No. 81, November 2021
DOI: 10.36071/clivar.81.2021

CLIVAR Exchanges

Special Issue: Advances in Climate Prediction Using Artificial Intelligence

WCRP World Climate Research Programme

WORLD METEOROLOGICAL ORGANIZATION | United Nations Educational, Scientific and Cultural Organization | Intergovernmental Oceanographic Commission | International Science Council

CLIVAR (Climate and Ocean: Variability, Predictability and Change) is the World Climate Research Programme's core project on the Ocean-Atmosphere System

Editorial

Guest editors: Jing-Jia Luo[1], and Swadhin Behera[2]

1. Institute for Climate and Application Research (ICAR), Nanjing University of Information Science and Technology, Nanjing, China
2. Application Laboratory, VAiG, JAMSTEC, Yokohama, Japan

Accurate weather and climate forecasts are essential for disaster preparedness, socio-economic development, and ecosystem management. This is especially important in the present time under the increased pressure of anthropogenic warming that have been threatening global sustainability. Tremendous efforts have been made to improve forecasts by means of innovative or upgraded statistical and numerical models. More sophisticated ocean-atmosphere models are being used with much complicated physics and finer resolutions for improving operational weather and climate forecasts. However, current skills in predicting weather and climate, in particular extreme events, are still limited. In this regard, the World Weather/Climate Research Programmes (WWRP/WCRP) have set up working groups to improve numerical models and seamless forecasts covering weather and climate from sub-seasonal to interdecadal timescales.

Meanwhile, with the rapid increase of model-simulated data as well as observational data collected from various *in situ* and remote techniques, we are heading towards a new era in the use of big data in weather and climate science. This calls for an innovative way to effectively deal with these big data and correctly dig out complex relations/physics hidden in them. Stimulated by Google AlphaGo, which was built with a deep machine learning algorithm and astonishingly beat the human world champion of Go in 2016, Artificial Intelligence (AI) has attracted more and more attention to improving weather and climate forecasts in recent years. Successful applications have demonstrated that performance of the forecast models built on AI, while still in its infancy stage, displays a promising advance, particularly for weather forecasts because of the availability of abundant model and observational data samples for training AI models. In addition, AI has also been widely used as alternative attempts to tackle difficulties in cloud and weather type classifications, disaster evaluation, ocean wave and storm prediction, green energy power forecast, data assimilation, model development including sub-grid scale process parameterization and solution of Navier-Stokes partial differential equations, numerical model output corrections, downscaling, and so on.

Compared to weather forecasts, building AI models for climate forecasts has been very challenging. This is because of rather limited records of climate observations available for training the AI models. Hence, additional efforts are expected for data augmentation since it is usually hard to accurately reproduce realistic physical features/relations. To encourage the endeavors in improving AI climate forecasts, Nanjing University of Information Science and Technology (NUIST), Alibaba DAMO Academy, National Climate Centre, National Marine Environmental Forecasting Center, and Anhui Meteorology Service in China jointly organized the "2021 AI Earth" prize competition: "Artificial Intelligence Innovation Challenge--AI Helps Accurate Weather and Ocean Forecast", with data supported by the Institute for Climate and Application Research (ICAR) (detailed information is available at https://tianchi.aliyun.com/competition/entrance/531871/introduction?spm=5176.12281957.1004.2.38b03eafcjKhrq&lang=en-us). This first world competition on AI-based prediction of climate phenomena like El Niño/Southern Oscillation (ENSO) attracted 2849 teams from 19 countries. Among them 47.4% were from Chinese universities, 12.4% from overseas universities, 29.2% from private enterprises, and

10.9% from research institutes.

This special issue records some of those studies reporting recent advances in the application of AI to ENSO forecast (Ham et al.; Mu et al.; Patil et al.; and the top team of the competition: Luo et al.), numerical weather forecast correction and downscaling (Huang et al.; Hu et al.), heavy precipitation forecast (Ma et al.), ocean wave height forecast (Li et al.), and ocean modeling (Song et al.). The study of climate forecasts using AI is still in its infancy, but we expect that AI will provide an innovative way to fundamentally improve the predictions and understanding of weather and climate events in the coming years. It is worth noting here that the World Meteorological Organization (WMO) has also recently launched a competition called "Prize Challenge to Improve Sub-Seasonal to Seasonal Predictions Using Artificial Intelligence" (see https://s2s-ai-challenge.github.io/). Hence, more contributions to this exciting topic are expected in near future.

We hope that this special issue in *CLIVAR Exchanges* will encourage more efforts and attempts to develop skillful AI-based climate forecast models with a high interpretability of the results. We would like to thank here the generous financial support from Machine Intelligence Lab, Team of AI EARTH, and TIANCHI Platform of Alibaba DAMO Academy, particularly Ting Gao for coordinating the "2021 AI Earth" competition, and Hao Li, Zhibin Wang, Lei Chen, Yuan Hu, and Yuan Hong for their excellent technical supports. Finally, we would also like to thank Jose Santos, Jing Li, Qian Zhao and Peng Lian at the International CLIVAR Project Office for their tireless support in the scheduling, organizing, and typesetting of this special issue.

Recent progress in ENSO forecast using deep learning

Yoo-Geun Ham[1], Jeong-Hwan Kim[1], and Jing-Jia Luo[2]

1. Department of Oceanography, Chonnam National University, Gwangju 61186, South Korea
2. Institute for Climate and Application Research (ICAR), Nanjing University of Information Science and Technology, Nanjing 210044, China

1. Introduction

El Niño/Southern Oscillation (ENSO) has substantial influence on global climate variability and is known to be responsible for not only extreme weather events such as heat waves, floods, droughts, and heavy rainfall (Timmermann et al. 2018; Cai et al. 2019; 2020) but also modifications to global ecosystems (Park et al. 2019). Therefore, the successful prediction of ENSO is of considerable interest to reducing potential economic and social damage. To meet the worldwide demand for reliable ENSO forecasting, more than 30 operational forecast systems are regularly operated to predict the Nino3.4 index, which is a representative index to diagnose the amplitude and phase of ENSO, as collected by IRI (International Research Institute for Climate and Society) ENSO forecasts (https://iri.columbia.edu/our-expertise/climate/enso/).

Although ENSO forecasts produced by the current state-of-the-art forecast systems are approximately reliable up to 6 months in advance, the one year advance forecasting of ENSO remains challenging. In addition, there is a limitation in that the forecasts started from the boreal spring season are less reliable than those issued from other seasons (Webster and Yang 1992). For example, in the boreal spring of 2014, most operational forecast systems predicted the likely occurrence of a super El Niño event at the end of 2014; however, the development of El Niño event was abruptly stalled after the boreal summer of 2014, and the smaller amplitude of El Niño event during 2014/15 resulted in a moderate warm event.

During the last few years, significant progress has been achieved in forecasting ENSO by applying deep learning algorithms. Ham et al. (2019) (hereafter referred to as H19) applied the convolutional neural network (CNN) algorithm, a widely used image recognition approach, to ENSO forecast. The CNN was selected in order to appropriately capture ENSO precursor signals beyond the Pacific. For the first time, using the deep learning-based model, they showed that the predicted Nino3.4 index is reliable by up to 1.5 years, exhibiting a significant improvement in comparison to current state-of-the-art ENSO forecast systems. After the pioneering efforts of H19, attempts have been made to apply various types of deep learning algorithms to ENSO forecast as we will introduce in more detail in the next section.

2. Recent progress

One of these follow-up studies to apply deep learning to ENSO forecasts introduced a time sequence in the input data (i.e., predictor) (Yan et al. 2020; Gupta et al. 2019; Cachay et al. 2021). The basic concept was to consider the past time evolution in the input of multiple time frames to eventually help predict future time evolution. The temporal convolutional network (TCN) is a variant of the CNN that adapts the notion of the convolutional process to formulate a sequence model. Similar to the recurrent neural network, which is commonly used for the prediction of time series, the employment of causal and dilated convolutions in the TCN facilitates the use of sequential data with temporality and large receptive fields.

Yan et al. (2020) proposed the TCN method combined with the ensemble empirical mode decomposition (referred to as the EEMD-TCN). This hybrid approach first decomposes the time series of ENSO index according to the frequency using the EEMD; then, each component is predicted using the TCN model, and finally combined to obtain the final prediction result. The all-season correlation skill during 1984–2017 in the EEMD-TCN model reached approximately 0.8 for the 12-month lead forecasts. Compared to the all-season correlation skill of the CNN model of H19 (i.e., 0.65 for the 12-month lead forecasts), one can consider that the EEMD-TCN exhibits better forecast skill than the CNN. However, it should be noted the EEMD is contentious to be applied for real-time forecasts as the smoothing process may be transferring information from the future (i.e., test set) to the past (Cachay et al. 2021). In other words, any type of time-filtering is not possible for the real-time forecasts because of their requirement of the information beyond the end of the time series (Wheeler and Hendon 2004). As the TCN without the EEMD is approximately 0.5 for the 12-month lead forecasts (Yan et al. 2020), it may not be helpful to introduce the time sequence in the input variables using the TCN to improve the ENSO forecast skill.

Convolutional long short-term memory (ConvLSTM) is one of the representative algorithms that considers a time sequence of the input data. Gupta et al. (2019) applied the ConvLSTM to predict the monthly mean Nino3.4 index and argued that the model can be skillfully predicted up to one year ahead. They showed

that the predicted time series using the ConvLSTM model were systematically closer than those predicted using the CNN model for major El Niño events (e.g., 1997/98, 2009/10, 2015/16, and 2018/19). However, the overall forecast performance of the CNN in their study is not as good as that of H19, probably because the version of CNN used by Gupta et al. (2019) might not present optimal usage of training data. Gupta et al. (2019) used only sea surface temperature (SST) data from the reanalysis to train the model, whereas H19 additionally utilized heat content to allow recognition of the subsurface temperature signal and centennial simulations from the CMIP5 multi-model climate models. Therefore, a comparison of the forecast performance between ConvLSTM and CNN based models is worthwhile to be performed by ensuring optimal usage of the training data.

The graphic neural network (GNN) is utilized for forecasting ENSO events (Cathay et al. 2021). They argued that the GNN is advantageous for forecasting ENSO as GNN links the information between remote grid points through edge-connections, whereas the convolutional process in the CNN only gathers information in the adjacent grid points. Thus, GNN generalizes convolutions to non-Euclidean data, and allows modeling of large-scale global connections as edges of a graph. This indicates that the GNN can appropriately consider nonlinear interactions between the remote precursors for ENSO. Note that their usage of training data is almost the same as that of H19; therefore, the ENSO forecast skill of GNN by Cathay et al. (2021) might be optimal, at least in terms of the training dataset. They showed that the all-season correlation skill of the Nino3.4 index in the GNN is systematically better than that in the CNN during the forecast lead months shorter than 6-months; however, the CNN exhibits better forecast skill at 9- and 12-month lead forecasts.

This comparable ENSO forecast skill between GNN and CNN implies that GNN is not the only technique to simulate the large-scale global connections. CNN also has the ability to partially simulate the interaction between the remote grid points as the information of each grid point in the feature map is gathered in each neuron in the fully-connected layer. Therefore, the nonlinear amplification of ENSO owing to the co-emergence of multiple ENSO precursor signals can also be partially captured in the CNN.

3. ConvLSTM model for short-term ENSO forecasts

Considering the aforementioned deep learning algorithms, we formulated the ConvLSTM model to compare forecast results with those of the CNN model of H19. For this purpose, we utilized the same training dataset of the CMIP5 model outputs to formulate the ConvLSTM model (Figure 1). The inputs (X), cell outputs (C), and hidden states (H) of the LSTM are one-dimensional vectors, while that of the ConvLSTM are three-dimensional tensors. Therefore, input (I), forget (F), output gates (O) are also derived as three-dimensional tensors. Furthermore, by replacing all dot products with convolution operators (*), the number of parameters can be significantly reduced compared to applying LSTM cells to all grid points (Figure 1a). These changes allow ConvLSTM to learn the changes or movements of objects in the input maps or feature maps. The ConvLSTM model consists of our ConvLSTM layers and two max-pooling layers between them. After the fourth ConvLSTM layer, there are two fully-connected layers which link to the final output (i.e., the predicted Nino3.4 index) (Figure 1b).

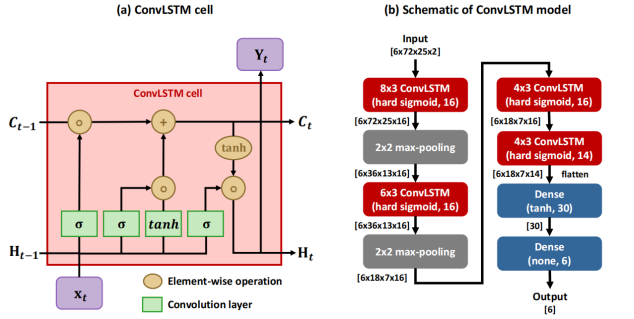

Figure 1: Description of the (a) convolutional long short-term memory (ConvLSTM) cell, and (b) ConvLSTM model for ENSO forecasts .

A noticeable difference between the CNN and the ConvLSTM model is the frequency of the input data; the 10-day averaged values from 60 days before the initial time (i.e., a total of six time frames) are used for the ConvLSTM, whereas monthly-averaged data from 3-months before the initial time (i.e., a total of three time frames) are used for the CNN model. This approach aims to appropriately and comprehensively consider the time evolution of the input variables in the ConvLSTM, and hopefully contributes to better prediction of the Nino3.4 index, particularly for the short-term forecasts (i.e., up to 6-months lead) to complement the CNN of H19 whose excellence lies in the forecasts at longer than 6-months lead. Note that, as the reanalysis output used as the training dataset in H19 is not provided in 10-day frequency (Carton and Giese 2008), the model is solely trained with the CMIP5 model outputs. The generated forecast results in the ConvLSTM were compared with those of the CNN model for up to 6-months leads for most recent ENSO events (i.e., 2019/20 El Niño event and 2020/21 La Niña event). This is not only to validate the ENSO forecast skill of the ConvLSTM, but also to ensure the superiority of the CNN in the H19 in real-time ENSO forecasts.

In addition, rather than the heat content anomaly as used by H19, the simulated 20° isotherm depth anomaly (D20) was used as an input variable for the ConvLSTM. The D20 data were used to obtain a 10-day averaged value from the daily outputs; we believe that the difference in the input variable between the ConvLSTM and the CNN might not cause a severe difference in the ENSO forecast skill.

The same simulated SST dataset was used to train the ConvLSTM and CNN models.

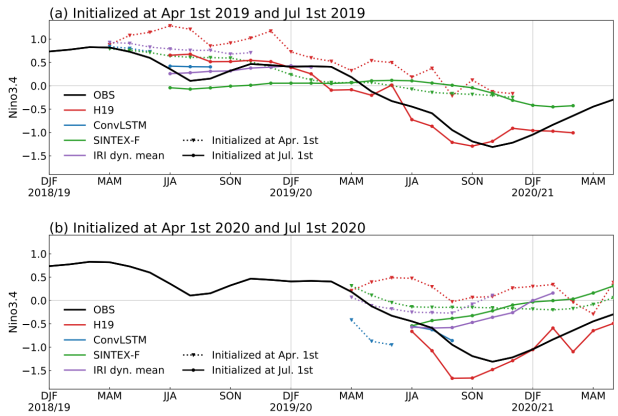

Figure 2: Predicted Nino3.4 index (unit: °C) in the convolutional neural networks (CNN) model developed in Ham et al. (2019) (red), convolutional long short-term memory (ConvLSTM) (blue), SINTEX-F (green), and multi-model ensemble average included in the International Research Institute for Climate and Society (IRI) ENSO forecasts. The black line shows the reference Nino3.4 anomalies obtained by global ocean data assimilation system (GODAS) reanalysis. Panel (a) shows the forecast results started on April (dotted lines) and July (solid lines) 1, 2019. Panel (b) is the same but for 2020.

Figure 2 shows the forecast results for the 2019/20 El Niño event and 2020/21 La Niña event. In addition to the forecast result of the CNN and the ConvLSTM models, the multi-model average of the dynamical forecast systems included in the IRI ENSO forecasts and the SINTEX-F model (Luo et al. 2008) are also shown. The observed Nino3.4 index was weakly decreased in time from the February-March-April to July-August-September (JAS) season, and then it intensified up to the boreal winter of 2019/20. As a result, the Nino3.4 index during December-January-February (DJF) of 2019/20 exhibited approximately 0.5°C. The weak El Niño signal during the boreal winter of 2019/20 was weakened after the boreal spring of 2020, and a subsequent La Niña event occurred during the early winter of 2020/21.

For the forecasts started from April 1, 2019 (Figure 2a), the forecasted Nino3.4 index in all of the ENSO forecast systems during 2019 tended to be larger than the observed. The predicted positive Nino3.4 anomaly was largest in the H19 model, and even increased in time up to the boreal summer of 2019. In contrast, the predicted Nino3.4 index was realistic to some extent for the first 2 months in the ConvLSTM, multi-model average of dynamical models collected by IRI, and SINTEX-F. However, none of the models reproduced the observed weakening of the positive Nino3.4 anomalies during the boreal summer season. It should be noted that the predicted Nino3.4 index in the ConvLSTM exhibited a more realistic temporal evolution than that of the CNN, indicating the possible improvement in simulating the ENSO by considering the time sequence of the input data.

For the forecasts started from July 1, 2019, the results at the early lead months were quite similar between the models. The predicted Nino3.4 index in all models did not exhibit any significant changes in time; therefore, none of the models predicted the minima of the Nino3.4 index during the JAS season of 2019. For the forecasts at lead times longer than 9 months, the differences between the models were obvious. The CNN model successfully predicted the observed weak positive Nino3.4 values during DJF 2019/20, whereas the other models predicted an amplitude weaker than that observed. More importantly, the phase transition to La Niña during 2020 was also well predicted in the CNN model. On the contrary, the forecasted Nino3.4 in the SINTEX-F exhibited continuous normal conditions during 2020; therefore, La Niña event in the boreal winter of 2020/21 was not realistically predicted. This is consistent with the main arguments of H19 that the forecast skill of the Nino3.4 index is significantly improved in the CNN.

For the forecasts started from April 1, 2020 (Figure 2b), the negative Nino3.4 index was hardly predicted in the CNN, SINTEX-F, and multi-model average of the IRI-collected dynamical models. This would imply that predicting the onset of La Niña event during the boreal spring of 2020 was quite challenging. Interestingly, the phase transition of the Nino3.4 index from the positive to the negative was partially predicted only in the ConvLSTM model. For the forecasts starting from July 1, 2020, the decrease in the Nino3.4 index was also well predicted in the ConvLSTM model. Although the decrease in the Nino3.4 index was predicted in the CNN model, the decrease was in excess of that observed. La Niña event in 2020/21 was successfully but partially predicted in the CNN and ConvLSTM models. In contrast, the dynamical models consistently predicted the gradual increase in the Nino3.4 index. Therefore, La Niña event during the boreal winter of 2020/21 was not predicted at all in the SINTEX-F and multi-model averaged of dynamical models collected by IRI. For example, the predicted Nino3.4 index during DJF of 2020/21 was nearly zero in the IRI multi-model average and SINTEX-F models, whereas that in the CNN was approximately -1°C, close to the observed value. This clearly demonstrates the advantage of the deep learning-based models with the introduction of the time sequence of the input variables.

The cause of the superior forecasts in predicting the 2020/21 La Niña event in the ConvLSTM model was examined in more detail. Figure 3 shows the spatial distribution of the heat content, and D20 which were used as the initial conditions for the forecasts started on April 1, 2020 in the CNN and ConvLSTM, respectively. The monthly-averaged heat content anomalies exhibited positive values over the equatorial central Pacific during January and February 2020, which was systematically weakened in March 2020. The degree of the weakening of positive heat content anomalies is not robust from January to February 2020,

whereas an abrupt decrease is shown from February to March 2020. This indicates that the positive heat content anomalies which were responsible for El Niño during the boreal winter of 2019/20 were diminished, and the subsequent La Niña event initiated from March 2020. The 10-day averaged values of the D20 anomalies also exhibited a similar time evolution. The degree of the decrease in the D20 anomalies over the equatorial central Pacific is relatively weak from the end of January to February 2020 (Figures 3d–3f); the negative time-tendency of the D20 anomalies is clearly exhibited during March 2020 (Figures 3h and 3i).

