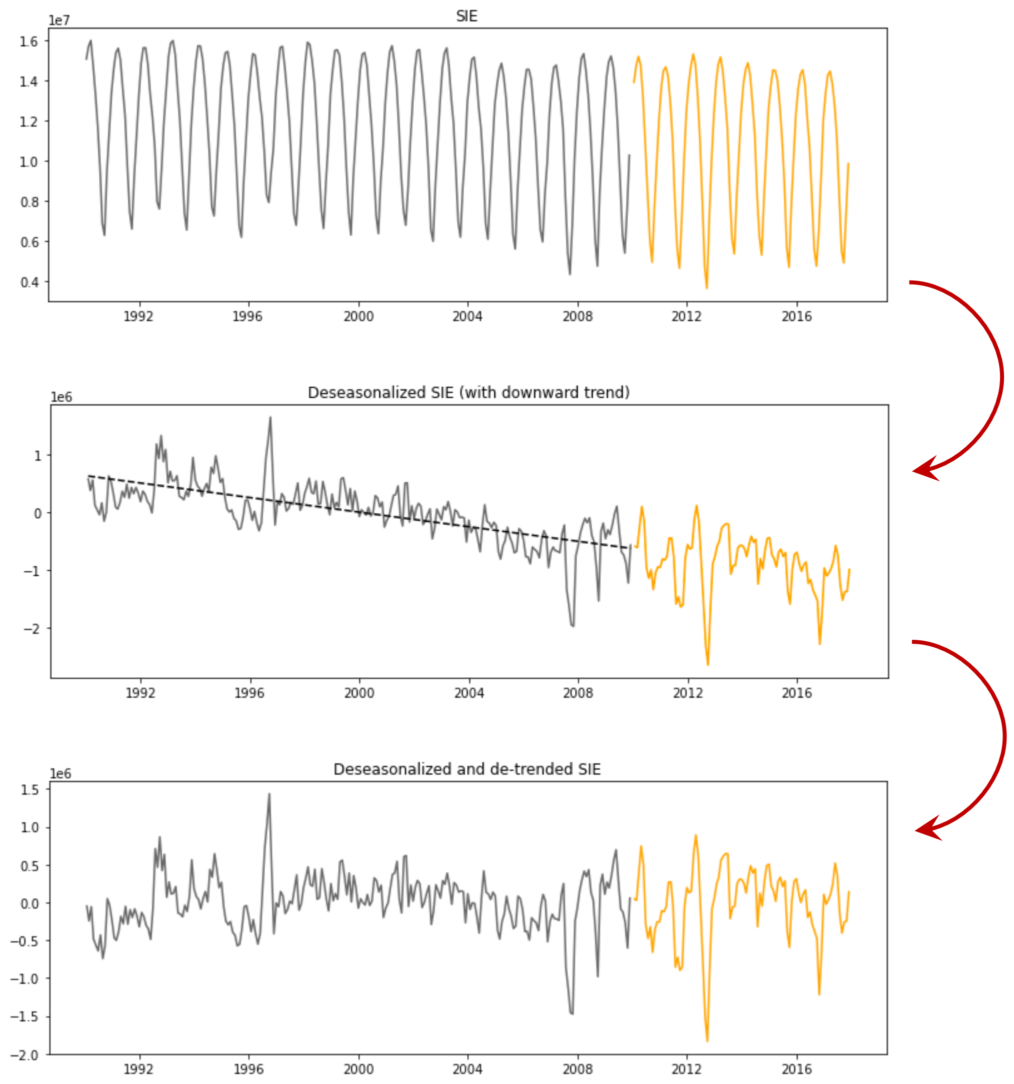


Model	R^2 Score	% RMSE
d-LSTM	0.983	4.15
m-LSTM	0.974	4.86
EA-LSTM	0.964	5.78
Baseline	**0.986**	**3.60**

- The baseline model performs best & can account for 98.6% of the variation!