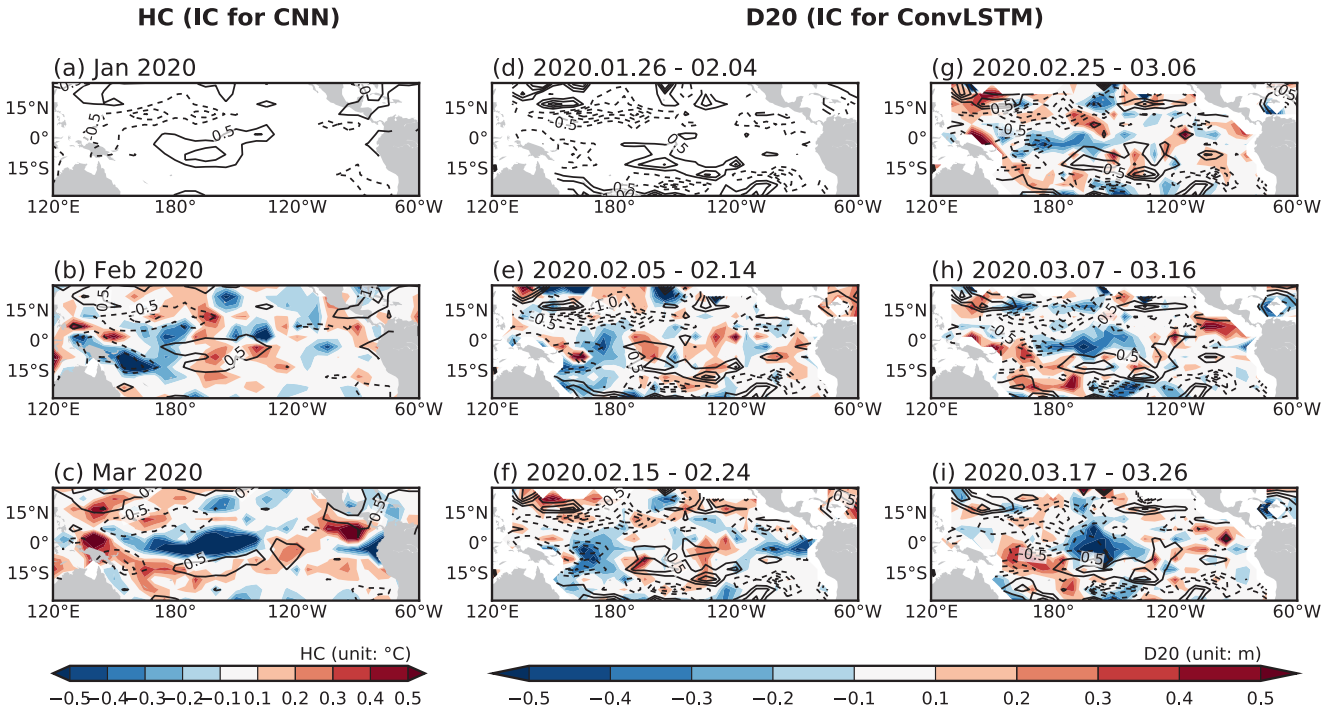

Figure 3: Monthly heat content anomalies (unit: °C) in (a) January, (b) February, and (c) March 2020, which were used as initial conditions for the forecasts started on April 1, 2020 in the convolutional neural networks (CNN) model. Panels (d) to (i) show the 10-day averaged 20°C isotherm depth anomalies (unit: m) during January to March 2020 for the initial conditions for the forecasts started on April 1, 2020 in the convolutional long short-term memory (ConvLSTM) model.

It is likely that the abrupt changes in the initial conditions at the latest time frames can be correctly recognized in the ConvLSTM model. As the initial conditions in the ConvLSTM are linked along a temporal sequence, the changes in initial conditions at the latest time frames might directly affect the forecasts. In other words, the abrupt decrease in the subsurface temperature anomalies over the equatorial central Pacific during March 2020 is likely to lead the continuous decrease in the Nino3.4 index in the forecasts, which contributes to development of La Niña event. The superiority of the ConvLSTM model can also be interpreted as the consideration of the relative importance between initial conditions at different time frames. In other words, the signals in initial conditions at a past time frame are relatively difficult to be transferred to the output node, as these are connected to the output through multiple time nodes. In contrast, the signals in the initial conditions with the latest time frame are directly linked to the output; therefore, their influence on the output would be relatively strong. As a result, the decrease in the heat content anomalies in the initial conditions at the latest time frame can be considered important in the ConvLSTM. This type of time sequence between initial conditions was not appropriately designed in the CNN model, which resulted in the failure in predicting the development of La Niña event initiated during boreal spring of 2020.

4. Concluding remarks

Since it was demonstrated that deep learning-based ENSO forecast models outperform current state-of-the-art dynamical forecast models, various deep learning techniques have been applied to further improve the forecast skill of the ENSO. ConvLSTM networks can potentially outperform CNNs by considering the time evolution of the input variables, and TCNs combined with the technique of separating the low- and high-frequency components of the input time series would be beneficial to correctly simulate an upscale feedback of high-frequency winds to ENSO. The GNN is designed to efficiently extract the remote ENSO-related precursor signals with complex spatial distributions.

In line with the efforts to adopt the aforementioned advanced deep learning algorithms for ENSO forecasts, we applied the ConvLSTM, in particular for ENSO forecasts at shorter than 6-month lead. The forecast quality in predicting the weak 2019/20 El Niño event was similar between the CNN and ConvLSTM models, whereas prediction of the 2020/21 La Niña event was superior in the ConvLSTM. In particular, among the forecasts started from April 1, 2020, the ConvLSTM was the only model among the four

model groups (i.e., ConvLSTM, CNN, SINTEX-F, and dynamical models included in the IRI forecasts) that successfully predicted the onset of La Niña signal. The abrupt decrease in the heat content anomalies over the equatorial Pacific during the boreal spring of 2020 was successfully recognized as the distinct developing signal of La Niña by the time sequencing algorithm in the ConvLSTM. This implies that the ConvLSTM can further improve ENSO forecast skill by appropriate consideration of the time sequence in the initial conditions.

Acknowledgements

This study is supported by the NRF under Grant No. NRF-2020R1A2C2101025. J.-J. L. is supported by National Key R&D Program of China (Grant No. 2020YFA0608004).

References

Cachay, S.R., E. Erickson, A.F.C. Bucker, E. Pokropek, W. Zotosnak, S. Bire, ... & B. Lütjens, 2021: The World as a Graph: Improving El Nino Forecasts with Graph Neural Networks. arXiv preprint arXiv:2104.05089.

Cai, W., L. Wu, M. Lengaigne, T. Li, S. McGregor, J.S. Kug, ... & P. Chang, 2019: Pantropical climate interactions. *Science*, **363**(6430).

Cai, W., M.J. McPhaden, A.M. Grimm, R.R. Rodrigues, A.S. Taschetto, R.D. Garreaud, ... & C. Vera, 2020: Climate impacts of the El Niño–southern oscillation on South America. *Nature Reviews Earth & Environment*, **1**(4), 215-231.

Carton, J.A., & B.S. Giese, 2008: A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA). *Monthly weather review*, **136**(8), 2999-3017.

Ham, Y.G., J.H. Kim, & J.J. Luo, (2019). Deep learning for multi-year ENSO forecasts. *Nature*, **573**(7775), 568-572.

Luo, J.J., S. Masson, S.K. Behera, & T. Yamagata, 2008: Extended ENSO predictions using a fully coupled ocean–atmosphere model. *Journal of Climate*, **21**(1), 84-93.

Timmermann, A., S.I. An, J.S. Kug, F.F. Jin, W. Cai, A. Capotondi, ... & X. Zhang, 2018: El Niño–southern oscillation complexity. *Nature*, **559**(7715), 535-545.

Webster, P.J., & S. Yang, 1992: Monsoon and ENSO: Selectively interactive systems. *Quarterly Journal of the Royal Meteorological Society*, **118**(507), 877 -926.

Wheeler, M.C., & H.H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly weather review*, **132**(8), 1917-1932.

Yan, J., L. Mu, L. Wang, R. Ranjan, & A.Y. Zomaya, 2020: Temporal convolutional networks for the advance prediction of ENSO. *Scientific reports*, **10**(1), 1 -15.

Improving long-lead predictions of ENSO using convolutional neural networks

Kalpesh Ravindra Patil, Takeshi Doi, Pascal Oettli, J. V. Ratnam, Swadhin Behera

Application Laboratory, VAiG, JAMSTEC, Yokohama, Japan

1. Introduction

El Niño/Southern Oscillation (ENSO) is a recurring natural climatic event associated with anomalous changes in the sea surface temperature (SST) of the central and eastern equatorial Pacific Ocean with a periodicity between 3 to 7 years. ENSO also exhibits significant inter-annual variation (Trenberth 1987) and is known to affect the global climate through large scale changes in the atmospheric teleconnections and the associated changes in the spatial distribution of global precipitation, surface temperature, frequency and path of tropical cyclones (Ropelewski and Halpert 1987). Prediction of ENSO events at a sufficient lead-time will help stakeholders to plan in advance to mitigate its adverse effects on agriculture, energy and health sectors (Kim et al. 2020a; 2020b). There are several indices to identify ENSO events based on SST anomalies (SSTA) over various regions of the equatorial Pacific Ocean, the most commonly used is the Nino3.4 index which is derived by averaging the SST anomalies over the Nino3.4 region 5N-5S; 170W-120W (Trenberth 1997). In this study, we attempt to forecast ENSO events through the Nino3.4 index using a deep learning technique, namely convolutional neural network (CNN).

Attempts to predict ENSO events can be dated back to the late 1980s. The forecasting techniques in general were either based on dynamical numerical models (Luo et al. 2008; Doi et al. 2016) or statistical models (Barnston et al. 1999). Even-though the dynamical models have fairly similar skill compared to the statistical models at shorter leads (6 months to 1 year), they generally exhibit lower skills at longer lead times (>1 yr) (Tangang et al. 1998; Barnston et al. 1999; Aguilar-Martinez and Hsieh 2009). The lower skill in dynamical models at longer leads is believed to be associated with the spring predictability barrier (Levine and McPhaden 2015). Recently, deep learning techniques, which are forms of advanced statistical models, are being widely used to predict climatic events with great success (cf. Reichstein et al. 2019). Deep learning techniques have also been used to successfully predict ENSO events at long lead times (Ham et al. 2019; Mahesh et al. 2019; Broni-Bedaiko et al. 2020; Cachay et al. 2020; Guo et al. 2020; Yan et al. 2020; Wang et al. 2021). These studies have shown good skill in forecasting ENSO up to 18 months lead time. The study of Ham et al. (2019; 2021) attempted to predict ENSO up to a lead time of 2 years, however they found a drop in the skill after 18 months of lead time. In this study we attempt to forecast ENSO events at a lead time of 2 years with skills higher than those reported in past studies using the CNN model.

2. Model and Methods

In the present study we attempt to forecast the Nino3.4 index at 18 to 23 months lead time using the CNN model. The CNN framework used in this study is similar to the one adopted by Ham et al. (2019) but with an addition of Hyper-Parameters (*HyPrs*) optimization. In this study we do not use the process of transfer learning as in Ham et al. (2019), as the CMIP models used in the transfer learning have biases (Kim et al. 2020b) and ENSO events are generally not well simulated in the CMIP models (Bellenger et al. 2013).

The inputs to the CNN model in this study are i) SSTA from the COBE analysis (Ishii et al. 2005) and ii) vertically averaged temperature anomalies (VATA; averaged temperature anomalies from 0 to 300m depth) from the SODA reanalysis data (Carton and Giese 2008). The sensitivity of the model to the source of the specified SSTA was checked by replacing the COBE SSTA with the SODA SSTA. Past three monthly global spatial fields of SSTA and VATA are considered as predictors to CNN lagged in the past by 18 to 23 months from the target Nino3.4 index season. For example, predictors of target season DJF-1997 at a lead time of 18 months are from May-Jun-Jul 1996, at a 19-month lead time are from Apr-May-Jun 1996, and so on.

CNN's were trained for a period of 101 years (1871-1973) and tested with Global Ocean Data Assimilation System (GODAS) data (Behringer and Xue 2004) for a period of 34 years (1984-2017). Both training and testing data were re-gridded to a 5°x5° spatial resolution using nearest neighbor method. GODAS data was kept aside of training and validation phases of CNN model development and a gap of 10 years was kept in the validation and testing phase to avoid the bias due to ocean memory persistence. This is done in order to verify the robustness of the method (cf. Ham et al. 2019). Hyperbolic-tangent (*tanh*) activation function is used for all convolutional and dense layers except the last output layer. Separate CNN models were developed for each season, as there is high seasonal variability in ENSO. However, attempts using a unified model to predict ENSO at all seasons together have shown poor performance compared

to the seasonal models (Ham et al. 2021). We use two fully connected layers in the model after convolutional layers for better feature extraction (Matsuoka et al. 2018). The dimension reduction in the model was done by using strides/average pooling. Various *HyPrs* were used to generate an ensemble of forecasts. We use *HyPrs* optimization (Table 1 shows the various *HyPrs* and their bounds) using random search method (Bergstra and Bengio 2012).

Table 1: Various *HyPrs* of CNN and their bounds used for ENSO forecasting

Hyper-parameter	Lower bound	Upper bound
Convolutional filter size - 1st layer	(3,3), (5,5), (3,5), (7,7), (4,8)	
Convolutional filter size - 2nd, 3rd layer	(1,1), (3,3), (1,4), (3,3), (2,4)	
Convolutional filters at 1st layer- doubled after every layer	8	48
Fully connected layers neurons	8	512
Drop out	0.2	0.75
L2 regularization	0.000001	0.001
Learning rate	0.00005	0.01
Convolutional and FC Layer initialization	Uniform, Gaussian and Constant	
Batch size	5 to 100% of train data	
Optimizer	RMSProp, Nadam, Adam	

3. ENSO forecasting skill

Several model experiments with different *HyPrs* were carried out. The performance of each model was tested and an ensemble average of the top ten performing models is presented in this study. The skill of predicted Nino3.4 index was evaluated by using the *correlation coefficient* (*'r'*) and *root mean square error* (*RMSE*) between the CNN model prediction and GODAS data during 1984 to 2017. Fig. 1 shows the correlation coefficients for ensemble mean of CNN's over all the seasons (reconstructed from seasonal CNN models) at 18 to 23 lead months. In Fig. 1 we have compared the *'r'* from proposed CNN with Scale Interaction Experiment-Frontier ver. 2 (hereafter referred as SINTEX-F2) 3DVAR dynamical model ensemble mean (Doi et al. 2017). It is evident from Fig. 1, that the *'r'* values of SINTEX-F2 are lower by about 0.2 at lead times of 18 to 23 months. The proposed CNN also exhibits the higher *'r'* compared to those reported in Ham et al. (2019 and 2021), approximately by a value of 0.1 to 0.2. Fig. 1 also suggests the model to be insensitive to the source of input SSTA. The seasonal *'r'* for the proposed CNN model and its comparison with SINTEX-F2 is highlighted in Fig. 2 in the form of forecast target season *vs* lead time. The *'r'* of SINTEX-F2 is greater than 0.5 for only few target seasons and a spring predictability barrier is evident in Fig 2. Whereas the proposed CNN has *'r'* greater than 0.5 at all the lead times and target seasons (except for few seasons). The CNN model seems to be less affected by the spring predictability barrier (Fig. 2).

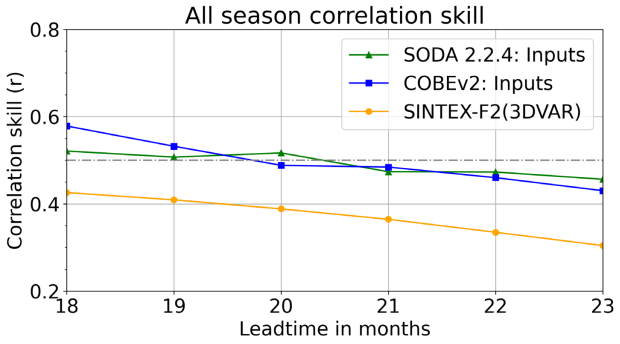

Figure 1: All season *'r'* (reconstructed from individual seasonal CNN models) of proposed method with SODA and COBE inputs. The validation is performed on GODAS inputs for a period of 1984 to 2017.

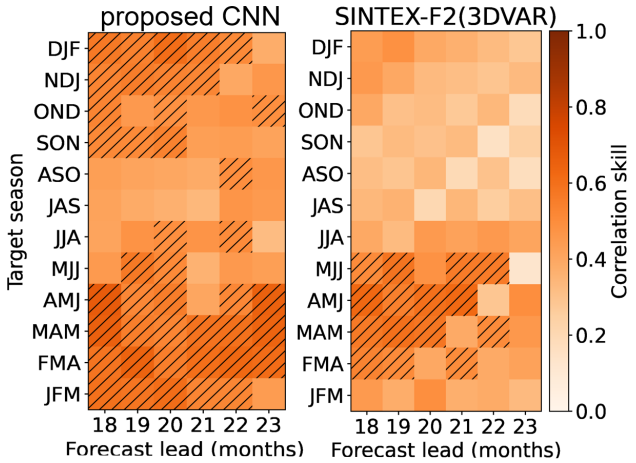

Figure 2: Variation in seasonal *'r'* of the proposed method (left) with SODA reanalysis inputs and its comparison with SINTEX-F2 (right) for a leadtime of 18 to 23 months. The validation is performed on GODAS inputs for a period of 1984 to 2017. Hatching highlights the forecasts with *'r'* exceeding 0.5.

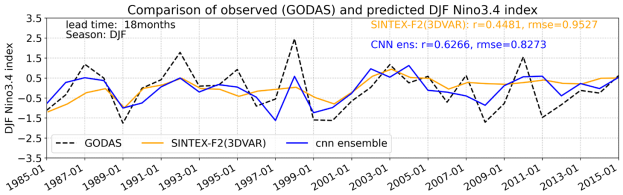

Figure 3: Nino3.4 DJF 18-month leadtime predicted index (from MJJ-1998) by proposed CNN model (blue) and its comparison with SINTEX-F2 (orange). The validation is performed on GODAS Nino3.4 index for a period of 1984 to 2017.

In addition, to compare the performance of the proposed method during the peak ENSO season (DJF), we compare the DJF predicted index of the CNN model and SINTEX-F2 ensemble mean with GODAS Nino3.4 index in Fig. 3. CNN models exhibit much higher *'r'* and comparatively low RMSE for the DJF predicted index compared to SINTEX-F2 as seen in Fig. 3. This is due to the fact that the proposed method is less affected when forecasts are initiated from spring season (Fig. 2). Along with deterministic skills,

we also evaluated the fidelity of the CNN model to provide warnings on specific events. The observed Nino3.4 index was segregated into terciles, and based on the occurrence of an event in respective tercile a contingency matrix was prepared to give *hit rate (HR)*, ratio of hits to number of events occurred, and *false alarm rate (FAR)*, ratio of false alarms to number of non-occurrence of event, as per Mason and Graham, (1999). It is evident from Table 2 that the CNN model has higher hit rates and lower false alarm rate compared to SINTEX-F2 over most of the terciles.

Table 2: *HR* and *FAR* comparison between CNN and SINTEX-F2 for lead time of 18 and 23 months (Bold values indicate respectively better performance)

Tercile	HR/FAR	Upper	Middle	Lower
CNN(18mn)	HR	**0.67**	**0.55**	**0.73**
	FAR	**0.19**	**0.22**	**0.13**
SINTEX-F2(18mn)	HR	0.64	0.50	0.50
	FAR	0.20	0.24	0.24
CNN(23mn)	HR	0.42	**0.37**	**0.64**
	FAR	0.32	**0.30**	**0.17**
SINTEX-F2(23mn)	HR	**0.45**	0.20	0.30
	FAR	**0.30**	0.38	0.33

4. Final remarks

In the present study we used CNN model to predict ENSO at long lead time of 23 months. The CNN models were trained using COBE and SODA re-analysis data on a period of ~101 years. The *HyPrs* of CNN were optimized using random search method. The proposed method was evaluated using deterministic skills as well as HR/FAR scores. Compared to a few past studies using dynamical and deep learning approaches, the proposed method shows higher skills with an all-season correlation skill nearly equal to 0.45 at 18 to 23 months lead time, and 0.6266 for the peak season of DJF at 18 months lead time. Apart from that, HR/FAR scores suggest better efficiency in providing a warning for important events. Analysis also shows that though, the proposed method has higher skill scores, it fails to capture the amplitude of the ENSO events realistically. We will be addressing the issue in future studies.

Data availability

Data related to this study can be downloaded from: COBE SST, https://psl.noaa.gov/data/gridded/data.cobe.html; SODA version 2.2.4, https://climatedataguide.ucar.edu/climate-data/soda-simple-ocean-data-assimilation; GODAS, https://www.esrl.noaa.gov/psd/data/gridded/data.godas.html.

References

Aguilar-Martinez, S., and W. Hsieh, 2009: Forecasts of Tropical Pacific Sea Surface Temperatures by Neural Networks and Support Vector Regression. *International Journal of Oceanography*, **2009**, 1-13. https://doi.org/10.1155/2009/167239.

Barnston A.G., M. H. Glantz, and Y. He, 1999: Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997–98 El Niño episode and the 1998 La Niña onset. *Bull. Amer. Meteor. Soc.*, **80**, 217–243. https://doi.org/10.1175/1520-0477(1999)080%3C0217:PSOSAD%3E2.0.CO;2.

Behringer D., and Y. Xue, 2004: Evaluation of the Global Ocean Data Assimilation System at NCEP: The Pacific Ocean. Preprints, Eighth Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface, Seattle, WA, *Amer. Meteor. Soc.*, 2.3. http://ams.confex.com/ams/84Annual/techprogram/paper_70720.htm.

Bellenger, H., E. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard, 2013: ENSO representation in climate models: from CMIP3 to CMIP5. *Climate Dynamics*, **42**, 1999-2018. https://doi.org/10.1007/s00382-013-1783-z.

Bergstra, J., and Y. Bengio, 2012: Random search for hyper-parameter optimization. *Journal of machine learning research*, **13**, 281-305. https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf.

Broni-Bedaiko, C., F. Katsriku, T. Unemi, M. Atsumi, J. Abdulai, N. Shinomiya, and E. Owusu, 2019: El Niño-Southern Oscillation forecasting using complex networks analysis of LSTM neural networks. *Artificial Life and Robotics*, **24**, 445-451. https://doi.org/10.1007/s10015-019-00540-2.

Cachay, S. R., E. Erickson, A. F. C. Bucker, E. Pokropek, W. Potosnak, S. Osei, and B. Lütjens, 2020: Graph Neural Networks for Improved El Nino Forecasting. arXiv preprint arXiv:2012.01598.

Carton, J.A. and B. Giese, 2008: A Reanalysis of Ocean Climate Using Simple Ocean Data Assimilation SODA. *Mon. Weath. Rev.*, **136**, 2999-3017. https://doi.org/10.1175/2007MWR1978.1.

Doi, T., S. Behera, and T. Yamagata, 2016: Improved seasonal prediction using the S INTEX-F2 coupled model. *Journal of Advances in Modeling Earth Systems*, **8**, 1847-1867, https://doi.org/10.1002/2016ms000744.

Doi, T., A. Storto, S.K. Behera, A. Navarra, and T. Yamagata, 2017: Improved prediction of the Indian Ocean dipole mode by use of subsurface ocean observations. *Journal of Climate*, **30**(19), 7953-7970. https://doi.org/10.1175/JCLI-D-16-0915.1.

Guo, Y., X. Cao, B. Liu, and K. Peng, 2020: El Niño Index Prediction Using Deep Learning with Ensemble Empirical Mode Decomposition. *Symmetry*, **12**, 893, https://doi.org/10.3390/sym12060893.

Ham, Y., J. Kim, and J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568-572. https://doi.org/10.1038/s41586-019-1559-7.

Ham, Y., J. Kim, E. Kim, and K. On, 2021: Unified deep learning model for El Niño/Southern Oscillation forecasts by incorporating seasonality in climate data. *Science Bulletin*, **66**, 1358-1366. https://doi.org/10.1016/j.scib.2021.03.009.

Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto, 2005: Objective Analyses of Sea-Surface Temperature and Marine Meteorological Variables for the 20th Century using ICOADS and the Kobe Collection. *Int. J. Climatol.*, **25**, 865-879. https://doi.org/10.1002/joc.1169.

Kim, Y., S. Min, X. Zhang, J. Sillmann, and M. Sandstad, 2020a: Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. *Weather and Climate Extremes*, **29**, 100269. https://doi.org/10.1016/j.wace.2020.100269.

Kim, Y. et al., 2020b: Publisher Correction: Malaria predictions based on seasonal climate forecasts in South Africa: A time series distributed lag nonlinear model. *Scientific Reports*, **9**. https://doi.org/10.1038/s41598-020-58890-y.

Levine, A., and M. McPhaden, 2015: The annual cycle in ENSO growth rate as a cause of the spring predictability barrier. *Geophysical Research Letters*, **42**, 5034-5041. https://doi.org/10.1002/2015gl064309.

Luo, J., S. Masson, S. Behera, and T. Yamagata, 2008: Extended ENSO Predictions Using a Fully Coupled Ocean–Atmosphere Model. *Journal of Climate*, **21**, 84-93. https://doi.org/10.1175/2007jcli1412.1.

Mahesh, A., M. Evans, G. Jain, M. Castillo, A. Lima, B. Lunghino, ... and V. Balaji, 2019: Forecasting El Niño with Convolutional and Recurrent Neural Networks. In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, Canada, 8-14. https://www.climatechange.ai/papers/neurips2019/40/paper.pdf.

Mason, S. J., and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels, *Weather Forecasting*, **14**, 713–725.

Matsuoka, D., M. Nakano, D. Sugiyama, and S. Uchida, 2018: Deep learning approach for detecting tropical cyclones and their precursors in the simulation by a cloud-resolving global nonhydrostatic atmospheric model. *Progress in Earth and Planetary Science*, **5**. https://doi.org/10.1186/s40645-018-0245-y.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, 2019: Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**, 195-204. https://doi.org/10.1038/s41586-019-0912-1.

Ropelewski, C. F., and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation, *Mon. Wea. Rev.*, **115**, 1606–1626.

Tangang, F., W. Hsieh, and B. Tang, 1998: Forecasting regional sea surface temperatures in the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors. *Journal of Geophysical Research: Oceans*, **103**, 7511-7522. https://doi.org/10.1029/97jc03414.

Trenberth, K., 1997: The Definition of El Niño. *Bulletin of the American Meteorological Society*, **78**, 2771-2777. https://doi.org/10.1175/1520-0477(1997)078<2771:tdoeno>2.0.co;2.

Wang, S., L. Mu, and D. Liu, 2021: A hybrid approach for El Niño prediction based on Empirical Mode Decomposition and convolutional LSTM Encoder-Decoder. *Computers & Geosciences*, **149**, 104695. https://doi.org/10.1016/j.cageo.2021.104695.

Yan, J., L. Mu, L. Wang, R. Ranjan, and A. Zomaya, 2020: Temporal Convolutional Networks for the Advance Prediction of ENSO. *Scientific Reports*, **10**. https://doi.org/10.1038/s41598-020-65070-5.

Spatio-temporal Semantic Decoupling Network for Improved ENSO Forecasting

Hailun Luo[1,*], Weiguang Sang[1], Jiakun Zhao[1], Yifei Xu[1]

1. School of Software, Xi'an Jiaotong University, Xi'an, Shaanxi, China, 710049

*Corresponding author. Email address: helen0303@stu.xjtu.edu.cn

I. Introduction

El Niño-Southern Oscillation (ENSO) is an irregular periodic variation in air pressure and sea surface temperatures among other atmospheric and oceanographic parameters over the tropical eastern Pacific Ocean, which not only causes wide-range catastrophic weather, but also poses serious harm to public safety, long-term effective ENSO prediction is therefore of great value. Numerous models, typically dynamic as well as machine learning ones, have been proposed to simulate and predict the development trends of ENSO. However, despite decades of efforts, predicting ENSO at lead times of more than one year remains a major challenge, and correlation coefficients begin to decline significantly usually after 6 months. Even though some excellent models can predict ENSO more than 12 months in advance, the output of their models is only one month predictive value (Ham et al. 2019), which is not very efficient. Moreover, the prediction of ENSO events is influenced by the spatial and the temporal information simultaneously, the existing models usually only consider the spatial or temporal effects, and use CNN-only or RNN-only models to extract features, which does not capture the semantic information of both two dimensions at the same time.

To solve the above problem, we propose Spatio-temporal Semantic Decoupling Network (SDNet) for improved ENSO forecasting. We evaluate our model on real dataset and the results demonstrate that SDNet is superior to other neural network models.

II. Methodology

A. Setup of ENSO Forecasting Problem

El Niño-Southern Oscillation (ENSO) prediction can be formulated as a spatiotemporal sequence prediction problem. As with the previous relevant work, we choose the oceanic Niño index (ONI) as the indicator of ENSO event. The goal of ENSO forecasting is to build an end-to-end deep learning model based on historical climate observation and simulation data to forecast future ENSO events.

If we record meteorological factors (e.g., sea-surface temperature) in Niño3.4 region in chronological order, we will get a tensor sequence $X_1, X_2, ..., X_t$ representing predictor and a tensor sequence $Y_{t+1}, Y_{t+2}, \ldots, Y_{t+M}$ representing indicator. Hence the prediction problem can be described as using spatiotemporal sequences (meteorological factors) over the past J month at t time to predict the Niño3.4 index in the next M months. Training task can be expressed as follows:

$$\hat{\theta} = arg \max_{\theta} P(Y_{t+1}, ..., Y_{t+M} | X_{t-J+1}, ..., X_t; \theta) \quad (1)$$

Where θ is the parameter of model and $\hat{\theta}$ is the optimal parameter.

B. Proposed Method

The purpose of this study is to leverage the semantic relationship between temporal and spatial to obtain the ENSO prediction over long time horizon. Fig.1 illustrates the pipeline of the proposed Spatio-temporal Semantic Decoupling Network (SDNet) which consists of two modules: SpatioTemporal Convolutional Module (STCM) and Threestream Temporal Scale Module (3sTSM). Firstly, the meteorological factors are fed into STCM which is composed of several Convolution Units (CU). For each CU, it alternately uses Temporal Convolutional Block (TCB) and Spatial Convolutional Block (SCB) to fully collect spatiotemporal semantic information features F_{st}. Subsequently, the F_{st} is sent to 3sTSM after global average pooling. Each stream yields hidden state features h_{t1} , h_{t2} , h_{t3} containing multiple time scales. Finally, these three features are connected to get the feature F_e, which contains multiscale spatial and temporal semantic information. The fully connected (FC) layer accepts F_e as input and eventually outputs ONI for M months. Next, we will introduce the proposed network in detail.

The key assignment of our work is to adequately extract temporal and spatial semantic information. To this end, we use nine Convolution Units (CU) to fully encode spatiotemporal semantics. In each CU, for the analysis of temporal dependencies, we employ temporal convolution which in our work can be regarded as the 1D convolutional neural network (1D CNN) to process the semantics of time dimension. As for spatial-correlated relationship, the data of local region may affect each other, which can be effectively dealt with by convolutional neural network (CNN) since it has powerful ability to hierarchically capture the spatial structural information. Therefore, we alternatively apply Temporal Convolutional Block

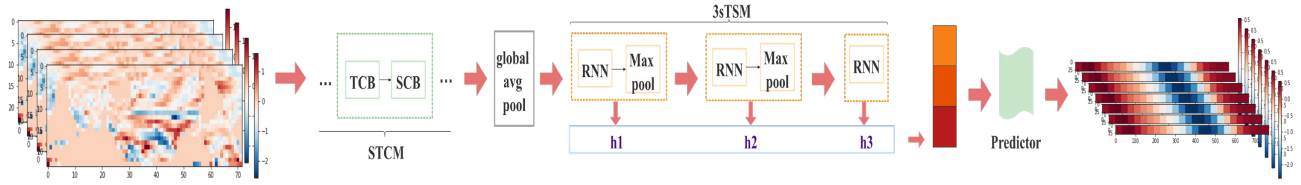

Figure 1: The pipeline of the proposed SDNet. Meteorological factors are fed to STCM to extract sufficient spatiotemporal semantic information. The last CU is connected to 3sTSM which is used to collect multiscale temporal semantics.

(TCB) and Spatial Convolutional Block (SCB) to capture the propagation of information flow in space and time.

To explore richer temporal semantics, we design 3sTSM to extract long-range, multiscale temporal information, which consists of three RNN blocks, and two pooling layers that change the temporal dimension step by step. For time dependencies, meteorological factors are regulated by both recent months and interannual climate changes. In addition, different time scales express distinct semantics. Therefore, we apply recurrent neural network (RNN) block to deal with time dependencies. Then the receptive field of the output is expanded by the Pooling layer, the result subsequently serves as input to the next RNN block. Note that the last hidden state h_{tn} of each RNN is temporarily stored. Finally, we connect the hidden layer features h_{t1} , h_{t2} , h_{t3} of the three RNN blocks and are able to obtain temporal semantic information with long-range and multiscale.

III. Experiment

A. Dataset and Evaluation Metrics

Historical Climate Observation and Simulation Dataset is provided by the Institute for Climate and Application Research (ICAR), which includes historical simulation data by CMIP5/6 model and historical assimilated observation data by American SODA model in nearly 100 years. Each sample has the following weather and spatiotemporal predictor: abnormal sea surface temperature (SST), abnormal thermal content (T300), abnormal zonal wind (Ua), and abnormal meridional wind (Va). The data dimension includes (*year, month, lat, lon, 4*). Year in each data sample shows the beginning year of the data. Month in each data sample shows matching month of the data, it is 36 for training data which include historical data for three years in a row since the beginning of the year. The 3rd and 4th dimension of each sample are longitude and latitude data over 0°– 360°E, 55°S – 60°N, and sampling every 5 degrees separately. The last dimension is the predictive factor. The label of each sample is the ONI for the corresponding month. For test data, there are 527 samples randomly selected from multiple global sea data, Month in each data sample is 12, and the rest are the same as the train data. The label of each test sample is the ONI of the next two years.

According to this dataset, we choose CMIP5/6 model for training, SODA model for validation, and multiple global sea data for testing in the experiment. The task is to predict the ONI index for the next M = 24 months using meteorological factors of current t time and the past 11 months. To evaluate the performance of our model, we calculate the Pearson Correlation Coefficient (PCC) and the Root Mean Square Error (RMSE) between the resulting predicted and actual values (Yan et al. 2020), The metrics are formulated as follows:

$$PCC = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sum_{i=1}^{n}(P_i - \bar{P})^2}} \tag{2}$$

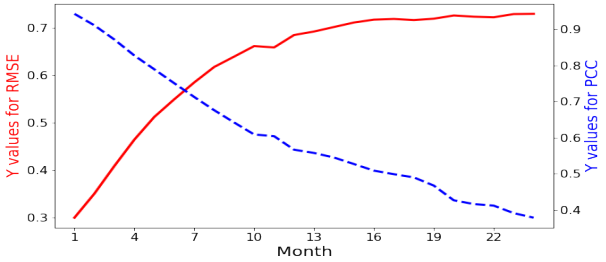

Figure 2: The RMSE and PCC values between the predictions and real labels in the SDNet-based ONI forecasting

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - P_i)^2} \tag{3}$$

Here, Y and $\bar{Y}$ denote the actual values and its mean value, P and $\bar{P}$ denote the predicted values and its mean value, n is the length of samples, respectively.

In order to evaluate the performance of the model more intuitively and verify its effectiveness in longer prediction horizon, we quantitatively score the forecast results based on PCC and RMSE. The calculation formula is:

$$Score = \frac{2}{3} \times PCCskill - RMSEskill \tag{4}$$

where:

$$PCCskill = \sum_{j=1}^{24} a \times \ln(j) \times PCC_j \tag{5}$$

$$a = \begin{cases} 1.5 & j \le 4 \\ 2 & 5 \le j \le 11 \\ 3 & 12 \le j \le 18 \\ 4 & 19 \le j \end{cases} \tag{6}$$

$$RMSEskill = \sum_{j=1}^{24} RMSE_j \tag{7}$$

Here, the longer the forecast time horizon, the higher the proportion it accounts of the score, and the higher scores mean more stable model performance.

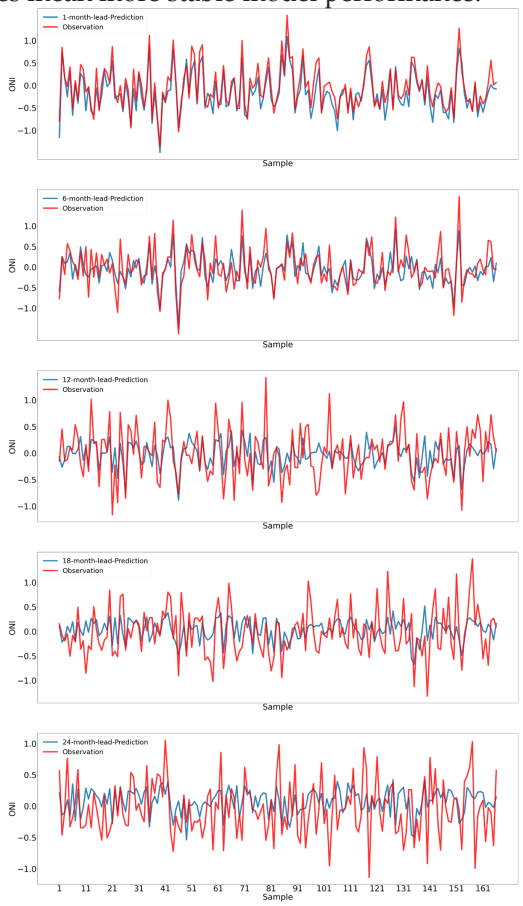

Figure 3: Predicted and real values of the SDNet-based ONI for different time horizon.

B. Results and Discussion

To validate the effectiveness of our proposed method on ENSO forecasting, we evaluate our SDNet with some excellent methods including CNN (Ham et al. 2019) and EEMD-TCN (Yan et al. 2020) on Historical Climate Observation and Simulation Dataset, where our model significantly outperforms other methods.

The prediction results of our model and its evaluation results are shown in Fig. 2 and 3, which demonstrate (1) All predicted curves and truth values have the same fluctuation and inflection points. (2) With the increase of forecast horizon, RMSE is gradually increasing and PCC is decreasing overall. (3) The 1-month predicted time horizon has the highest correlation and the least deviation, where PCC is 0.9225 and RMSE is 0.3202. (4) The curve obtained from the 24-month predicted time horizon is least matched with the real curve, where PCC is 0.3582 and RMSE is 0.7488 respectively. The predicted value has the same growth trend of the real value, though the amplitude is lower. (5) There are 16 forecast months with correlation coefficient exceeding 0.5 and the deviations are within acceptable range.

Fig. 4 and 5 show the comparison of SDNet-based, SCB-based and TCB-based CNN-based and EEMD-TCN-based ENSO predictions on RMSE and PCC. From these figures: (1) As the prediction time increases, the RMSE of all models gradually rises and the PCC gradually decreases. (2) For RMSE metric, SDNet is a little worse than the CNN at lead times of 6 to 15 months, probably due to the limitations of the model prediction. (3) For PCC metric, SDNet is superior to other networks especially at lead times of 1 to 9 months and 13 to 24 months, and slightly better than the CNN at lead times of 9 to 13 months.

Since STCM is the core module of our network, here we also verify the effectiveness of TCB and SCB. Overall, TCB performs better on this task than SCB, especially in the longer prediction horizon, which indicates that the semantic of time are more significant than space in later stage.

Moreover, we also calculate the *Score* of these five methods according to Eq.(4) to directly compare the effectiveness of these methods. It is noteworthy that SDNet has multi scale receptive field both in spatial and temporal dimensions, and the results verify the advancement of our method. Table 1 reveals that our model outperforms existing deep learning methods by a sizable margin.

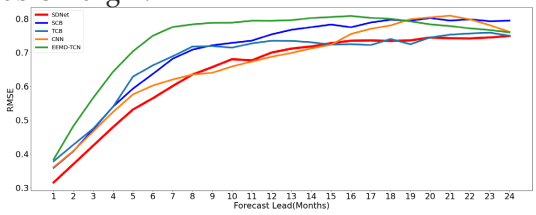

Figure 4: RMSE comparison of predictions and real labels of ONI obtained using SDNet, SCB, TCB, CNN, and EEMD-TCN methods.

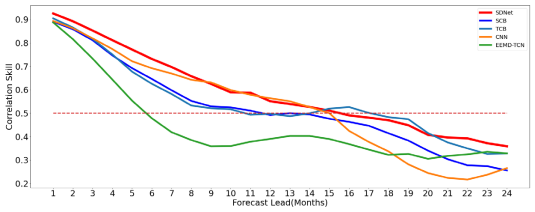

Figure 5: PCC comparison of predictions and real labels of ONI obtained using SDNet, SCB, TCB, CNN, and EEMD-TCN methods.

Table 1. Score Comparison of Different.

Methods	Score
SDNet	**38.784**
TCB	34.667
SCB	30.609
CNN	28.358
EEMD-TCN	22.357

IV. Conclusion

In our work, we propose two modules to provide the efficient and long time horizon ENSO prediction: the Spatiotemporal Convolutional Module profoundly gathers the spatiotemporal semantic information

of the meteorological factors and the Three-stream Temporal Scale Module further collects multi-scale temporal semantics. By coupling these methods, we derive SDNet, a powerful feature extractor that captures multiscale spatiotemporal features neglected by single meteorological factor modeling. After experimenting with the real dataset, we show that our model is superior to the existing methods in a considerable range, not only providing effective ENSO predictions, but also expanding the time horizon of each prediction.

Our model is an experiment on an integrated dataset which can serve as a baseline for subsequent related work.

Acknowledgment

Thanks are due to the Institute for Climate and Application Research (ICAR) for provision of dataset.

Refferences

Ham, Y.G., J.H. Kim, and J.J. Luo, 2019. Deep learning for multi-year enso forecasts. *Nature*, **573** (7775): 568–572.

Yan, J., L. Mu, L. Wang, R. Ranjan, and A.Y. Zomaya, 2020. Temporal convolutional networks for the advance prediction of ENSO. *Scientific Reports*, **10**(1): 8055.

ENSO Deep Learning Forecast Model: A Survey

Bin Mu[1], Bo Qin[1], Shijin Yuan[1*], Yuehan Cui[1], Guansong Wang[1], Xiaoyun Qin[1], Xin Feng[1]

1. School of Software Engineering, Tongji University, Shanghai, China
* Corresponding author (yuanshijin@tongji.edu.cn)

I. Introduction

ENSO (El Niño-Southern Oscillation) (Philander 2018) is one of the most dominant ocean-atmosphere cycles around the world, which has profound impacts on multiple fields such as climate change, socio-economy and sustainability (Patz 2002; McPhaden et al. 2006; Glantz 2001; Chen et al. 2017). Since the 1980s, there has been a major bloom in ENSO research (e.g. McCreary 1981; Philander et al. 1987; Neelin et al. 1992). Meteorological researchers have been arduously devoted to studying dynamical process analysis (Ren et al. 2020), physical mechanism (Wang et al. 2017a) and forecasting (Zhang et al. 2020) of ENSO events, and have made considerable progress.