Beating the baseline

- A skillful model should be able to predict variations in the target variable after having removed seasonality and linear trend (i.e. be able to predict anomalies)
- Removing the seasonality and linear trend from the target data allows the model to focus on capturing non-obvious variations
- **Can we use Ali et al.'s architecture to predict these anomalies?**



Beating the baseline

- Can we use Ali et al.'s architecture (& modifications) to predict anomalies?
  - Modifications:
    - Remove seasonality from inputs and target variables
    - Add several new timeframes (3-day, 10-day, 15-day) to the ensemble
    - Train the ensembles separately before combining with a dense layer
    - Increase/decrease number of LSTM layers & cells

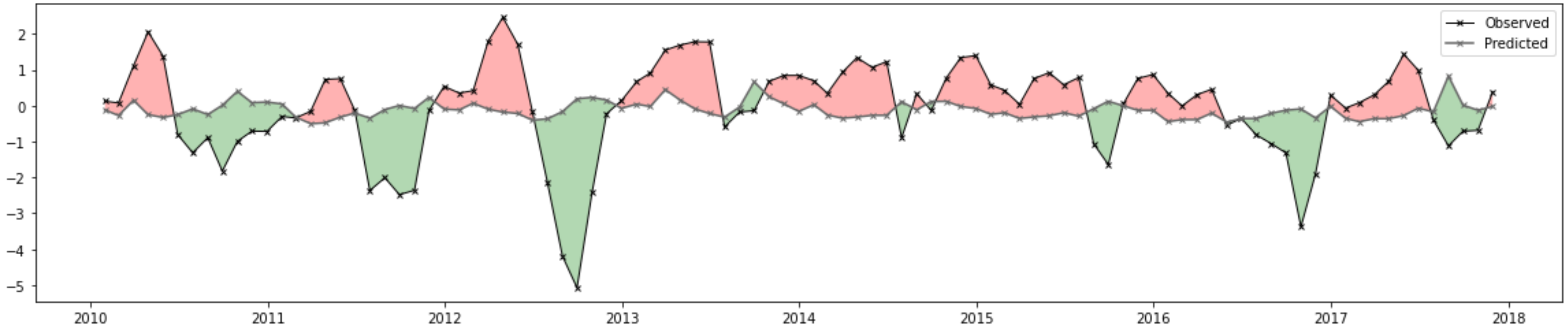


Figure #. Arbitrary example of a poorly-performing model

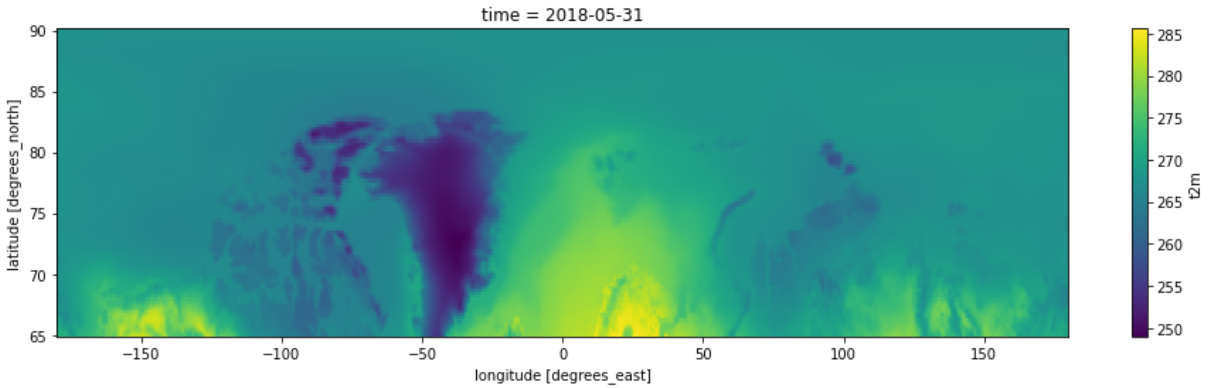
- **Short answer: No.**

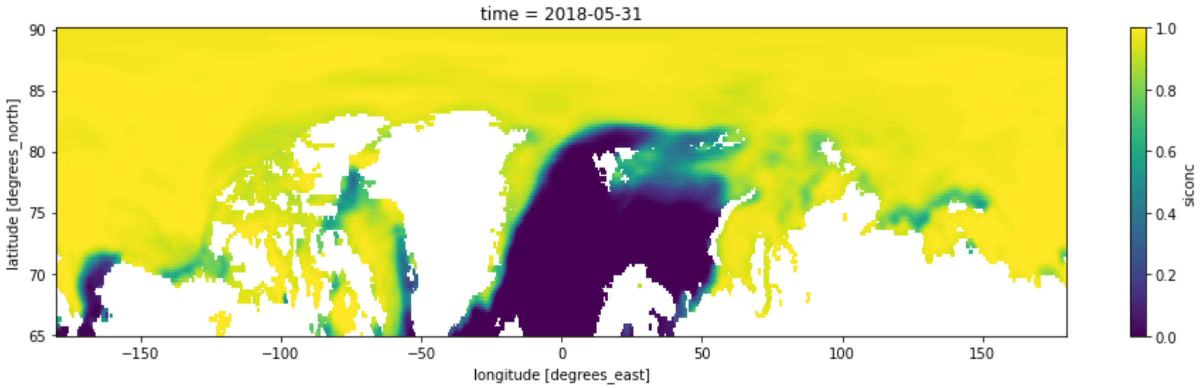
IMPROVEMENTS

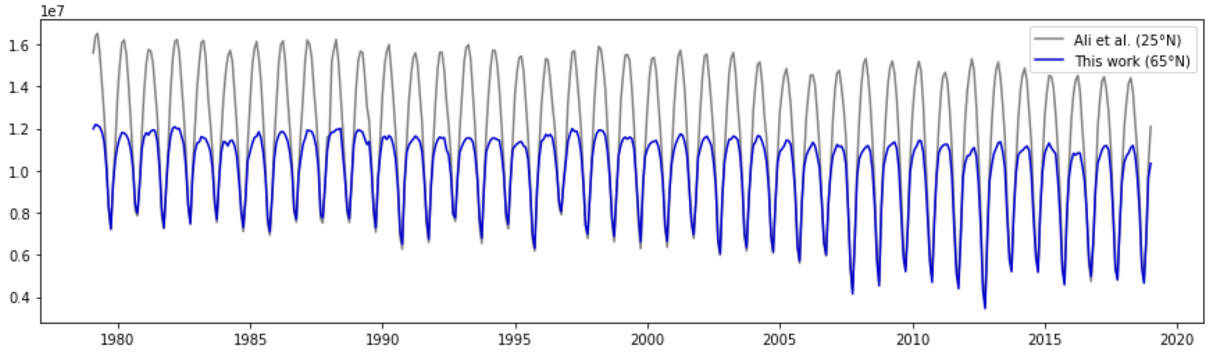
Can we do better?

Introducing spatial information

- Ali et al.'s model uses 1-D inputs, thereby ignoring any spatial information
- New input dataset: ERA5 data North of 65N* & aggregated to monthly means to keep data size manageable
- Note that we are using ERA5 sea ice concentrations to calculate SIE