The conventional approach to investigate ENSO is to construct dynamical partial differential equations to simulate the multivariate propagation of ENSO amplitudes, based on which the numerical climate model can be established (Luo et al. 2008; Tang et al. 2018). However, it is worth noting that the prediction skill of traditional numerical climate models for ENSO has gradually dropped, especially since the beginning of the 21st century. This is primarily because the numerical climate models are constructed based on the oceanic-atmospheric characteristics in a specific period (Battisti 1988; Suarez and Schopf 1988; Jin 1997; Picaut et al. 1997; Weisberg and Wang 1997), while the climate migration of the past decades makes the numerical models no longer suitable representations for the current ocean-atmosphere status. Not to mention that there are many simplifications and approximations in the numerical climate model programming itself, or even crucial neglect (Jin et al. 2008; Bianucci et al. 2018), which leads to forecast errors.

In recent years, with the development of computing equipment and storage facilities, data-driven deep learning (DL) technology has been playing a promising role in many frontiers of scientific research. As for ENSO forecast, researchers have made a natural analogy between the ENSO index (SST pattern) forecast and time-series (spatial-temporal sequence) prediction problem, using transfer advanced computer vision (CV) techniques to tackle these challenges, which has produced excellent results (Ham et al. 2019; He et al. 2019; Geng and Wang 2021; Zhang et al. 2017). Benefiting from the structure-modularity and sample-efficiency, most ENSO DL forecast models can achieve high prediction skills at a lower computational cost (Ham et al. 2019; Zhang et al. 2019a), compared with developing the enormous time-/resource-consuming numerical climate model, especially when the latent physical processes involving ENSO are not comprehensively understood. This is also known as the black-box characteristic of the DL model (Rudin 2019). These novel features provide appreciable new and valuable insights into ENSO forecasting.

However, the requirement of ENSO forecast is not only pursuing the quantified accuracy, but exploring the causes of skill improvements, furthermore revealing and grasping the underlying nature law for ENSO dynamic mechanisms. The DL-only ENSO forecast models cannot be considered sufficient for scientific investigation and engineering applications. Therefore, the consensus of the research community is to enhance the interpretability of DL models. There are usually two novel ways to achieve this purpose: the first is to incorporate the prior ENSO dynamical knowledge into DL modeling process (Geng and Wang 2021; Ham et al. 2021; Rojo Hernández et al. 2020) (referred to as physics-informed DL models), which is the direct way to increase the credibility of DL models; the second is to couple the DL models into the existing numerical climate models to replace crucial procedures, referred to as intelligent numerical models, naturally retaining convinced physical mechanisms. In general, these two approaches both blur the boundary between statistical and dynamical methods and usually yield state-of-the-art results to a large extent, which requires interdisciplinary cooperation between computer and atmospheric science.

In this article, following the above clues, we first introduce the ENSO DL forecast models according to different forecast objects in ENSO research, including ENSO index (Nino3, Nino4 and Nino3.4) prediction and physical variable grid prediction. Then, we highlight some encouraging milestones of physics-informed ENSO DL forecast models. Subsequently, we give an insight into the intelligent numerical models. Finally, we provide a discussion about present progresses and challenges, and make future prospects on possible development directions for the new generation of ENSO forecast models.

II. Naïve ENSO DL Forecast Model

1. Time-series Forecast Model for Index prediction

Index prediction can be considered as the regression for El Niño indexes, such as Nino3, Nino4, and Nino3.4, by which the occurrence and intensity of

ENSO events can be measured. This can be modeled as a time-series forecast problem (TFP):

$$x_{M+1}, ..., x_{M+N} = F_{TFP}(x_1, ..., x_M) \quad (1)$$

where $X_1,...,X_M$ are the indexes in historical M timestamp and $X_{M+1},...,X_{M+N}$ are the future N timestamp indexes to be predicted, $F_{TFP}(\cdot)$ is the forecast system. Many favorable neural networks have already made accurate predictions 6, 9, and 12 months ahead for Niño indexes, for example, ensemble QESN (McDermott and Wikle 2017), BAST-RNN (McDermott and Wikle 2019) and LSTM (long short-term memory) (Broni-Bedaiko et al. 2019) are representative works. These models produce very similar Nino3.4 developments for the entire 2015/2016 ENSO cycle. Though the peak phase of ENSO is slightly underestimated during Dec. 2015, they display superiority to many other operational forecasts.

The major idea of these models is to utilize the recurrent neural network (RNN), which can mine the potential correlations in temporal memory and make future predictions deductively. Nowadays, an increasing number of improved RNN are being proposed, which absorbs more neurological opinions and customizes more interpretable structures. For example, deliberation gate, forget gate, and global and local memory gate are applied in an improved LSTM unit (Jiang et al. 2020). These structures can help extract more nonlinear temporal correlations and process longer-term information if applied to ENSO forecast.

2. Spatial-temporal Forecast Model for Grid Prediction

There is a simple way to transfer the method of index prediction to all grids over the Pacific Ocean or globe for ENSO prediction. That is, building the individual forecast model for every grid in the focused regions. Some above-mentioned ENSO index prediction models have also made such preliminary attempts. This approach is referred to as the isolation model in this article, and the advantage is that it can focus on the specific evolution characteristics of each grid and avoid the negative interference and redundant noise induced by other grids.

However, the isolation model is not a comprehensive approach for sophisticated ENSO events. On the one hand, the input of isolation model is usually confined to a single physical index (or predictor), which is obviously inconsistent with the physical ENSO multivariate air-sea coupling. On the other hand, the isolation model ignores the global characteristics of ENSO and the spatial-temporal teleconnections, which seriously hampers the understanding of multi-scale dynamical interactions. Therefore, in analogizing the video frame forecast, the researchers began to formulate ENSO grid prediction as a spatial-temporal forecast problem (STFP) and migrate advanced computer vision (CV) techniques to build models.
The scheme for ENSO grid prediction can be described as:

$$X_{M+1}, ..., X_{M+N} = F_{STFP}(X_1, ..., X_M) \quad (2)$$

which is similar to Eq. (1). It directly changes the input and output from the time-series scaler value to the spatial-temporal grid (2D or 3D) value, and correspondingly, $F_{STFP}(\cdot)$ utilizes other deep learning modules to extract spatial information like convolution network (CNN). CNN uses a shared kernel to scan the entire physical pattern by row and column like a sliding window (Krizhevsky et al. 2012). Here, the size of kernel is also regarded as the receptive field, which determines how far the spatial remote correlations in ENSO amplitudes will be considered in the model during scanning.

Many studies have established ENSO grid prediction model successfully. Mu et al. (2019) and He et al. (2019) both built a ConvLSTM (Convolution LSTM) (Xingjian et al. 2015) model to capture the spatial-temporal dependencies of ENSO SST patterns over multiple time horizons and obtain better predictions than traditional numerical models. The correlation skill for Nino 3.4 is beyond 0.75/0.65 within 6-/12-month lead time on average. The forecast skill does not drop much in the eastern equatorial Pacific and has only a slight decrease in the central equatorial Pacific with the increase of lead time. But overall, the performances of these deep learning models almost approach or even exceed the numerical climate model, which demonstrates that they can be adequate alternative tools for ENSO research.

Why can the convolution operator effectively improve the ENSO forecast skill? This is mainly due to spatial division and kernel sharing; the former means that the kernel only focuses on the adjacent features in the receptive field in a round of calculation, which holds the locality; the latter means that all the local spatial areas share one common kernel in the whole pattern, which holds the globality. Subsequently, the learned locality and globality contain the local and remote spatial correlations, which improves the physical consistency extraction in the space.

Besides CNN, some other similar DL modules can also be used to process spatial features, such as pooling operator (Ciresan et al. 2011), transformation convolution operator (Noh et al. 2015), and dilated convolution operator (Yu and Koltun 2015). In addition, the ConvLSTM is a CNN-RNN coupled DL model actually, which is specially designed to extract spatial-temporal characteristics simultaneously. Nowadays, it is mainstream to build such models to predict future patterns, such as PredRNN (Wang et al. 2017b, 2018), MotionRNN (Wu et al. 2021), PhyDNet (Guen and Thome 2020), and STMFANet (Jin et al. 2020). These structures should be customized in receptive fields, memory units and fore-propagation processes to adapt to appropriate scenarios for ENSO forecast.

During the transition from index prediction to grid prediction, another enhancement is also considered in

the ENSO DL forecast model. That is, using multiple physical variables together as the model input. The simplest way is to cascade different variables on the channel (He et al. 2019) and customize different loss functions for each variable in the model output from the perspective of meteorology (discussed in Section III). Theoretically and practically, this approach has further improved ENSO forecasting skills.

III. Physics-informed ENSO DL Forecast Model

Though the grid prediction brings the data-driven ENSO forecast forward by a big step, there seems to be a pitfall among such pure statistical DL models: A blind pursuit of the prediction accuracy without the comprehensive grasp of the underlying dynamical mechanisms, results in a lack of generalizability to out-of-sample scenarios and physical-inconsistency forecast results (Karpatne et al. 2017). This mainly lies on the black-box systems of DL models, which are internally agnostic.

To tackle this problem, incorporating prior knowledge into DL models hand-crafted (Reichstein et al. 2019) is an effective method to make the black box transparent to some extent, further raising the interpretability of the model and producing credible forecast results, which can be adaptively applied in both ENSO index and grid predictions.

For example, Ham et al. (2019) incorporate seasonality in a CNN model to predict ENSO events, resulting in the correlation skill for Nino3.4 over 0.75/0.65/0.5 within 6-/12-/17-month forecast respectively, which outperforms the current numerical predictions. Zheng et al. (2020) construct a pure satellite data-driven deep learning considering the multi-scale spatial characteristics to forecast the evolutions of tropical instability waves, which is closely related to ENSO phenomena, and obtain accurate and efficient results. The forecast errors are lower than 0.1°C at the 30th recursive step. These pioneering studies have achieved further progresses on model predictability and a long-term forecast across multiple timescales.

This kind of DL modeling method is also blooming in forecasting other weather and climate phenomena, such as typhoons (Zhang et al. 2019b) and the North Atlantic Oscillation (Mu et al. 2020a; Yuan et al. 2019), achieving great progress. In general, in order to raise the interpretability as much as possible, there are three common methods to incorporate the prior physical knowledge.

Data interpretability. To some extent, the upper limit of the performance of DL models is dependent on training samples. On the one hand, more and more targeted data pre-processing methods are applied to make the model focus on the meteorological meta features. For example, EOF or SVD decomposition (Björnsson and Venegas 1997), temporal stationarity and multivariate correlation analysis (Geiβ and Einax 1996) are the representative approaches to filter the important predictors that are highly correlated with the forecast. Furthermore, researchers tend to separate the features with different frequencies in the spatial-temporal series data by DWT (discrete wavelet transform) (Mu et al. 2020a) and EEMD (ensemble empirical mode decomposition) (Yuan et al. 2019), and establish individual models for simulating the evolutions of different frequencies, achieving the coordination of spatial-temporal and frequency domains. These methods can greatly help separate signal and noise, improving the models' attention on the major features. On the other hand, many DL data augmentation algorithms have been transferred for forecast model training (Weiss 2020; Zhang et al. 2020), which must specifically maintain the physical consistency of the generated data.

Structure interpretability. Due to the modularity of neural networks, it is convenient to construct a well-designed data flow in the models. The contextual cues lying in the spatial-temporal data are the most concerning issues, which is considered to dominate the development of ENSO. So it is proposed to enhance the ability to extract longer-term spatial-temporal correlations using multi-scale spatial-temporal memory extraction and feature fusion structures (Mu et al. 2021b), multivariate air-sea coupler for feature interaction (Mu et al. 2021a), and neural computing unit fused with partial differential equations (Raissi et al. 2019), etc. Apart from this, more flexible attention modules (Vaswani et al. 2017) are widely used in the various models, which helps the model to automatically find the most critical temporal memory, spatial connection, and variable contribution. On the other hand, the loss function of the model can also be incorporated into prior physics knowledge to enhance the physical consistency of the forecast. For example, conservation of flux (energy) (Yang et al. 2019), boundary conditions (Pfau et al. 2020) and intermediate variable constraints (Muralidhar et al. 2020), are effective approaches to regulate the evolution direction of the model during training. Constructing a physically interpretable model is a challenging frontier in current research, where an in-depth interdisciplinary of meteorology and computer science is urgently needed.

Result interpretability: More importantly, knowing the origin for the forecast skill improvement (or decline) of ENSO DL forecast models, can steadily increase understanding of nature's laws, and drawing conclusions from forecast results is always the most direct feedback to extract the interpretable information from deluges of geoscience data. During the optimization of a constructed neural network, the program of "automatic derivative" (Baydin et al. 2018) helps to find the training direction. According to this characteristic, there are two fundamental approaches that can be used to discover knowledge; the first is the salience analysis (Borji 2019), which can obtain the most significant area of the input variable via

gradient back-propagation, this is highly consistent with the concept of sensitive area in meteorological predictability and can help improve the operational target observation. The second is perturbation analysis (Goodfellow et al. 2014). By superimposing perturbations on the input of the DL models, such as Monte Carlo (Andrieu et al. 2003) and CNOP (conditional nonlinear optimal perturbation) (Mu et al. 2003) methods, the initial patterns that can most likely change the status of the original events can be found (i.e., optimal precursor, optimal growth initial errors, etc.). Furthermore, advanced interpretability methods exhibit promising potentials and should be migrated into meteorological research, such as counterfactual inference (Johansson et al. 2016), causal representation learning (Schölkopf et al. 2021; Yang et al. 2021), etc. These methods can provide a reliable theoretical basis for revealing the knowledge learned by the model.

IV. ENSO Intelligent Numerical Model

Intelligent numerical models utilize the advanced DL models to replace crucial components in original numerical models, which can fully blend the dynamic mechanisms and statistical effectiveness. As mentioned above, there have been many excellent numerical climate models that can provide high-resolution simulations and accurate forecasts. At the same time, to further improve the representations of smaller-scale crucial physical processes, lots of parameterizations have been proposed, such as cloud-system resolving model and planetary boundary-layer parameterization. However, besides the benefits of forecast skills, they also tremendously increase the time and computation consumption, and inevitably and empirically approximate the sophisticated dynamics and physical parameters (Zhang et al. 2005) due to incomplete knowledge and heterogeneity in the underlying processes, which introduces bias. Therefore, there have been many successful attempts to use the DL models to replace parameterizations in numerical models, taking full advantage of fastness and effectiveness (Wang et al. 2020; Bolton and Zanna 2019).

For example, Tao et al. (2020) build a neural network to optimize spatial representations of soil organic carbon (SOC) in a climate model, substantially correcting the overestimated SOC storage. Han et al. (2020) introduce a residual convolutional neural network as a new moist physics parameterization in a superparameterized general circulation model, accurately simulating the timing and intensity of convective events. In general, the DL models that are specifically constructed to replace parameterizations always consider the unambiguous physical processes and have reliable sampling strategies to collect adequate training sets, covering all the cases. In addition, such DL models can usually maintain high generalization and stability (Yuval and O'Gorman 2020) while improving efficiency, which are the major demonstrations of intelligent numerical models. Meanwhile, the DL models have also been widely coupled within other modules of numerical climate models, such as DL data assimilation (Pawar et al. 2020; Huang et al. 2021), DL forecast result correlation (Le et al. 2020; Han et al. 2021), DL climate downscaling (Mu et al. 2020b), etc.

Moreover, a general-purpose adapter is also required to fill the gap between two different engineering environments to maintain the data flow, which can conveniently integrate the Python-based DL models into the Fortran-based numerical models (i.e., Fortran-Python adapter).

Although there is little work specifically for ENSO studies, the existing researches have already indicated the significantly promising synergistic collaborations between computer science and meteorological studies, and will tap great potential into ENSO research.

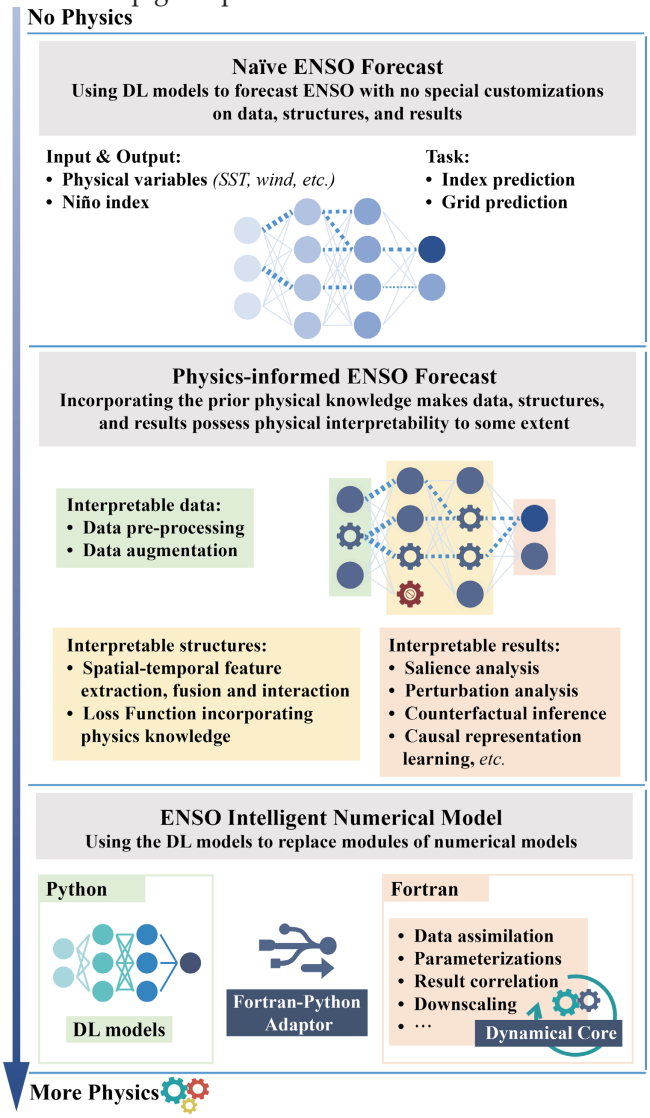

Figure 1: The current progress on ENSO forecasts using DL models from the purely data-driven naïve ENSO forecast method to the two statistical-dynamical coupling methods. One is the method based on statistics and supplemented by physics (the physics-informed ENSO forecast), another is the method based on physics and supplemented by statistics (the ENSO intelligent numerical model). They both effectively improve the forecast skills and physical interpretability to some extent.

V. Current Progresses

In Figure 1, we summarize the current progress on ENSO forecasts using DL models according to the above Sections. It firstly describes the purely data-driven ENSO forecast models (the naïve ENSO forecast), and then illustrates two statistical-dynamical coupling models (the physics-informed ENSO forecast and ENSO intelligent numerical model) via different forms of incorporating physics, which can effectively improve the forecast skills and interpretability of DL models.

Naïve ENSO forecast: This is the purely data-driven and the most fundamental ENSO DL forecast model firstly leading the wave. At this stage, the model is quite simple, and the researchers usually use the existing CNN or RNN directly without special customizations.

Physics-informed ENSO forecast: This is the model based on statistics and supplemented by physics, which incorporates the prior physical knowledge into data, structures, and results of ENSO DL forecast models, increasing the interpretability and reliability to some extent. For example, Mu et al. (2021a) design a multivariate air-sea coupler for ENSO DL forecast model mathematically according to the crucial ENSO dynamical mechanisms, which exhibits higher correlation skills of Niño indexes and physically consistent multivariate patterns. Mu et al. (2021b) use the saliency map method to investigate the subsurface precursor signals of different types of ENSO in a DL model, the results of which are consistent with numerical models. In the future, it is expected that advanced models and universal methods will be proposed to discover the spatial-temporal correlations, such as remote teleconnections and sensitive areas, and causality, etc.

ENSO Intelligent numerical model: This is the model based on physics and supplemented by statistics, which uses the DL models to replace crucial modules of numerical models, such as data assimilations, parameterizations, result correlations and downscaling. There are relatively few associated cases specifically for ENSO at present, and the development of ENSO intelligent models needs amounts of devotion. There are still many unknowns in climate changes, migrations, and varieties, which means that the ENSO intelligent numerical models should not only aim at improving efficiency, but also focus on maintaining the physical consistency and exploring potential physical mechanisms, such as accurately distinguishing the EP (eastern Pacific) and CP (central Pacific) types of El Niño events and reliably representing strong ENSO events. Besides the physics-informed models, this is a more vibrant research topic for applying DL models in ENSO forecasts.

VI. Final Remarks

In this article, we introduce the development footprints of the deep learning forecast model in ENSO research. Overall, the statistical-dynamical coupling methods, including the physics-informed DL models and intelligent numerical models, will further improve the ENSO forecast skills and lead the direction of ENSO research together in the future, laying the foundation for the development of a new generation of the climate forecasting systems.

Acknowledgments

This study is supported in part by the National Key Research and Development Program of China under Grant 2020YFA0608002, in part by the National Natural Science Foundation of China under Grant 42075141, in part by the Key Project Fund of Shanghai 2020 "Science and Technology Innovation Action Plan" for Social Development under Grant 20dz1200702, and in part by the Fundamental Research Funds for the Central Universities under Grant 13502150039/003.

References

Andrieu, C., N. De Freitas, A. Doucet, and M.I. Jordan, 2003: An introduction to MCMC for machine learning. *Machine learning*, **50**, 5–43.

Battisti, D.S., 1988: Dynamics and thermodynamics of a warming event in a coupled tropical atmosphere–ocean model. *Journal of Atmospheric Sciences*, **45**, 2889–2919.

Baydin, A.G., B.A. Pearlmutter, A.A. Radul, and J.M. Siskind, 2018: Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, **18**.

Bianucci, M., A. Capotondi, R. Mannella, and S. Merlino, 2018: Linear or nonlinear modeling for ENSO dynamics? *Atmosphere*, **9**, 435.

Björnsson, H., and S. Venegas, 1997: A manual for EOF and SVD analyses of climatic data. *CCGCR Report*, **97**, 112–134.

Bolton, T., and L. Zanna, 2019: Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, **11**, 376–399.

Borji, A., 2019: Saliency prediction in the deep learning era: Successes and limitations. *IEEE transactions on pattern analysis and machine intelligence*.

Broni-Bedaiko, C., F.A. Katsriku, T. Unemi, M. Atsumi, J.-D. Abdulai, N. Shinomiya, and E. Owusu, 2019: El Niño-Southern Oscillation forecasting using complex networks analysis of LSTM neural networks. *Artificial Life and Robotics*, **24**, 445–451.

Chen, C., M.A. Cane, A.T. Wittenberg, and D. Chen, 2017: ENSO in the CMIP5 simulations: Life cycles, diversity, and responses to climate change. *Journal of Climate*, **30**, 775–801.

Ciresan, D.C., U. Meier, J. Masci, L.M. Gambardella,

and J. Schmidhuber, 2011: Flexible, high performance convolutional neural networks for image classification. Twenty-second international joint conference on artificial intelligence.

Geiβ, S., and J. Einax, 1996: Multivariate correlation analysis-a method for the analysis of multidimensional time series in environmental studies. *Chemometrics and intelligent laboratory systems*, **32**, 57–65.

Geng, H., and T. Wang, 2021: Spatiotemporal Model Based on Deep Learning for ENSO Forecasts. *Atmosphere*, **12**, 810.

Glantz, M.H., 2001: Currents of change: impacts of El Niño and La Niña on climate and society. Cambridge University Press.

Goodfellow, I.J., J. Shlens, and C. Szegedy, 2014: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

Guen, V.L., and N. Thome, 2020: Disentangling physical dynamics from unknown factors for unsupervised video prediction. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11474–11484.

Ham, Y.G., J.H. Kim, and J.J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572.

Ham, Y.G., J.H. Kim, E.S. Kim, and K.W. On, 2021: Unified deep learning model for El Niño/Southern Oscillation forecasts by incorporating seasonality in climate data. *Science Bulletin*.

Han, L., M. Chen, K. Chen, H. Chen, Y. Zhang, B. Lu, L. Song, and R. Qin, 2021: A deep learning method for bias correction of ECMWF 24–240 h forecasts. *Advances in Atmospheric Sciences*, **38**, 1444–1459.

Han, Y., G.J. Zhang, X. Huang, and Y. Wang, 2020: A moist physics parameterization based on deep learning. *Journal of Advances in Modeling Earth Systems*, **12**, e2020MS002076.

He, D., P. Lin, H. Liu, L. Ding, and J. Jiang, 2019: Dlenso: A deep learning enso forecasting model. *Pacific Rim International Conference on Artificial Intelligence*, Springer, 12–23.

Huang, L., H. Leng, X. Li, K. Ren, J. Song, and D. Wang, 2021: A Data-Driven Method for Hybrid Data Assimilation with Multilayer Perceptron. *Big Data Research*, **23**, 100179.

Jiang, X., J. Yu, Y. Sun, Z. Qin, Z. Zhu, Y. Hu, and Q. Wu, 2020: DAM: Deliberation, Abandon and Memory Networks for Generating Detailed and Non-repetitive Responses in Visual Dialogue. arXiv preprint arXiv:2007.03310.

Jin, B., Y. Hu, Q. Tang, J. Niu, Z. Shi, Y. Han, and X. Li, 2020: Exploring spatial-temporal multi-frequency analysis for high-fidelity and temporal-consistency video prediction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4554–4563.

Jin, E.K., and Coauthors, 2008: Current status of ENSO prediction skill in coupled ocean–atmosphere models. *Climate Dynamics*, **31**, 647–664.

Jin, F.F., 1997: An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. *Journal of the atmospheric sciences*, **54**, 811–829.

Johansson, F., U. Shalit, and D. Sontag, 2016: Learning representations for counterfactual inference. *International conference on machine learning*, PMLR, 3020–3029.

Karpatne, A., and Coauthors, 2017: Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on knowledge and data engineering*, **29**, 2318–2331.

Krizhevsky, A., I. Sutskever, and G.E. Hinton, 2012: Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, **25**, 1097–1105.

Le, X.H., G. Lee, K. Jung, H. An, S. Lee, and Y. Jung, 2020: Application of convolutional neural network for spatiotemporal bias correction of daily satellite-based precipitation. *Remote Sensing*, **12**, 2731.

Luo, J.J., S. Masson, S.K. Behera, and T. Yamagata, 2008: Extended ENSO predictions using a fully coupled ocean–atmosphere model. *Journal of Climate*, **21**, 84–93.

McCreary, J., 1981: A linear stratified ocean model of the equatorial undercurrent. Philosophical Transactions of the Royal Society of London. *Series A, Mathematical and Physical Sciences*, **298**, 603–635.

McDermott, P.L., and C.K. Wikle, 2017: An ensemble quadratic echo state network for non-linear spatio-temporal forecasting. *Stat*, **6**, 315–330.

McDermott, P.L., and C.K. Wikle, 2019: Bayesian recurrent neural network models for forecasting and quantifying uncertainty in spatial-temporal data. *Entropy*, **21**, 184.

McPhaden, M.J., S.E. Zebiak, and M.H. Glantz, 2006: ENSO as an integrating concept in earth science. *Science*, **314**, 1740–1745.

Mu, B., C. Peng, S. Yuan, and L. Chen, 2019: ENSO forecasting over multiple time horizons using ConvLSTM network and rolling mechanism. 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 1–8.

Mu, B., J. Li, S. Yuan, and X. Luo, 2020a: Prediction of North Atlantic Oscillation Index Associated with the Sea Level Pressure Using DWT-LSTM and DWT-ConvLSTM Networks. *Mathematical Problems in Engineering*, **2020**.

Mu, B., B. Qin, S. Yuan, and X. Qin, 2020b: A Climate Downscaling Deep Learning Model considering the Multiscale Spatial Correlations and Chaos of Meteorological Events. *Mathematical Problems in Engineering*, **2020**.

Mu, B., B. Qin, and S. Yuan, 2021a: ENSO-ASC 1.0. 0: ENSO Deep Learning Forecast Model with a Multivariate Air–Sea Coupler. *Geoscientific Model Development Discussions*, 1–33.

Mu, B., Y.H. Cui, S.J. Yuan, 2021b: Spatiotemporal Prediction, Precursor Analysis and Sensitive Area Identification of ENSO using Deep Neural Network. *Journal of Advances in Modeling Earth Systems*. Submitted.

Mu, M., W.S. Duan, and B. Wang, 2003: Conditional nonlinear optimal perturbation and its applications. *Nonlinear Processes in Geophysics*, **10**, 493–501.

Muralidhar, N., J. Bu, Z. Cao, L. He, N. Ramakrishnan, D. Tafti, and A. Karpatne, 2020: PhyNet: Physics Guided Neural Networks for Particle Drag Force Prediction in Assembly. *Proceedings of the 2020 SIAM International Conference on Data Mining*, SIAM, 559–567.

Neelin, J., and Coauthors, 1992: Tropical air-sea interaction in general circulation models. *Climate Dynamics*, **7**, 73–104.

Noh, H., S. Hong, and B. Han, 2015: Learning deconvolution network for semantic segmentation. *Proceedings of the IEEE international conference on computer vision*, 1520–1528.

Patz, J. A., 2002: A human disease indicator for the effects of recent global climate change. *Proceedings of the National Academy of Sciences*, **99**, 12506–12508.

Pawar, S., S. E. Ahmed, O. San, A. Rasheed, and I. M. Navon, 2020: Long short-term memory embedded nudging schemes for nonlinear data assimilation of geophysical flows. *Physics of Fluids*, **32**, 076606.

Pfau, D., J.S. Spencer, A.G. Matthews, and W.M.C. Foulkes, 2020: Ab initio solution of the many-electron Schrödinger equation with deep neural networks. *Physical Review Research*, **2**, 033429.

Philander, S., W. Hurlin, and A. Seigel, 1987: Simulation of the seasonal cycle of the tropical Pacific Ocean. *Journal of Physical Oceanography*, **17**, 1986–2002.

Philander, S.G., 2018: 9. El Niño, La Niña, and the Southern Oscillation. Princeton University Press.

Picaut, J., F. Masia, and Y. du Penhoat, 1997: An advective-reflective conceptual model for the oscillatory nature of the ENSO. *Science*, **277**, 663–666.

Raissi, M., P. Perdikaris, and G.E. Karniadakis, 2019: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, **378**, 686–707.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and others, 2019: Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**, 195–204.

Ren, H.L., F. Zheng, J.-J. Luo, R. Wang, M. Liu, W. Zhang, T. Zhou, and G. Zhou, 2020: A review of research on tropical air-sea interaction, ENSO dynamics, and ENSO prediction in China. *Journal of Meteorological Research*, **34**, 43–62.

Rojo Hernández, J.D., Ó.J. Mesa, and U. Lall, 2020: ENSO dynamics, trends, and prediction using machine learning. *Weather and Forecasting*, **35**, 2061–2081.

Rudin, C., 2019: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, **1**, 206–215.

Schölkopf, B., F. Locatello, S. Bauer, N.R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio, 2021: Toward causal representation learning. *Proceedings of the IEEE*, **109**, 612–634.

Suarez, M.J., and P.S. Schopf, 1988: A delayed action oscillator for ENSO. *Journal of Atmospheric Sciences*, **45**, 3283–3287.

Tang, Y., and Coauthors, 2018: Progress in ENSO prediction and predictability study. *National Science Review*, **5**, 826–839.

Tao, F., and Coauthors, 2020: Deep learning optimizes data-driven representation of soil organic carbon in Earth system model over the conterminous United States. *Frontiers in big Data*, **3**, 17.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Łukasz Kaiser, and I. Polosukhin, 2017: Attention is all you need. *Advances in neural information processing systems*, 5998–6008.

Wang, C., C. Deser, J.-Y. Yu, P. DiNezio, and A. Clement, 2017a: El Niño and southern oscillation (ENSO): a review. *Coral reefs of the eastern tropical Pacific*, 85–106.

Wang, X., W. Wang, and B. Yan, 2020: Tropical Cyclone Intensity Change Prediction Based on Surrounding Environmental Conditions with Deep Learning. *Water*, **12**, 2685.

Wang, Y., M. Long, J. Wang, Z. Gao, and P.S. Yu, 2017b: PredRNN: Recurrent neural networks for predictive learning using spatiotemporal lstms. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 879–888.

Wang, Y., Z. Gao, M. Long, J. Wang, and S.Y. Philip, 2018: PredRNN: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. *Proceedings of the 35th International Conference on*

Machine Learning, PMLR, 5123–5132.

Weisberg, R.H., and C. Wang, 1997: A western Pacific oscillator paradigm for the El Niño-Southern Oscillation. *Geophysical research letters*, **24**, 779–782.

Weiss, B.M., 2020: Artificial enhancement of remote sensing data using generative adversarial networks. Utrecht University.

Wu, H., Z. Yao, J. Wang, and M. Long, 2021: MotionRNN: A flexible model for video prediction with spacetime-varying motions. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15435–15444.

Xingjian, S., Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W. Woo, 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 802–810.

Yang, M., F. Liu, Z. Chen, X. Shen, J. Hao, and J. Wang, 2021: CausalVAE: disentangled representation learning via neural structural causal models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9593–9602.

Yang, Z., J.L. Wu, and H. Xiao, 2019: Enforcing deterministic constraints on generative adversarial networks for emulating physical systems. arXiv preprint arXiv:1911.06671,.

Yu, F., and V. Koltun, 2015: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122,.

Yuan, S., X. Luo, B. Mu, J. Li, and G. Dai, 2019: Prediction of North Atlantic Oscillation index with convolutional LSTM based on ensemble empirical mode decomposition. *Atmosphere*, **10**, 252.

Yuval, J., and P. A. O'Gorman, 2020: Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature communications*, **11**, 1–10.

Zhang, H., P.S. Chu, L. He, and D. Unger, 2019a: Improving the CPC's ENSO Forecasts using Bayesian model averaging. *Climate Dynamics*, **53**, 3373–3385.

Zhang, Q., H. Wang, J. Dong, G. Zhong, and X. Sun, 2017: Prediction of sea surface temperature using long short-term memory. *IEEE Geoscience and Remote Sensing Letters*, **14**, 1745–1749.

Zhang, R., Q. Liu, and R. Hang, 2019b: Tropical cyclone intensity estimation using two-branch convolutional neural network from infrared and water vapor images. *IEEE Transactions on Geoscience and Remote Sensing*, **58**, 586–597.

Zhang, R.H., R. Kleeman, S. E. Zebiak, N. Keenlyside, and S. Raynaud, 2005: An empirical parameterization of subsurface entrainment temperature for improved SST anomaly simulations in an intermediate ocean model. *Journal of climate*, **18**, 350–371.

Zhang, R.H., and Coauthors, 2020: A review of progress in coupled ocean-atmosphere model developments for ENSO studies in China. *Journal of Oceanology and Limnology*, **38**, 930–961.

Zhang, Y., P. Chen, J. Huang, and Y. Chen, 2020: Enhancing Underwater Image Using Multi-scale Generative Adversarial Networks. International Symposium on Parallel Architectures, *Algorithms and Programming*, Springer, 259–269.

Zheng, G., X. Li, R.H. Zhang, and B. Liu, 2020: Purely satellite data–driven deep learning forecast of complicated tropical instability waves. *Science Advances*, **6**, eaba1482.

A climate downscaling method based on deep back-projection neural network

Jing Hu[1], Chuan Tian[1], Buhong, Ge[1], Xiaomeng Huang[2], Xi Wu[1]

1. Department of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China
2. Ministry of Education Key Laboratory for Earth System Modeling and Department of Earth System Science, Tsinghua University, Beijing, China

(Corresponding authors: hxm@tsinghua.edu.cn, xi.wu@cuit.edu.cn)

1. Introduction

Climate downs-scaling is a procedure to take information known at large scales to make predictions at local scales, which is an important procedure for weather and climate applications. Typically, climate downscaling methods nowadays propose to refine coarse resolution forecasts to meet a desired resolution. Generally, there are two types of downscaling methods: dynamical downscaling and statistical downscaling. Dynamical downscaling methods account for developing high-resolution general circulation models (GCMs) or regional climate models (RCMs) to simulate local meteorological processes, which include embedding sub-grid of key regions into GCM or using the output of GCM as the boundary conditions in RCMs for high-resolution projections. The merit of dynamical downscaling methods lies in the clear physical causalities but these methods are quite time-consuming and modifying GCMs to a certain resolution is not convenient.

Statistical downscaling methods learn the relationships between the low-resolution (LR) and high-resolution (HR) climate patterns. Regression is a popular way to estimate the mapping projections and has made preliminary progress in downscaling sea temperatures and precipitation (Hessami et al. 2008). However, Regression methods are limited by its linearity. In recent years, deep learning has gained popularity in climate science. For example, Rodrigues et al. (2018) explore climate downscaling using several convolution layers. But these methods mostly use basic and early-stage deep learning strategies with simple convolutional neural frameworks.

In this article, a deep-learning based method based on deep back-projection neural network (DBPNN) is proposed, in which an enhanced residual network structure is leveraged for retrieving fine-scale details in the reconstructed HR images.

2. The proposed method

As shown in Fig. 1, the network structure of this method is composed of three modules: the shallow feature extraction module, the back-projection residual module and the reconstruction module. In the following, these three modules are described.

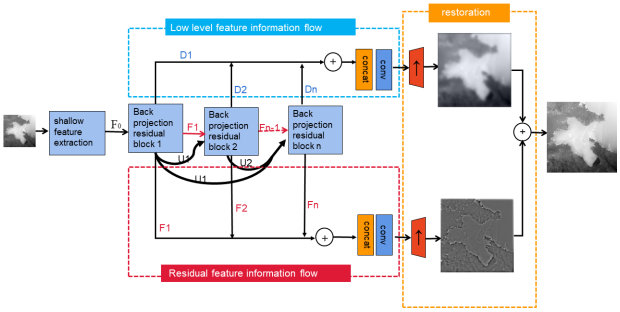

Figure 1: Framework of the proposed method

2.1 Shallow feature extraction module

The shallow feature extraction module is a light-weight network structure which adopts the dual-branch network structure instead of stacking too many layers in a simple chain way. The structure of the module is shown in Fig. 2.

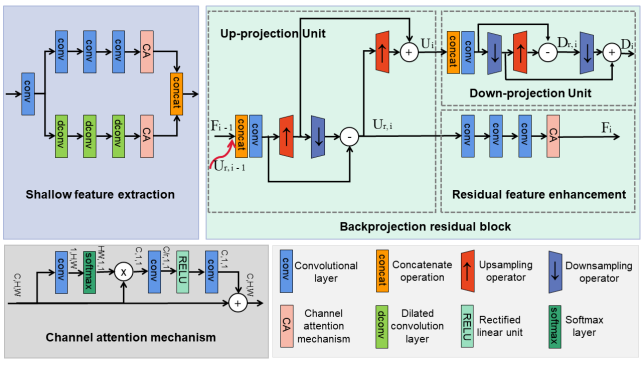

Figure 2: DBPNN specific structure information

The merit of this dual-branch network is its capability to obtain features from different receptive fields, enhancing the richness and diversity. One of the dual-branch is a plain network composed of several standard convolution layers and a channel attention, which is used to extract the most relevant pixels around the target one. The channel attention mechanism bypasses normalization operations, but uses convolution kernels to flexibly and adaptively learn the correlation between channels.

Another branch consists of several dilated convolution layers, which is used to expand the receptive field with as few layers as possible, providing more contextual information.

2.2 Deep back projection residual module

Back-projection residual module is the core of the proposed algorithm, which guides the network structure to focus on the recovery of residual information and effectively solves the problem of underfitting in the reconstruction of high-frequency details. It consists of three sub-units: up-projection, down-projection and residual feature enhancement.

The first up-projection unit uses the output of the shallow feature extraction module. Subsequently, to enhance the propagation and reuse of residual feature information, the inputs of the remaining up-projection units are the LR residual feature U_r, the output of the previous n-1 up-projection units and the enhanced LR residual feature F_{n-1}, the output of the previous residual feature enhancement unit.

The down-projection unit is used to obtain the low-frequency details, and the low-level feature flow. The nth down-projection unit takes the up-sampling reconstruction features, which are obtained by the first n-1 up-projection units, and outputs the LR features D_n after the down projection.

In addition, the LR residual feature output by the up-projection unit is passed into the residual enhancement unit. After the residual features passing through the channel attention mechanism, the network structure obtains more detailed information of the target that needs to be paid attention into. Thus, the enhanced LR residual feature flow is obtained.

The expression of back-projection residual module based on dense connection is shown in Eq. (1):

$$F_n, D_n = H_{RB,n}([U_{r,1}, \cdots U_{r,n-1}, F_{n-1}]) \quad (1)$$

where $H_{RB,n}(\cdot)$ is the back projection residual block function. D_n and F_n respectively represent the back-projection LR feature and enhanced LR residual feature output by the nth back-projection residual block. $U_{r,n}$ represents the LR residual feature of the up-projection unit of the nth back projection residual block.

2.3 Reconstruction module

This module separately upscales the residual feature flow and the low-level feature information flow, then adds the elements of the high and low frequency feature information after up-sampling, and obtains the HR image with fine details as follows:

$$I_{SR}=F+D \quad (2)$$

where D represents the reconstructed image with low frequency details, F represents the reconstructed image with high-frequency details, and I_{SR} represents the reconstructed HR image.

3. Experiment

3.1 Dataset

In this study, we use 12,000 simulation results of T2 data to test the effectiveness of the proposed method, which are obtained by running the WRF model (weather research & forecasting model). These data were within the range of Tianjin Port from 16.83157-124.031525°E to 35.909744-41.37973°N. We treat the low spatial resolution grid data (WRF; 9km; 1-hrly) with a size of 67×67 as the LR image and high spatial resolution grid data (WRF; 3km; 1-hrly) with the size of 201×201 as the HR image. Considering that the geographic range of the 9km data includes the 3km data, a pre-processing step of cropping was carried out on 9km data.

3.2 Implementation Details

Model Setting: In the Shallow feature extraction module, the sizes of all convolution kernels are fixed as 3×3, the rate of dilation of each layer is set in a specific order (e.g., 1, 3, 5, respectively). The input of this module is the LR image of 3 channels (RGB space) and the output is of 32 channels. 7 back projection residual modules are used and all the feature channels are set to 64. In our experiment, all up-sampling operations adopt the convolution layer with convolution kernel size 7×7, stride = 3 and padding = 2.

Training Setting: We randomly select 4800 HR-LR image pairs from the T2 dataset to train the proposed network structure. In addition, we select another 1200 image as testing data. Peak signal to noise ratio (PSNR), structural similarity index (SSIM) (Wang et al. 2004) and root mean square error (RMSE) are used as the objective evaluation standards. The images are randomly cropped as 192×192 patches and these patches are fed into the neural network with the batch size of 16. These experiments are run at the system of Ubuntu 16.04, and NVIDIA Corporation GP100GL. The learning rate is initialized as 1e-4 and halved every 200 epochs. To optimize our algorithm, we use L1 as the loss function and Adam optimizer (Kingma and Jimmy 2014) with momentum of 0.9.