*65N is a more common definition of the pan-Arctic domain

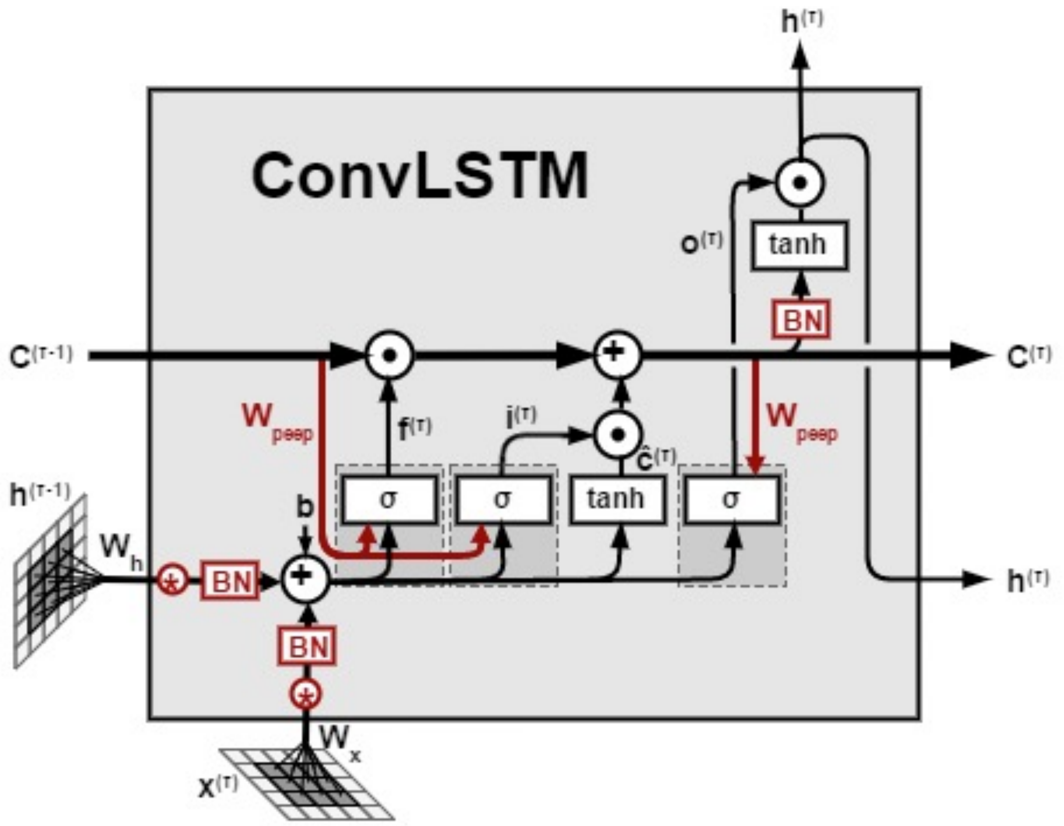
Climatological variables (19)

Variable	Units
u-component of wind (u10)	m s^{-1}
v-component of wind (v10)	m s^{-1}
Dewpoint temperature (d2m)	K
2m temperature (t2m)	K
Total precipitation (tp)	m
Snowfall (sf)	m of water eq.
Sea surface temperature (sst)	K
Sea ice area fraction (siconc)	%
Surface solar radiation downwards (ssrd)	J m^{-2}
Surface sensible heat flux (sshf)	Jm^{-2}