3.3 Comparisons with the State-of-the-Art Methods

In order to verify the effectiveness of this algorithm, we compare our network with two state-of-the-art super-resolution algorithms: EDSR (Lim et al. 2017), ESRGAN (Wang et al. 2018) and use bicubic interpolation and kriging interpolation as the baseline. Table 1 shows the downscaling results. Fig. 3 shows the visual comparison of the experimental results

Table 1. Quantitative evaluation of down sampling algorithms on 3× magnification. Bold letter indicates the best results

Algorithms	PSNR	SSIM	RMSE
Bicubic	28.306	0.886	9.820
Kriging	28.426	0.829	9.685
EDSR	35.438	0.935	5.432
ESRGAN	34.030	0.951	5.308
DBPNN (Ours)	**37.973**	**0.961**	**4.209**

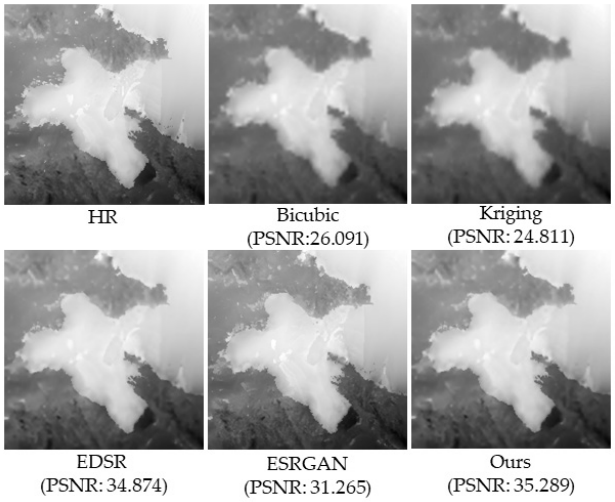

Figure 3: Visual comparison between different methods

4. Conclusion

This study presents a deep-learning based method to address the climate downscaling problem. A deep back-projection neural network is proposed to enhance the high-frequency information in the reconstructed HR image. Compared to the conventional downsampling methods such as bicubic interpolation and kriging interpolation, the proposed method has a significant improvement in terms of PSNR, SSIM and RMSE, which indicates the effectiveness of applying deep learning techniques in climate downsampling scenario.

Acknowledgments

This work is supported by National Key Research and Development Program of China (2020YFA0608001), National Natural Science Foundation of China under Grant 61602065, Sichuan province Key Technology Research and Development project under Grant 2021YFG0038, Sichuan Science and Technology Program(2020JDTD0020), National Science Foundation of China 42075142.

References

Hessami, M., P. Gachon, T.B. Ouarda and A. St-Hilaire, 2008: Automated regression-based statistical downscaling tool. *Environmental Modelling & Software*, **23**(6), pp.813-834.

Rodrigues, E.R., I. Oliveira, R. Cunha, and M. Netto, 2018: DeepDownscale: a deep learning strategy for high-resolution weather forecast. In *2018 IEEE 14th International Conference on e-Science (e-Science)*, pp. 415-422.

Wang, P., P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, 2018: March. Understanding convolution for semantic segmentation. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pp. 1451-1460.

Huynh-Thu, Q. and M. Ghanbari, 2008: Scope of validity of PSNR in image/video quality assessment. *Electronics letters*, **44**(13), pp.800-801.

Wang, Z., A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, 2004: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, **13**(4), pp.600-612.

Willmott, C.J. and K. Matsuura, 2005: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, **30**(1), pp.79-82.

Kingma, D.P. and J. Ba, 2014: Adam: A method for stochastic optimization. arXiv preprint arXiv: 1412.6980.

Keys, R., 1981: Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, **29**(6), pp.1153-1160.

Lim, B., S. Son, H. Kim, S. Nah, and K. Mu Lee, 2017: Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136-144.

Wang, X., K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, 2018: Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pp. 1-16.

Forecasting of Time Series Significant Wave Height based on ConvLSTM

Xiaojie Li[1], Lihui Zhang[1], Jiuke Wang[2], Xiaomeng Huang[3*], Qi Zhong[4], Xi Wu[1*]

1. Department of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China
2. National Marine Environmental Forecasting Center, Beijing, China
3. Ministry of Education Key Laboratory for Earth System Modeling and Department of Earth System Science, Tsinghua University, Beijing, China
4. China Meteorological Administration Training Center, Beijing, China
*Corresponding authors

1. Introduction

Significant wave height (SWH) is important in marine engineering construction, marine transportation, environmental protection and military operations (Mahjoobi et al. 2009). Thus, accurate forecasts of significant wave height can reduce the occurrence of disasters and reduce economic losses. Many studies have proposed different forecasting methods (empirical models, numerical simulations, and machine learning methods) to improve the accuracy of ocean wave forecasting. However, prior assumption-based models (i.e., autoregressive moving average methods) has limited predictive power; on the other hand, methods such as: numerical models (WAVE Model (WAM) (Group et al. 1988) , Simulated Waves Nearshore (SWAN) (Booij et al. 1999) , the third-generation wave model (WAVEWATCH III), and wave values Model (Marine Science and Numerical Modeling, WASNUM) (Tolman et al. 1999), which solve the wave spectrum governing equations to simulate and predict wave elements, have significant predictive capabilities for vast sea areas, but they require high-performance computing and large sample data sets. Fortunately, machine learning is not limited by high-performance computing and large-scale data sets, and thus it is widely used for wave prediction. They generally extract mathematical expressions or find empirical relationships between input and target variables from analysis of available time series (Solomatine and Ostfeld 2008) and have been widely used for real-life applications of different fields such as discharge and river flow prediction, evaporation estimation, flood management and wind speed prediction (Shamshirband et al. 2020). Currently Long short-term memory (LSTM), Gated recurrent unit (GRU) and other models predict the significant wave height, only using the time series information of the data, and failing to use the spatial information. In this article, we construct Convolutional Long the Short-Term Memory (ConvLSTM) based method to establish a short-term (24h) prediction for the significant wave height in coastal sites in or near China. Several parameters can affect significant wave height in nearshore and offshore areas such as water depth, swell, refraction, diffraction, terrain and winds. Therefore, we design a variety of significant wave height prediction sensitivity experiments (combining different attribute information). Experimental results suggests that our model has achieved good results.

2. Method

Our purpose is to study the effect of wave attributes and spatial information on SWH. We carried out two kinds of sensitivity experiments for significant wave height prediction. One is to explore the sensitivity of different wave properties to SWH. Another is to explore the accuracy of spatial information for predicting SWH of ocean waves. LSTM is capable of processing time series data, but it can't handle high-dimensional data. Therefore, we add a convolution operation with LSTM as backbone, and consider spatial information while processing timing information (denoted as ConvLSTM). It replaces matrix multiplication with the convolution operation of each gate in the LSTM unit and captures basic spatial features by performing convolution operations in multi-dimensional data. Since the input of classical LSTM is one-dimensional, it is not suitable for spatial sequence data, such as video, satellite, and radar image data sets, we design our ConvLSTM using 3D data as its input. ConvLSTM consists of two networks: encoding network and prediction network. Both networks are superimposed by several ConvLSTM layers. We take the output of ConvLSTM layers (coding network) as the input of ConvLSTM layers (corresponding prediction network) (Xingjian et al. 2015). Finally, all the states of each layer in the prediction network are connected to generate the final prediction. Because of its multi-layer superposition, it has a strong representation ability and is suitable for precipitation, significant wave height prediction and other complex dynamic systems. Compared with the newer Informer (Zhou et al. 2021) method only considering time series, our ConvLSTM simultaneously considers time and spatial information. We use Root Mean Square Error (RMSE) to verify the importance of spatial information. Our experiment found that the original ConvLSTM produces severe overfitting on the ocean wave dataset. Therefore, we made some improvements. First, we add a code to reduce the overfitting. Moreover, we found that Batch Normalization in ConvLSTM

(BatchNorm) batches the entire batch, which is related to the batch size setting, does not work well. We get the best results in ConvLSTM with InstanceNorm batch instead of BatchNorm batch, which can improve the prediction accuracy.

3. Experiments

Data Processing: SWH data comes from the French ocean wave WAVERYS with a spatial resolution of 0.2° and a time interval of 3h. In this article, the local area is used (108.2°E-150.2°E, 6.4°N-40°N). Moreover, two other elements are also needed: the wind speed at 10m (from EAR5 reanalysis field with a spatial resolution of 0.25 degrees and a time resolution of 3h), and the terrain (from National Centers for Environmental Information with a spatial resolution of 0.0166 degrees). Note that the time resolution of all data is 3h, and the time interval is 2018-2019. Since the resolution of wave height and wind speed data is inconsistent, the nearest neighbor interpolation method is used to unify the resolution to 0.2°. We divide the dataset into a training set of 4,656 samples and a test set of 920 samples. To study the reliability of the forecasting, we choose four months (corresponding to spring, summer, autumn and winter) in the period 2018-2019 for forecasting.

Data standardization: To make the features of different measures comparable, we employ Z-score standardization to standardize input features and perform scaling transformations on the feature vectors of different attributes. After processing, each attribute (i.e., dimension of the feature vector) confirms to the standard normal distribution. Formally,

$$X\_h=(X-M)/DS$$

where X is the original data, M is the mean of the original data, DS is the standard deviation of the original data, and X_h is the standardized data.

Evaluation metrics: Two types of criteria are used to measure the performance of our method: 1) Root mean square error (RMSE) and 2) correlation coefficient (R). RMSE represents the deviation between the forecast and the observed significant wave height, and R reflects the correlation between the observed value and the forecast wave height. Formally,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2},$$

$$R = \frac{\frac{1}{N}\sum_{i=1}^{N}(P_i - P_a)(O_i - O_a)}{\left[\sum_{i=1}^{N}(P_i - P_a)^2 \sum_{i=1}^{N}(O_i - O_a)^2\right]^{\frac{1}{2}}}$$

where P_i is the prediction value, O_i is the observed value, N is the number of samples, P_a is the average of the prediction values, and O_a is the average of the observed values. The smaller the RMSE, the closer the R value is to 1, which means the higher the accuracy of the model, and the closer to the observed value.

4. Results and analysis

We compared our method (ConvLSTM) with the state-of-the-art method Informer model with different attribute combinations. Tables 1-4 show the qualitative results with different elements.

Table 1. RMSE and Correlation coefficient (R) of SWH with Previous SWH as Input

Forecast Time(/h)	ConvLSTM		Informer	
	RMSE	R	RMSE	R
3h	0.342	0.933	0.420	0.909
6h	0.343	0.931	0.428	0.906
9h	0.348	0.927	0.434	0.899
12h	0.353	0.924	0.459	0.888
15h	0.363	0.921	0.483	0.875
18h	0.371	0.916	0.505	0.862
21h	0.372	0.911	0.526	0.849
24h	0.381	0.904	0.547	0.835

Table 2. RMSE and Correlation coefficient (R) of SWH with Previous SWH and Previous wind as Input

Forecast Time(/h)	ConvLSTM		Informer	
	RMSE	R	RMSE	R
3h	0.321	0.946	0.206	0.978
6h	0.334	0.945	0.241	0.970
9h	0.338	0.943	0.276	0.960
12h	0.339	0.941	0.314	0.947
15h	0.344	0.938	0.355	0.932
18h	0.346	0.935	0.389	0.918
21h	0.347	0.931	0.419	0.904
24h	0.349	0.926	0.447	0.889

Table 3. RMSE and Correlation coefficient (R) of SWH with Previous SWH, wind and terrain as Input

Forecast Time(/h)	ConvLSTM		Informer	
	RMSE	R	RMSE	R
3h	0.330	0.944	0.215	0.976
6h	0.334	0.942	0.249	0.968
9h	0.340	0.940	0.2835	0.958
12h	0.340	0.937	0.321	0.945
15h	0.345	0.934	0.3609	0.930
18h	0.351	0.930	0.394	0.916
21h	0.352	0.925	0.423	0.902
24h	0.353	0.919	0.452	0.888

Table 4. RMSE and Correlation coefficient (R) of SWH with Previous SWH and current wind as Input

Forecast Time(/h)	ConvLSTM		Informer	
	RMSE	R	RMSE	R
3h	0.179	0.988	0.204	0.978
6h	0.182	0.988	0.240	0.970
9h	0.186	0.987	0.276	0.959
12h	0.190	0.987	0.3155	0.947
15h	0.195	0.986	0.356	0.932
18h	0.203	0.985	0.390	0.917
21h	0.218	0.983	0.420	0.903
24h	0.239	0.98	0.448	0.889

From Tables 1 and 2, we can see that the wind speed

information has the greatest influence in predicting the significant wave height. Table 3 suggests that topographic factors have no significant effect on the prediction of significant wave height. From the above table, we can see that Table IV (using current wind speed) achieves the best prediction results (RMSE decreases to 0.24 for 24h and the correlation coefficient reaches 1) when the same model of ConvLSTM is used, indicating that the wind speed at the current moment is the most important factor affecting the significant wave height. However, Informer is not sensitive to the current wind speed, and we can see from Table IV that the prediction results of ConvLSTM are much higher than those of Informer when the same current wind speed is used at the same time. In addition, it can be seen that Informer has a large difference in predicting the significant wave height in long series, while our ConvLSTM has very little. Informer is inferior to ConvLSTM in long-term prediction. However, the short-term prediction of Informer is better than that of ConvLSTM. We can conclude that the combination of spatial information and current wind speed is very important for predicting the significant wave height, which greatly improves the accuracy of the prediction. Compared with Informer, ConvLSTM using spatial information performs superiorly in predicting the significant wave height.

Figure 1 shows our best results (24h) predicted using ConvLSTM with a time interval of 3h. The first and third columns are the observed values for 3-24h, and the second and fourth columns are the predicted values for 3-24h.

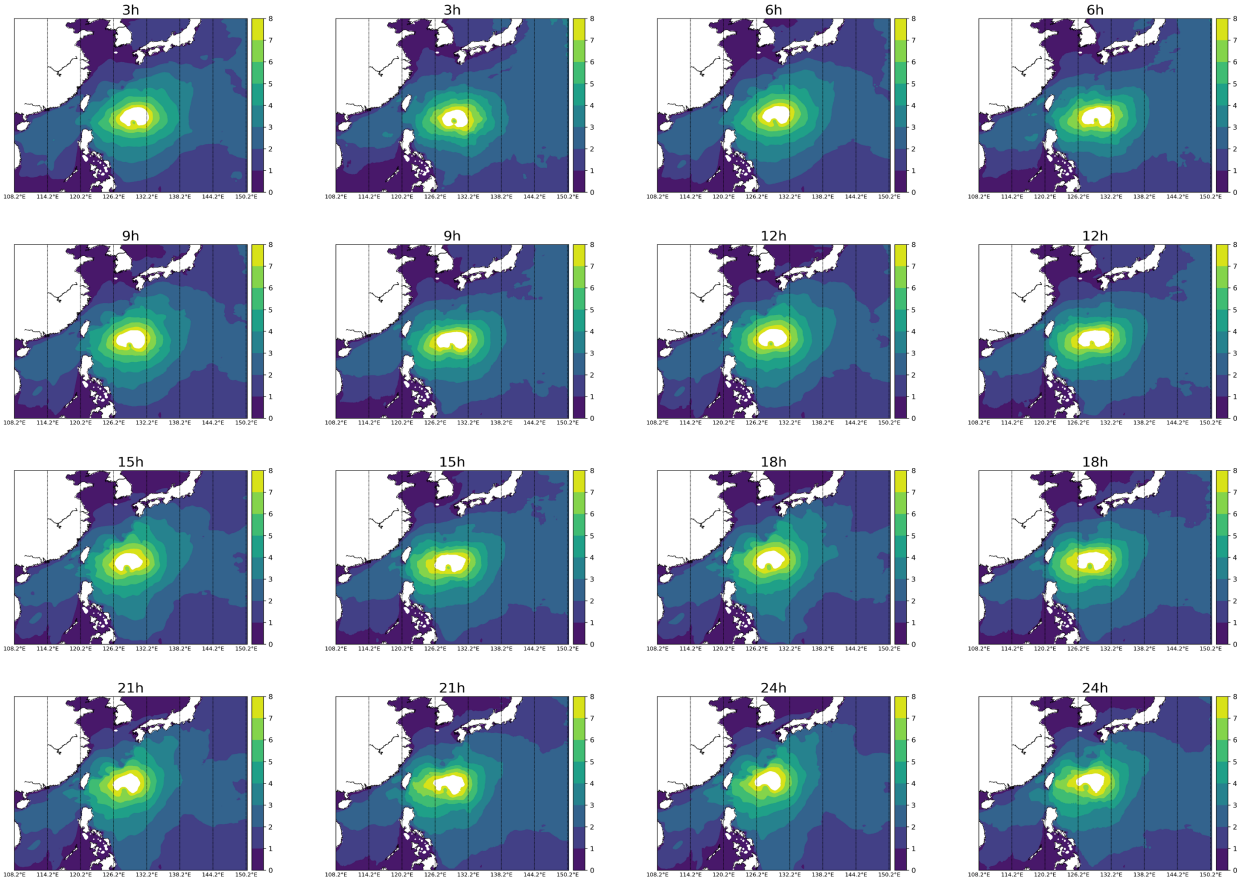

Figure 1: left is observed value, right is prediction

5. Conclusion

This article establishes a short-term prediction model of significant wave height based on ConvLSTM. Experimental results show that wind speed is the most important element that affects the significant wave height prediction. Using wind speed information and wave height information, we can accurately predict the 3h and 6h significant wave height. As the forecasting timeliness increases, the forecast accuracy of 21h and 24h is rapidly reduced. Meanwhile, we can conclude that if model input is not significantly correlated with the prediction result, they cannot effectively improve the predictive ability of the model. However, our ConvLSTM model has higher learning ability and adaptability. It provides an effective machine learning method for short-term prediction of significant wave height in complex terrain and can be widely used in several Marine environment applications.

Acknowledgments

This work is supported by National Key Research and Development Program of China (2020YFA0608001) and National Science Foundation of China (42075142, 41505079, 42130608 and 42030611).

References

Mahjoobi, J., & E.A. Mosabbeb, (2009). Prediction of significant wave height using regressive support vector machines. *Ocean Engineering*, **36**(5), 339-347.

Group, T.W. (1988). The WAM model—A third generation ocean wave prediction model. *Journal of Physical Oceanography*, **18**(12), 1775-1810.

Booij, N.R.R.C., R.C. Ris, & L.H. Holthuijsen, (1999). A third-generation wave model for coastal regions: 1. Model description and validation. *Journal of geophysical research: Oceans*, **104**(C4), 7649-7666.

Tolman, H.L. (1999). User manual and system documentation of WAVEWATCH-III version 1.18. NOAA/NCEP Tech. Note 166, 110 pp.

Shamshirband, S., A. Mosavi, T. Rabczuk, N. Nabipour, & K.W. Chau, (2020). Prediction of significant wave height; comparison between nested grid numerical model, and machine learning models of artificial neural networks, extreme learning and support vector machines. *Engineering Applications of Computational Fluid Mechanics*, **14**(1), 805-817.

Shi, X., Z. Chen, H. Wang, D.Y. Yeung, W.K. Wong, & W.C. Woo, (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems* pp. 802-810.

Zhou, H., S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, & W. Zhang, (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of AAAI*.

Forecasting Significant Wave Height based on Transformer

Xiaojie Li[1], Hao Zhang[1], Jiuke Wang[2], Xiaomeng Huang[3*], Qi Zhong[4], Xi Wu[1*]

1. Department of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China
2. National Marine Environmental Forecasting Center, Beijing, China
3. Ministry of Education Key Laboratory for Earth System Modeling and Department of Earth System Science, Tsinghua University, Beijing, China
4. China Meteorological Administration Training Center, Beijing, China
*Corresponding authors

1. Introduction

Effectively forecasting wave height is essential for marine activities such as marine transportation, offshore oil development, marine scientific research, and fishery production. Since wave height is affected by several uncertain factors the issue of how to accurately and quickly forecast wave height remains a key and complex problem.

Before the 20th century, researchers used physical numerical wave models (Group et al. 1988) to forecast wave height. It was preliminarily applied along coastal cities to help people avoid disasters. However, the internal calculation of these numerical models is enormous and complex, which leads to a large number of calculation resources that are unable to produce timely prediction results. Therefore, many researchers began to use machine learning methods to forecast significant wave height (SWH). James et al. (2018) used support vector machine (SVM), multi-layer perceptron (MLP), and other machine learning methods to quickly and effectively forecast SWH. The successful application of machine learning in wave height prediction also demonstrates the prediction capability of machine learning methods in marine science (MacKay et al. 2003). In addition, machine learning method can also predict SWH on a wider area (Deo et al. 2003). Compared with traditional numerical models, the prediction results have better generalization ability and robustness, and the computation takes less time. With the development and application of machine learning in marine science, Recurrent Neural Network (RNN) (Zaremba et al. 2014), Gated Recurrent Unit (GRU) (Cho et al. 2014), Long-Short Term Memory (LSTM) (Olah et al. 2015), and other deep learning methods have been used to improve the accuracy of wave height prediction (VS et al. 2020). However, problems in the processing of long-term data series remain, such as long-term dependence, gradient vanishing and exploding, and easy over-fitting of the model. These methods have limitations in learning complex dynamic relations from time-series data, thus better methods need to be developed to replace them. Convolutional LSTM (Xingjian et al. 2015) structure is also considered by us to solve the wave height prediction task. It can capture more characteristic information in space and adapt to a larger area. However, its prediction accuracy is relatively low.