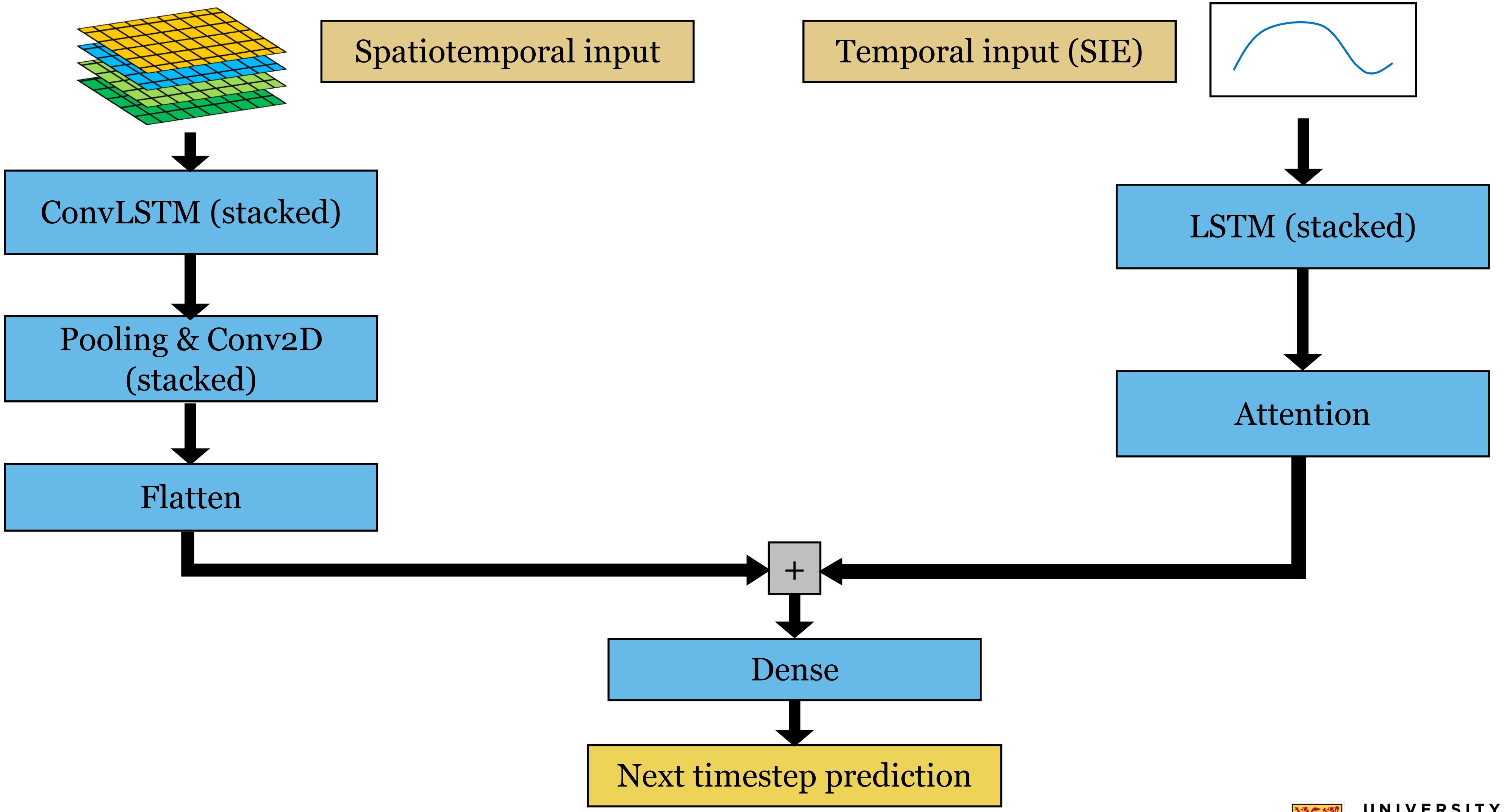
Variable	Units
Surface latent heat flux (slhf)	Jm^{-2}
Mean sea level pressure (msl)	Pa
Evaporation (e)	m of water eq.
Snowmelt (smlt)	m of water eq.
Total cloud cover (tcc)	%
Significant wave height (swh)	m
Mean wave period (mwp)	s
Mean wave direction (mwd)	Deg
10 metre wind gust (fg10)	m s^{-1}

Convolutional LSTMs (ConvLSTM)

- LSTMs cannot directly be used on spatial data, therefore we introduce convolutional LSTMs
  - ConvLSTM = LSTM where cell inputs and outputs, hidden states and gates are 3D tensors where the first two dimensions are the spatial dimensions.