In this article, we propose a Transformer-based machine learning model to predict wave height. The model can accurately forecast wave height (3-24h) by using the self-attention mechanism to learn complex dyna mic relations from time-series data. Experimental results show that our prediction result is better than state-of-the-art methods.

2. Method

Classical Transformer, proposed by Ashish Vaswani et al. (2017), aims to solve machine translation tasks. It differs from other forecast models (e.g. RNN or CNN) in that it relies on the attention mechanism. In this article, we use Transformer as the backbone for forecasting wave height. This method is still a Seq2Seq structure, including an encoder and decoder, as explained below:

Encoder: It consists of an input layer, a position-coding layer, and eight identical coding layers. We first map the time series data (previous day's mode data) into multi-dimensional feature vectors through a fully connected neural network and then encode the position information of the time series data into the feature vector by sine and cosine functions in the position-coding layer. Finally, we learn the complex dynamic relationship in the multi-dimensional feature vector through eight identical coding layers and send it into the decoder.

Decoder: It consists of eight identical decoding layers and output layers. The former aims to fuse the wave time series data and the multi-dimensional feature vector (generated by the encoder) through the self-attention layer and the feedforward neural network. Finally, output forecasting wave height by the output layer.

3. Experiment design

Datasets: French wave reanalysis data from 2018 to 2019 is used in our experiment. It has a spatial

resolution of 0.25° and a time interval of 3h. Selecting January 2018, October 2018, April 2019, and July 2019 (corresponding to spring, summer, autumn and winter) as test data and remaining data as training data. Besides the global area, we extract part of China's coastal data (accuracy range: 108.2°-150.4°, latitude range: 6.4°-40.0°) as partial area prediction experiment data.

Data feature selection: We use the Mann-Kendall method to select the most helpful 14 features which are relevant to the prediction results and convenient to accurately predict the wave height in the next 3-24 hours. This aims to reduce the amount of calculation of the model and obtain the prediction results more accurately and quickly.

Data standardization: We scale data with different dimensions to the same latitude, making them comparable. We use zero-mean normalization to map the data to 0-1 based on the mean and standard deviation of the original data (with mean value 0 and standard deviation 1). This means that the processed data satisfies the normal distribution. Formally:

$$Xt = \frac{Xs - M}{STD}$$

where Xt is the processed data, Xs is the original data, M and STD are the mean and the standard deviation of the original data respectively.

Training & Testing details: Experiments are implemented on the Pytorch framework (Paszke et al. 2019) with Tesla T4GPU. AdamW optimizer (Loshchiloy et al. 2017) is used with an initial learning rate of 0.1, gradually decreasing to 1e-4, and with the momentum of (0.9, 0.99) until it converges. The mean square error (MSE) is chosen as the cost function of training, and its calculation result can guide the model learning. Formally:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}[pre(i) - obs(i)]^2$$

where *pre* and *obs* denote forecasting wave height and real wave height respectively.

We test the performance of all models on local and global test datasets. We compared our Transformer method with the state-of-the-art method RNN, LSTM, GRU models with different attribute combinations.

4. Results and analysis

Root-mean-square error (RMSE) and Pearson correlation coefficient (CC) are used to measure the performance of different methods. RMSE represents the deviation between the real and the forecast wave height, and CC reflects the correlation between the real value and the forecast wave height. Formally:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[pre(i) - obs(i)]^2},$$

$$CC = \frac{\sum_{i=1}^{N}[pre(i) - \overline{pre}][obs(i) - \overline{obs}]}{\sqrt{\sum_{i=1}^{N}[pre(i) - \overline{pre}]^2 \sum_{i=1}^{N}[obs(i) - \overline{obs}]^2}}$$

where *pre* and *obs* denote forecasting and real wave height respectively. $\overline{pre}$ and $\overline{obs}$ represent the mean of the forecast and real wave height. The lower the RMSE value and the higher the correlation coefficient show the higher accuracy of the prediction result.

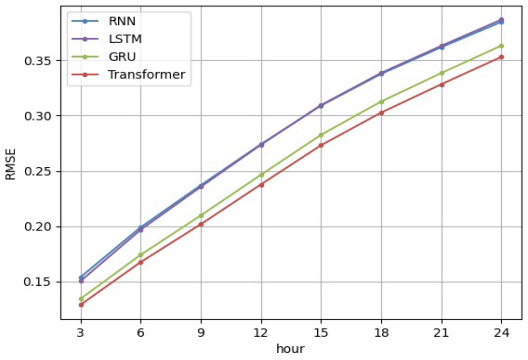

Figure 1: RMSE of SWH in local area forecast (24h)

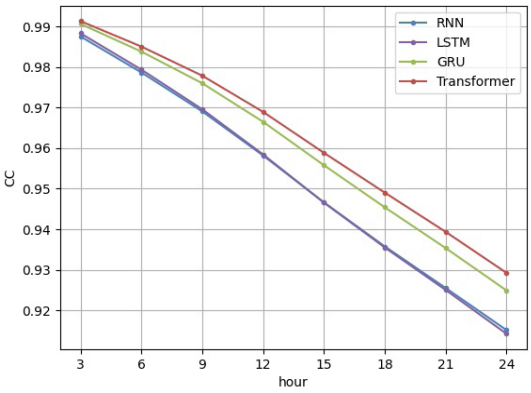

Figure 2: Correlation coefficient (CC) of SWH in local area forecast (24h)

Figures 1 and 2 show the RMSE and CC of SWH in the local area forecast and suggests that our method has reached the optimal performance at each moment, and its performance degradation over time is also significantly lower than other methods. Furthermore, RMSE increases with the extension of the forecast time, and CC gradually decreases. That's because the time series forecasting model when predicting future results depends on the pre-predicted results.

Figure 3 shows the changes in wave height (next 3 hours) predicted by our Transformer in a partial area. The black line represents the L1 (Manhattan) distance between the actual and forecast values, which is not more than 0.25m. Moreover, the curve change trend of the predicted wave height is very similar to the actual wave height change trend, so it further verifies that the Transformer has made excellent achievements in the predicted wave height.

Table 1 illustrates the 24-hour forecast results of all models in each season of the partial area. We can see that our Transformer in each season has reached the best results with the lowest RMSE and the highest

correlation coefficient (close to 1). Note that the prediction results of each model fluctuate with the change of seasons. This is because of different climatic conditions in each season. In a follow-up work, the wind speed can be input as a feature to the model for training to reduce this problem. Table 2 shows the performance of all models in predicting the height of global waves. This further proves that our Transformer has a strong generalization ability in predicting the global wave height.

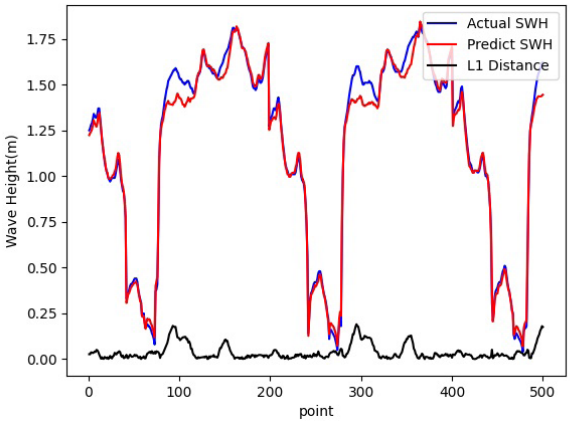

Figure 3: Time series plot for 3h prediction of SWH (partial area)

Table 1. RMSE and Correlation coefficient (CC) of SWH in partial area forecast (24h)

Methods	January 2018		October 2018		April 2019		July 2019	
	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC
RNN	0.456	0.887	0.449	0.909	0.313	0.894	0.290	0.907
LSTM	0.460	0.885	0.450	0.909	0.315	0.893	0.291	0.907
GRU	0.438	0.897	0.414	0.923	0.293	0.908	0.278	0.915
Transformer	**0.433**	**0.900**	**0.397**	**0.929**	**0.285**	**0.913**	**0.267**	**0.922**

Table 2. RMSE and Correlation coefficient (CC) of SWH in global area forecast (24h)

Methods	January 2018		October 2018		April 2019		July 2019	
	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC
RNN	0.565	0.894	0.554	0.909	0.573	0.905	0.554	0.930
LSTM	0.555	0.898	0.547	0.912	0.563	0.909	0.541	0.934
GRU	0.521	0.912	0.505	0.925	0.514	0.924	0.490	0.946
Transformer	**0.515**	**0.914**	**0.489**	**0.930**	**0.504**	**0.927**	**0.476**	**0.949**

5. Final remarks

Machine learning has gradually replaced the traditional numerical model to forecast wave height. In this article, we propose a transformer-based method to predict the wave height in the next 3-24 hours and reach state-of-the-art in many time series forecasting models. Compared with other methods, our method effectively solves the problem of long-distance dependence in time series forecasting because it quickly finds the characteristics, trends, and development laws of variable changes from time-series data through its self-attention mechanism. Experimental results demonstrate that our method is better than RNN, LSTM, GRU in terms of RMSE and CC. In a following work, we would obtain more accurate prediction results by adding the depth of the ocean, wind speed, and other factors in our experiment.

Acknowledgments

This work is supported by National Key Research and Development Program of China (2020YFA0608001) and National Science Foundation of China (42075142, 41505079, 42130608 and 42030611).

References

Group, T.W., 1988. The WAM model—A third generation ocean wave prediction model. *Journal of Physical Oceanography*, **18**(12), 1775-1810.

James, S.C., Y. Zhang, and F. O'Donncha, 2018. A machine learning framework to forecast wave conditions. *Coastal Engineering*, **137**, 1-10.

MacKay, D.J. and D.J. Mac Kay, 2003. Information theory, inference and learning algorithms.

Deo, M.C. and C.S. Naidu, 1998. Real time wave forecasting using neural networks. *Ocean engineering*, **26**(3), 191-203.

Zaremba, W., I. Sutskever, and O. Vinyals, 2014. Recurrent neural network regularization. arXiv preprint arXiv:1409.2329.

Cho, K., B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Olah, C., 2015. Understanding lstm networks. VS, F.E., 2020, May. Forecasting Significant Wave Height using RNN-LSTM Models. In *2020 4th International Conference on Intelligent Computing and Control Systems* (ICICCS) (1141-1146), http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

Shi, X., Z. Chen, H. Wang, D.Y. Yeung, W.K. Wong, and W.C. Woo, 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems* (802-810).

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, and I. Polosukhin, 2017. Attention is all you need. *Advances in neural information processing systems*, 5998-6008.

Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, and A. Desmaison, 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, **32**, 8026-8037.

Loshchilov, I., and F. Hutter, 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

Heavy Precipitation Nowcasting Based on Deep Learning

Jianbing Ma, Xianghao Cui, and Xi Wu

Chengdu University of Information Technology, Chengdu, China

1. Introduction

Nowcast typically means weather forecast of a very short term, usually less than two hours. Precipitation nowcasting is one of the most important nowcasting application areas since it is broadly applicable in agriculture, transportation, outdoor activities, and aviation control, to name a few. Intensive research efforts have been proposed in this area (Prudden et al. 2020; Patel et al. 2018; Shi et al. 2015; Shi et al. 2017; Soman et al. 2010). Whilst the traditional Numerical Weather Prediction (NWP) technique have been proved to be infeasible in precipitation nowcasting (Prudden et al. 2020; Patel et al. 2018;), linear extrapolation methods (Woo and Wong 2017; Bowler et al. 2004) have been introduced and showed better performance than NWP. However, these methods could not capture the underlying patterns of aberrations and trends from the historical data as the dynamic and non-linear nature of nowcasting. So nowadays, scientists have been beginning to deploy deep neural networks (Patel et al. 2018; Shi et al. 2015; Shi et al. 2017; Vaswani et al. 2017; Hochreiter et al. 1997) to deal with the spatio-temporal inputs of precipitation nowcasting, at the same time, deep neural networks are also much more efficient in processing the big amount of data involved.

In a typical precipitation nowcasting scenario, there are a series of radar data in chronological order. Not surprisingly, methods that handle temporal data, such as Long Short-Term Memory (LSTM) (Shi et al. 2015; Shi et al. 2017; Hochreiter et al. 1997), Gated Recurrent Unit (GRU) (Shi et al., 2017), and 3-Dimensional Convolutional Neural Network (3D-CNN) (Patel et al. 2018) were widely used in the applications. Recently, the Transformer based on the attention mechanism (Vaswani 2017) has shown to be more effective in dealing with time series data such as in natural language processing applications, therefore it could be a potential tool in nowcasting. In this article, we will introduce the transformer network into the nowcasting task.

Moreover, the current deep learning approaches applied to precipitation nowcasting typically focus on all-levels of precipitations, evaluated by the usual Mean Squared Error (MSE) or Root Mean Squared Error (RMSE) indicators, as can be observed for instance in Cao et al. (2019) and Wei et al. (2020). Clearly, in real applications, people are usually more concerned about heavy precipitations as they are responsible for the most disasters. Whilst concentrating on heavy precipitations seem a practical demand, one of the main reasons why it has been almost neglected could be that the actual precipitation data are usually unbalanced with much less heavy precipitations.

To handle this issue, in our article, we introduce two lines of approaches to solve the data imbalance and hence improve the prediction accuracy of heavy rainfalls. First, we introduce the Synthetic Minority Oversampling Technique (SMOTE) algorithm to oversample the strong rainfall. Simply speaking, it uses adjacent positive samples with random imputations to generate new positive strong rainfall samples to alleviate the imbalance of the data. Second, we propose a customized loss function to adjust the weight of heavy precipitation related loss.

To validate our proposed approaches, we experiment on a public precipitation dataset (Yao and Li 2017) and compare our result with those of some common methods.

The article is organized as follows: in section 2, we present background knowledge; section 3 depicts the details of the experiment dataset, including dimensions, data significance, etc.; in section 4 we describe the principles of each deep learning model and how they are combined, and we analyze and discuss the experimental results; finally, we provide conclusions in section 5.

2. Related Work

2.1 Machine Learning Methods

The current deep learning-based rainfall prediction models roughly follow two research directions. The first solution approach is a conventional one. It tries to extract features from the radar echo maps and then use the features for prediction, for instance, the top one approach in the CIKM AnalytiCup 2017 challenge (Yao and Li 2017). In contrast, the second line of research is to predict the future radar maps from the previous sequence of echo maps. In Shi et al. (2017), they treated precipitation nowcasting as a spatio-temporal sequence forecasting problem and hence combined the CNN and LSTM model to form a Convolutional LSTM (ConvLSTM) model, which contains convolutional structures in both the input-

to-state and state-to-state transitions, so as to handle the spatio-temporal information. The ConvLSTM model was shown to outperform the traditional LSTM in capturing the spatio-temporal correlations in the radar echo sequences and provide more accurate predictions than the Real-time Optical flow by Variational methods for Echoes of Radar (ROVER) algorithm (Woo and Wong 2017) currently used by the Hong Kong Observatory (HKO). The ConvLSTM model was then extended to the Trajectory Gated Recurrent Unit (TrajGRU) model, in which the recurrent connections are dynamically determined. These methods suffer from the weakness that the predicted radar maps become more and more blurred as time goes on, and so are hard to give sharp and accurate predictions of the whole radar maps, which is less suitable in predicting heavy precipitations. The same image sequence prediction idea was also implemented in U-Net (Ronneberger et al. 2015) and its variant RainNet (Ayzel et al. 2020) to predict the future radar maps. Unfortunately, this method also has the problem of smoothing emergence, and was reported not suitable to be used for predictions longer than 1 hour.

2.2 SMOTE for data processing

In almost all precipitation datasets, heavy precipitations are rare. Whilst heavy precipitations are actually what people are interested in, the dataset imbalance has to be resolved. SMOTE is one of the most effective ways to solve the data imbalance based on a random oversampling algorithm. It generates new minority samples by random interpolation of the original minority samples, as shown by Fig. 1.

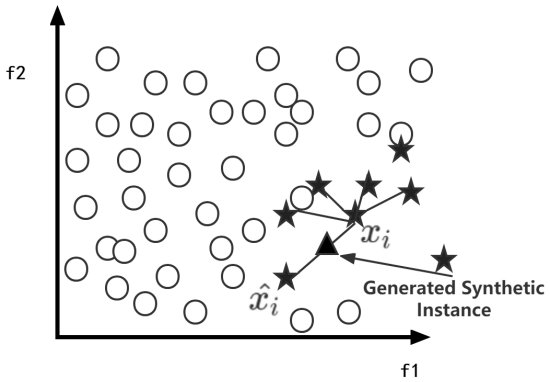

Figure 1: How SMOTE generates new samples. The 'asterisk' represents the minority sample, the 'circle' represents the majority sample, xi represents the current selected minority sample, and the 'square' represents the current randomly generated new sample.

More precisely, the algorithm can be described as follows:

A. For each sample x in the minority class, use the Euclidean distance to calculate the distance from all samples in the minority class sample set to obtain its k nearest neighbors.
B. Set a sampling ratio according to the sample imbalance ratio to determine the sampling magnification N. For each minority sample x, randomly select several samples from its k nearest neighbors, assuming that the selected nearest neighbor is x_n.
C. For each randomly selected neighbor xn, construct a new sample with the original sample according to the following formula.

$$x_{new} = x + rand(0,1) * (\tilde{x} - x)$$

3. The ShengZhen Precipitation Dataset

Provided by the ShenZhen Meteorological Bureau, this dataset contains 11000 samples, with 8000 in the training set and 3000 in the test set. Each sample contains 60 radar echo maps measured in 15 time series with an interval of 6 minutes, and in 4 different altitudes from 0.5km to 3.5km with an interval of 1km. Each sample covers the radar reflectivity of the surrounding 101 km × 101 km areas of the target site, and each training sample is labeled with a precipitation amount of the target site. The dataset contains three years of precipitation data, with the first two years for training and the last year for testing.

Since we do not have the label data for the test set. We partition the training set into new training and test sets, where we used 5000 for the training set and the remaining 3000 for the test set. The percentage of heavy precipitation samples (>30mm/h) in each dataset is 18.54% and 17.23%, respectively.

With a pre-experiment, in our experiments below, we deliberately used the fourth altitude (3.5km) of the radar echo data in training since the pre-experiment showed that using the fourth altitude gives a higher result accuracy compared to those of other altitudes. A pre-experiment also showed that using transformer gave us a better result than using LSTM, so in this article, we only report the results using the transformer structure.

4. Experiment Model Design

4.1 CNN + Transformer + Fully Connected (FC) Neural Network

In this model, we introduce the transformer network to replace the working part of LSTM. The transformer block has no recursion and convolution, and its ability to process time series is entirely due to its attention mechanism. It has been shown (Vaswani et al. 2017) that the transformer block has higher accuracy and efficiency than LSTM on many natural language tasks such as translation.

In our experiments, we made a few changes to the transformer block in Vaswani et al. (2017): we remove the decoding layer of transformer since in our scenario we only need the encoding layer to get the location feature vectors. The model structure is illustrated by Fig. 1.

The input to the model is a time-series radar maps with dimension (15,101,101). The convolution operation is then separately performed for each radar map so as to obtain the 15 temporal feature vectors. These 15 vectors are merged with the location features by position encoding to make them have temporal

sequential correlation, and then the information is fused in parallel for each location feature vector by the encoding layer, and finally the FC layer integrates all the results.

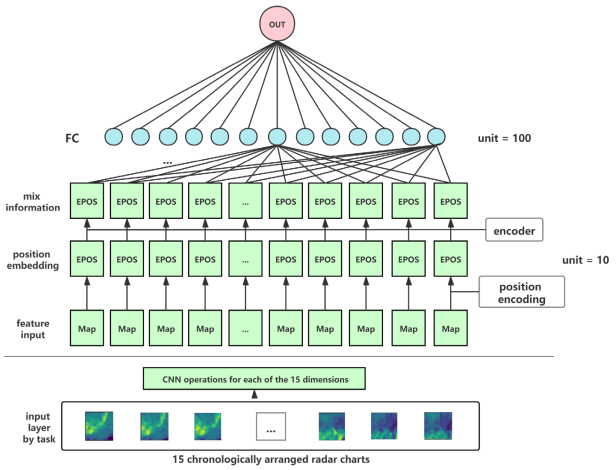

Figure 2: The model of CNN+Transformer+FC The input to the model is a time-series radar maps with dimension (15,101,101). The convolution operation is then separately performed for each radar map so as to obtain the 15 temporal feature vectors. These 15 vectors are merged with the location features by position encoding to make them have temporal sequential correlation, and then the information is fused in parallel for each location feature vector by the encoding layer, and finally the FC layer integrates all the results. The loss function is the MSE loss on precipitations. In the CNN block, six convolutional layers of the same structure are used to extract features, which are then compressed and output by a GlobalAveragePooling layer. **Note:** the following two models uses almost the same model structure but with augmented inputs and customized loss function, respectively.

4.2 SMOTE + CNN + Transformer + FC

Now there is a consideration on heavy precipitation, namely, whether using SMOTE oversampling, or heavy-precipitation oriented loss functions, can help increase the prediction accuracy on heavy precipitations.