General architecture



Model Improvements and Hyperparameter Tuning

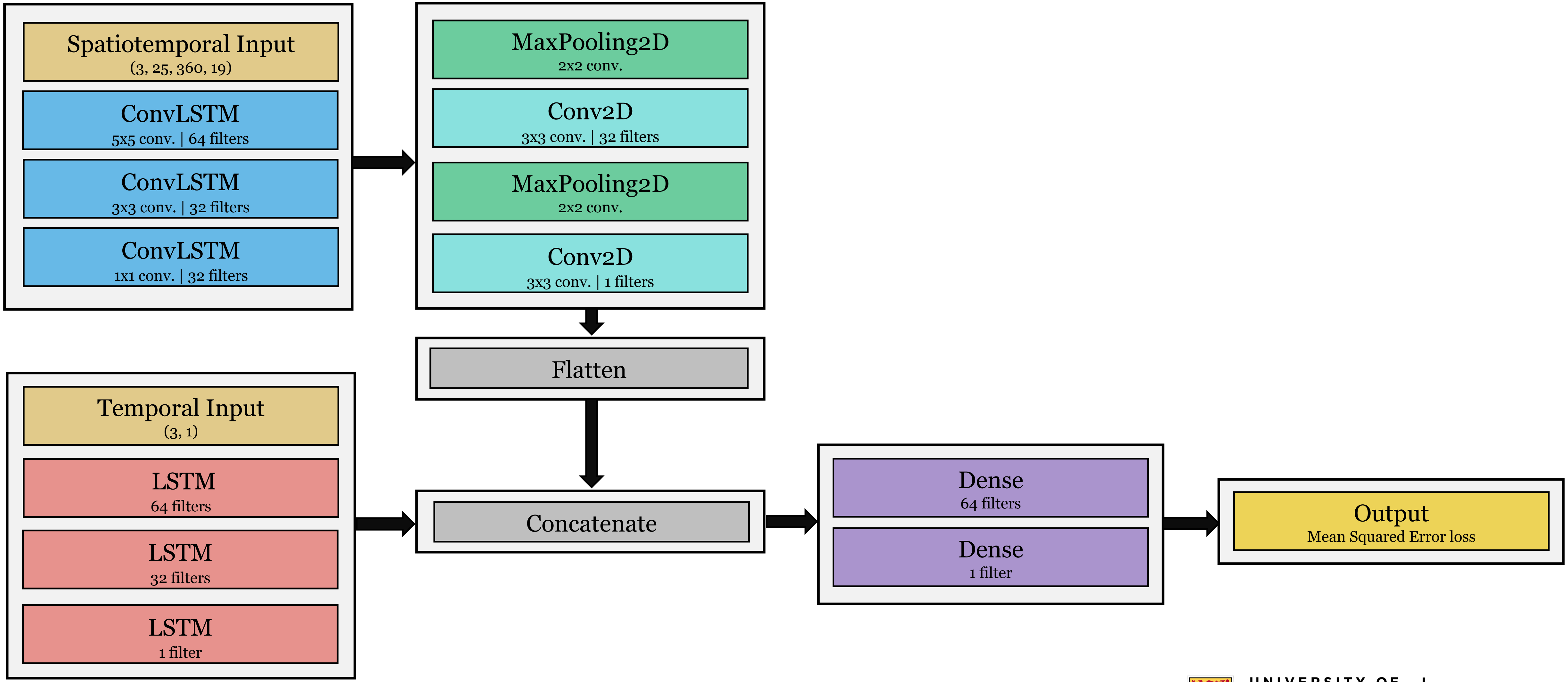
Hyper-parameter	Range
Image size (i.e. degree of downsampling)	[1440×101, 720×50, 360×25]
Choice of variables	[19 variables]
Number of ConvLSTM layers	[1 - 4]
Number of Conv2D layers	[1 - 4]
Number of LSTM layers	[1 - 4]
Number of hidden units	[16 - 256]
Kernel size	[1x1, 3x3, 5x5]
L1, L2 kernel regularization	[1e-4, 1e-3, 1e-2, 1e-1]
Dropout	[0.1, 0.2]
Global average pooling	[True, False]
Learning rate	[1e-4 – 1e-1]

Metrics

- Root mean squared percent error (RMSPE)
  - $RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$
- Mean absolute percent error (MAPE)
  - $MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{\hat{y}_i}\right|$
- Geometric mean relative absolute error (GMRAE)
  - GMRAE = $\sqrt[N]{\prod_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{|\hat{y}_i - b_i|}}$, where $b_i = y_{i-m}$

$\hat{y}_i = Predicted\ SIE$
$y_i = True\ SIE$

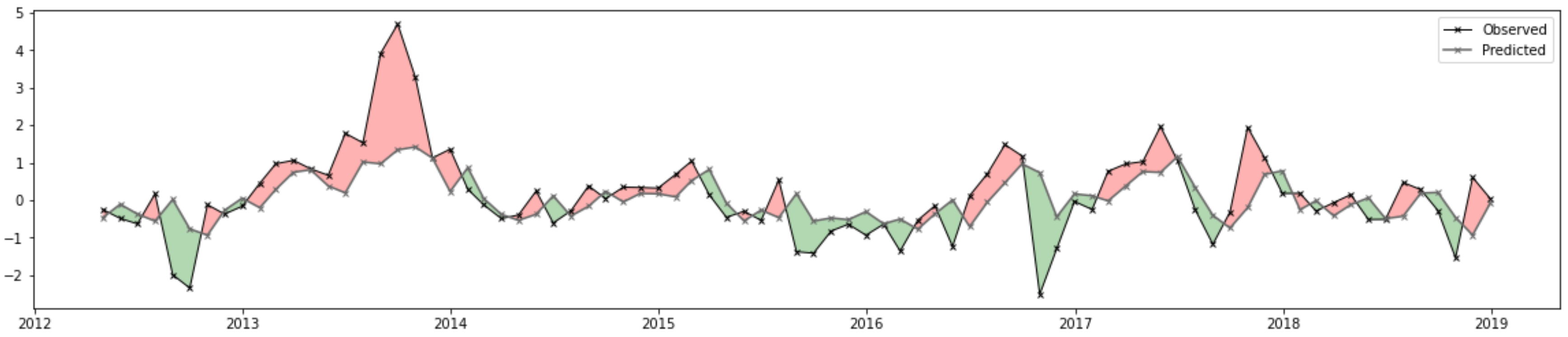
Sample architecture



Preliminary results

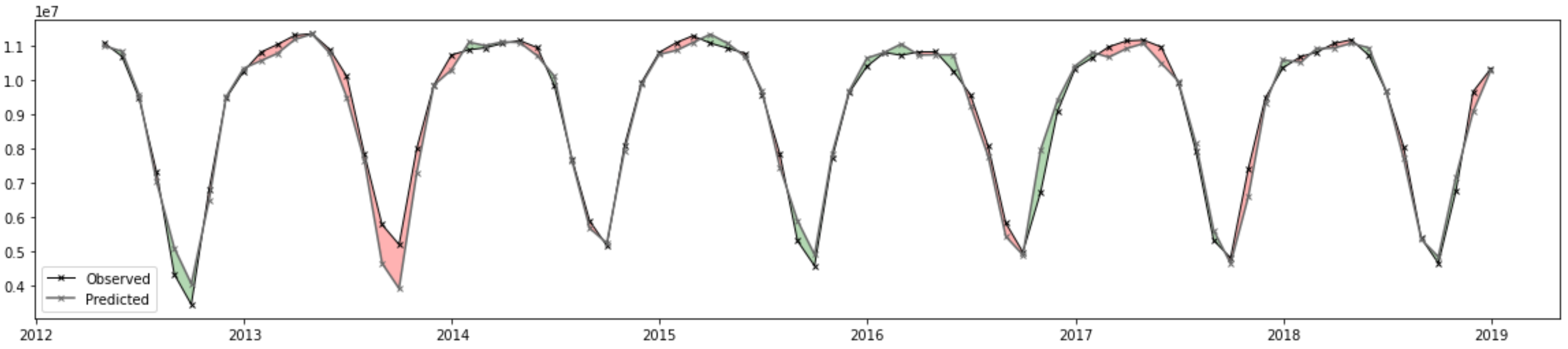
Anomalies

Model	RMSE	MAE
ConvLSTM	**0.961**	**0.678**
Persistence	0.995	0.730



Sea ice extent

Model	RMSPE	MAPE
ConvLSTM	**6.40**	**3.67**
Climatology	8.70	4.65



References

- Original paper: https://arxiv.org/abs/2108.00853
- ERA5: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5
- ConvLSTM: https://arxiv.org/abs/1506.04214

UNIVERSITY OF
WATERLOO

FACULTY OF ENGINEERING