In this subsection, we introduce the SMOTE algorithm to expand the strong rainfall samples. By using SMOTE, we generated 600 new heavy precipitation samples. We also generated their corresponding label values. The generated label values were calculated using the same random value we used to generate new sample images, namely,

$$y_{new} = y + random\_alpha * (\tilde{y} - y)$$

where y_{new} is precipitation label for the generated image, and $\tilde{y}$ are precipitation labels of the two original images, and the random_alpha value is the same value used in generating the corresponding image.

4.3 CNN + Transformer + FC + Customized loss function

There are many studies (Hu et al. 2018; Jaderberg et al. 2015; Bello et al. 2019; Oktay et al. 2018; Zhang et al. 2018) showing that the mechanism of amplifying the wanted signals and suppressing the unwanted signals can direct the network to focus on the features that are important for the task at hand. So here we also introduce this technique to focus on heavy precipitation. We use a customized loss function that increases the loss value of the strong precipitation samples, which makes the gradient of the strong rainfall samples increase, and thus increases the ability of the model to predict the strong precipitation samples.

The customized loss function is defined as follows:

loss = alpha *(True_precipitation - Pred_precipitation)^2

where alpha is 2 when True_precipitation >=30 and 1 otherwise.

4.4 Experiment Evaluation and Discussion

Now we evaluate the above two models.

Table 1. Impact of SMOTE and Customized Loss Function Concentrating on Heavy Precipitation

Dataset	RMSE	
	All-levels	Heavy precipitation
CNN+Transformer+FC	*17.46*	*28.9*
With SMOTE Extension	*18.9*	*10.3*
With Customized loss function	*18.3*	*18.6*

Table 1 describes the comparison of two extensions proposed to improve the prediction accuracy of heavy rainfall samples based on the use of the transformer model. It can be seen that by introducing SMOTE or customized loss function, the RMSE values on heavy precipitations decrease a lot. The SMOTE extended method reduces the RMSE from 28.9 to 10.3, which is a 64% improvement in prediction accuracy, and the use of the customized loss function also drops the RMSE to 18.6, which makes the best MSE result drop to 347, which is also a 36% improvement. So clearly, we can see that both extensions are very useful to enhance the prediction accuracy on heavy precipitations, especially the SMOTE extension. Nevertheless, due to the time and computational cost, and to avoid overfitting, we do not elaborately adjust the customized loss function parameters, otherwise the prediction RMSE could be even better.

Moreover, we can see that the prediction performance of the two extensions on all-levels of precipitations do not decrease too much, with about 6% increase on the RMSE values.

As we can see, the original model almost has no effect on predicting heavy precipitations, whilst the latter two do much better work. However, in the customized loss function case, we can still see that the predicted precipitation amount is almost always smaller than the true value, suggesting that it may have to adjust the weight on stressing heavy precipitations in the

customized loss function, whilst in the SMOTE extension, the predicted values catch the truths much tighter, but may bring the risk to fire false alarms. To this end, the customized loss function extension is a cautious approach in the sense that it seldom triggers false alarms whilst the SMOTE one is venturous but more accurate.

The following figures illustrate the results more visually.

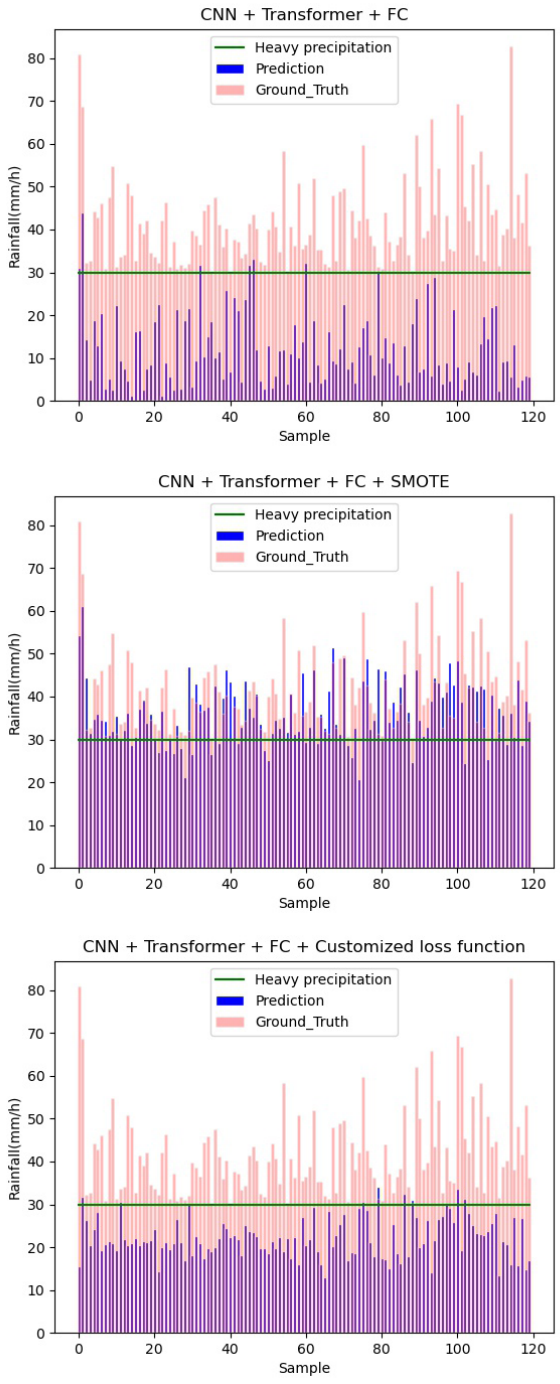

Figure 3: The prediction of heavy precipitations of the three models Note that in all the three subfigures, we only show the heavy precipitation samples, as can be seen from the orange-colored ones. The blue-colored results show the prediction performance on the heavy precipitation samples.

5. Conclusion

In this article, we introduced the transformer network into precipitation nowcasting and showed that the transformer-based network performed better than the LSTM-based network. Moreover, to improve the prediction accuracy on heavy precipitations, we introduced the SMOTE algorithm and customized loss functions to extend the transformer-based network. Both extensions were validated to be able to improve the prediction accuracy of the heavy precipitations by 64% and 36% respectively. Also, our result may suggest that the customized loss function extension is a cautious approach in the sense that it seldom triggers false alarms whilst the SMOTE one could be venturous.

As a future work, we will fine-tune the model to further improve the prediction accuracy. The Generative Adversarial Network (GAN), which has recently achieved great results, can also be used to augment the dataset to reduce data imbalance.

References

Prudden, R., S. Adams, D. Kangin, N. Robinson, S. Ravuri, S. Mohamed, and A. Arribas, 2020: A review of radar-based nowcasting of precipitation and applicable machine learning techniques. arXiv preprint arXiv:2005.04988.

Patel, M., A. Patel, and D. Ghosh, 2018: Precipitation nowcasting: Leveraging bidirectional lstm and 1d cnn. arXiv preprint arXiv:1810.10485.

Shi, X., Z. Chen, H. Wang, D. Y. Yeung, W.K. Wong, and W.C. Woo, 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pp. 802-810.

Shi, X., Z. Gao, L. Lausen, H. Wang, D.Y. Yeung, W. K. Wong, and W.C. Woo, 2017: Deep learning for precipitation nowcasting: A benchmark and a new model. arXiv preprint arXiv:1706.03458.

Marshall, J. S. and W. Palmer, 1948: The distribution of raindrops with size. *J. Meteor.*, **5**(4), 165-166.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and I. Polosukhin, 2017: Attention is all you need. In *Advances in neural information processing systems*, pp. 5998-6008.

Soman, S.S., H. Zareipour, O. Malik, and P. Mandal, 2010: A review of wind power and wind speed forecasting methods with different time horizons. In *North American Power Symposium*, pp. 1-8.

Hering, A.M., C. Morel, G. Galli, S. Sénési, P. Ambrosetti, and M. Boscacci, 2004: Nowcasting thunderstorms in the Alpine region using a radar based adaptive thresholding scheme. In *Proceedings of ERAD*, **Vol. 1**, No. 6.

Hu, J., L. Shen, and G. Sun, 2018: Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132-7141.

Jaderberg, M., K. Simonyan, and A. Zisserman, 2015: Spatial transformer networks. *Advances in neural information processing systems*, **28**, 2017-2025.

Yao, Y., and Z. Li, 2017: Short-Term Precipitation Forecasting Based on Radar Reflectivity Images. In *Proceedings of the CIKM*, Singapore.

Oktay, O., J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, and D. Rueckert, 2018: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

Zhang, X., X. Zhou, M. Lin, and J. Sun, 2018: Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6848-6856.

Woo, W.C., and W.K. Wong, 2017: Operational application of optical flow techniques to radar-based rainfall nowcasting. *Atmosphere*, **8**(3), 48.

LeCun, Y., Y. Bengio, and G. Hinton, 2015: Deep learning. *Nature*, **521**(7553), pp. 436-444.

Hochreiter, S. and J. Schmidhuber, 1997: Long short-term memory. *Neural computation*, **9**(8), 1735-1780.

Bowler, N.E., C.E. Pierce, and A. Seed, 2004: Development of a precipitation nowcasting algorithm based upon optical flow techniques. *Journal of Hydrology*, **288**(1-2), 74-91.

Cao, Y., Q. Li, H. Shan, Z. Huang, L. Chen, L. Ma, and J. Zhang, 2019: Precipitation nowcasting with star-bridge networks. arXiv preprint arXiv:1907.08069.

Wei, C.C. and C.C. Hsu, 2020: Extreme gradient boosting model for rain retrieval using radar reflectivity from various elevation angles. *Remote Sensing*, **12**(14), 2203.

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234-241.

Ayzel, G., T. Scheffer, and M. Heistermann, 2020: RainNet v1.0: a convolutional neural network for radar-based precipitation nowcasting. *Geoscientific Model Development*, https://doi.org/10.5194/gmd-2020-30.

Review on a data-driven framework for the convergence of ocean numerical simulation and machine learning

Zhenya Song[1, 2], Mengxuan Chen[3], Zhiyuan Kuang[1, 2], Quan Jin[1], Lingxiao Li[2, 4], Weiguo Liu[4], Haixing Liu[1, 2]

1. First Institute of Oceanography and Key Laboratory of Marine Science and Numerical Modeling, Ministry of Natural Resources, Qingdao 266061, China
2. Laboratory for Regional Oceanography and Numerical Modeling, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao 266237, China
3. Department of Earth System Science, Tsinghua University, Beijing 100084, China
4. School of Software, Shandong University, Jinan 250101, China

1. Introduction

Due to the collection of massive amounts of data, scientific research has progressed from experimental science, theoretical science, and computational science to a fourth paradigm called data-intensive scientific discovery. At the same time, breakthroughs in science and technology along with various observation and simulation methods have led to a continuous increase in the types and quality of data. Earth system data, characterized by its huge amount and rapid growth, i.e., they are massive (volume), fast acquisition (velocity), diversity in the type (variety), and quality (veracity), driving earth system research to enter into the "big data" era. The effective extraction of information from these data to increase our understanding of the Earth system, and ultimately to improve our forecasting ability, constitutes a new challenge.

The rapid development of high-performance computers and deep learning has provided both opportunities and challenges for innovating on ocean numerical models. Machine learning (including shallow and deep learning) algorithms can automatically extract patterns from data for use in predicting unknowns. These algorithms are essentially statistical learning methods. Therefore, there are six approaches to integrating machine learning and ocean numerical simulation. These approaches are new statistical forecasting/prediction, physical mechanism detection, parameterization optimization, model result correction, sub-model replacement, and numerical model surrogates. The implementation of these six approaches for data-driven deep integration of ocean numerical simulation and machine learning is discussed in this article, as are exploratory studies carried out within this integration framework.

2. Integration framework of ocean numerical simulation and machine learning

Reichstein et al. (2019) reviewed the application of machine learning to Earth science and proposed five approaches for integrating deep learning and earth system science. These approaches include improving parameterizations, replacing physical sub-models with machine learning models, analyzing model-observation mismatches, constraining sub-models, and surrogate modeling or emulation.

Concurrent with the work of Reichstein et al. (2019), six approaches driven by mass data for machine learning and ocean modeling convergence are proposed by us (Figure 1), including new statistical forecasting/prediction, physical mechanism detection, parameterization optimization, model result correction, sub-model replacement, and numerical model surrogates (Song et al. 2019). This convergence can be realized using a single approach or a combination of approaches. The six approaches are briefly described below.

(1) Statistical forecasting/prediction. Following the statistical method used in traditional forecasting, it is improved by applying machine learning instead of manual computations to perform statistical analyses of observational data. Due to the absence of oceanographic spatiotemporal data, however, this approach only applies to specific oceanographic phenomena (Gao et al. 2018; Ham et al. 2019) or forecasting in a small area (Deo and Naidu 1998; Mandal and Prabaharan 2006) now. It cannot provide continuous and large-scale coverage of the ocean as ocean numerical models do.

(2) Physical mechanism detection. The continuously increasing volume of oceanographic data has made it increasingly difficult to extract oceanographic phenomena and patterns. Machine learning is good at discovering complex relationships in high-dimensional data, such as, the relationship between typhoon track changes and background fields (Zhang et al. 2013; Park et al. 2016). This could help to identify and elucidate physical phenomena and laws of the ocean, improve our understanding of oceanographic processes, and promote further development and improvement of ocean models. However, machine learning mainly focuses on the correlation among two or more indicators and does not consider the

intrinsic causal relationship and physical mechanism among them. Oceanographers need to further analyze physical phenomena using theoretical analysis or ocean models.

(3) Parameterization optimization. Many parameterization schemes are available to approximate physical processes using ocean models, which introduce biases and uncertainties into simulation results. Parameter optimization is indispensable for developing and applying ocean models, but it currently relies mainly on manual debugging and heuristic experience. Machine learning can identify patterns in data obtained from many experiments and predict optimal parameters (Zhang et al. 2016; Jiang et al. 2018). It can also extract rules from large sets of observed data or high-resolution model results and establish a new parameterization scheme (Bolton and Zannal 2018), which improves the simulation and prediction ability of an ocean model. Parameter optimization is still applied to ocean model in recent studies rarely.

(4) Model result correction. Currently, simulated biases is still a serious problem in the ocean or atmosphere models. Like correction methods being explored for atmospheric models (Men et al. 2019), machine learning can identify patterns in the deviations of existing simulation and forecast results of ocean models to establish a deviation prediction model. Simulations and forecasts of ocean models can thus be corrected to reduce the deviation between model results and observations. However, model results can only be post-processed to improve ocean model simulation and forecasting capabilities. To develop and improve the models themselves, oceanographers still need to conduct in-depth analyses on the deviation prediction models to identify error sources and improve the modeling process. In addition, the model result correction also can be assumed as one of model assimilation methods and a trail based on a simple Lorentz model showed its bright future (Zhu et al. 2019).

(5) Sub-model replacement. The physical processes, such as nonlinear wave-wave interaction in the ocean wave model or the barotropic process in the ocean general circulation model, have been simplified as much as possible in models, but they still consume extensive computing resources and become computational bottlenecks. Reducing the number of calculations for these processes could negatively impact the accuracy of the simulation results. Machine learning, which can approximate any continuous nonlinear function, provide a solution to reduce the calculation cost. Studies on atmospheric and chemical models (Brenowitz and Bretherton 2018; Willis and von Stosch 2017) have shown that machine learning can be applied to an ocean model to fit functions to data for these processes. However, high accuracy requirements for the results of these processes in ocean models present a challenge for machine learning.

(6) Numerical model surrogate. This is similar to the new statistical forecasting/prediction method, but this approach uses the output of the numerical ocean model, which has larger coverages both in time and space compared to observations. An ocean model can be viewed as a mapping relationship between input and output variables, which can be formulated by machine learning with training datasets. The model based on machine learning is extremely important for emergency forecasting because it can calculate faster than traditional ocean models (i.e., ocean wave prediction, James et al. 2018). However, it should be noted that since the model is based on machine learning and the traditional ocean model output, its simulation and forecast accuracy cannot usually exceed those of ocean models.

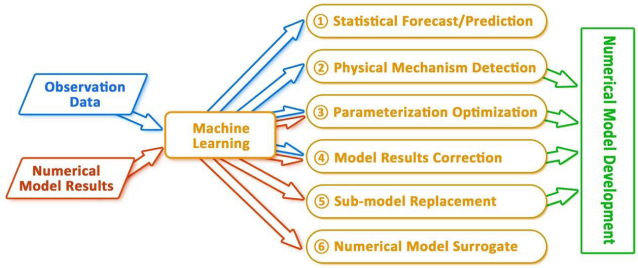

Figure 1: Schematic of an integration framework for ocean numerical simulation and machine learning

3. Explorative case study

Machine learning currently plays an important role in ocean numerical simulation and marine science. Integrating machine learning with ocean numerical simulation involves direct application and coordinated design to incorporate actual physical phenomena and their characteristics. The following two examples are used to illustrate.

3.1 Simulated Sea Surface Temperature (SST) bias correction

More than half a century of development has made the climate model an indispensable tool for elucidating the mechanism of climate change and forecasting the future climate. However, deficiencies in the ability of models to simulate climate change affect the accuracy of climate prediction.

Recently, ensemble empirical mode decomposition (EEMD) and a backpropagation (BP) neural network were used to correct a climate model of the global monthly average SST (Kuang et al. 2019). The First Institute of Oceanography Earth System Model version 2 (FIO-ESM v2.0) was run with Extended Reconstructed Global Sea Surface Temperature data (ERSST v5) from the US National Oceanic and Atmospheric Administration (NOAA) for historical and future scenarios of the Coupled Model Intercomparison Project 6 (CMIP6). The results were used to correct errors in the global monthly SST simulated using the historical scenarios, and the correction effect was evaluated. On this basis, a revised model was established and used to predict

and correct the global average SST for 2015–2100 under three Shared Socioeconomic Pathways (SSPs) called SSP1-2.6, SSP2-4.5, and SSP5-8.5.

Nine intrinsic mode function (IMF) components were obtained by performing EEMD on the observed and simulated global monthly average SSTs during the historical periods (1929–2014). Six new time signals were created using the seasonal, annual, interannual, decadal, interdecadal, and over-80-year scales, and each combined time signal was input into a BP neural network for deviation correction. The results showed that the revised model effectively reduced the deviation in the global monthly average SST simulated by the model; the root mean square error (RMSE) was reduced by 76.1% from 0.401°C to 0.096°C, the mean absolute deviation (MAD) was reduced by 77.2% from 0.338°C to 0.077°C, and the correlation coefficient for the observations was increased from 0.332 to 0.952.

The bias correction model based on the EEMD and BP neural network was then used to predict and correct the global average SST under the SSP1-2.6, SSP2-4.5, and SSP5-8.5 emission scenarios. Using the observations as a benchmark, the corrected results for 2015–2019 significantly surpassed the uncorrected results. Compared with the uncorrected predicted warming trend, the rate of increase in the global average SST for 2015–2100 was reduced to 0.424°C/100a, 1.325°C/100a, and 3.185°C/100a for SSP1-2.6, SSP2-4.5, and SSP5-8.5, respectively. The average increase in the temperature over the final 20 years of this century (2081–2100) compared with that for the last 20 years (1995–2014) was predicted to be 0.608°C, 1.183°C, and 2.409°C for SSP1-2.6, SSP2-4.5, and SSP5-8.5, respectively.

3.2 Fast simulation of the significant wave height

Ocean surface waves are the most common physical phenomenon in the ocean and correspond to ocean motion with the highest energy. They can propagate several kilometers with heights ranging from a few centimeters to tens of meters, and they play a vital role in navigation safety, coastal activities, and climate systems. Accurate simulation and prediction of ocean surface waves is necessary to better respond to changes in the climate system and ensure the safety of navigation and coastal activities.

Numerical simulation is a core technique used for ocean surface waves prediction. It can be categorized into two methods: empirical statistical methods and numerical wave models based on physical processes.

Traditional empirical statistical methods involve using mathematical statistics and fitting wind field and wave elements based on many simplifications and assumptions to produce empirical formulas for the simulation and forecasting of wave elements, which offer the advantage of a fast calculation speed. Ocean surface waves are composed of local wind waves and swells propagating from other sea areas. Empirical formulas obtained using long-term observations of ocean surface waves in early stages, can describe the growth of wind waves but cannot accurately show swells. Thus, there is a relatively large deviation between the simulation and prediction of sea waves and observations.

Numerical models of ocean surface waves are formulated based on the fundamental physical laws and processes governing such waves. The third-generation wave numerical model is based on a wave spectral energy balance equation, and it iteratively calculates each wave element using the wind, which is the main driving factor of the wave, and the terrain with fixed boundary conditions. This numerical model can be used to simulate and predict sea waves with relatively high accuracy; however, the calculation process for this five-dimensional problem has a high computational cost and is time-consuming. This does not facilitate high-resolution simulation and prediction. The simulation and prediction of wave elements via an ocean numerical model are based on wave conditions at one or more previous moments, so error accumulation may lead to large inaccuracies over the long term.

The rise of machine learning and other methods has resulted in the emergence of new empirical statistical methods for simulating and forecasting ocean surface waves elements with improved results at a single point or small area. However, the complexity associated with the coexistence of wind waves and swells continues to make simulation/predictions for global and oceanic basins a challenge.

A new neural network model, named OSWave-CNN (Ocean Surface Wave-Convolutional Neural Network), has been established for predicting the globally significant wave height. The global wind field can predict the globally significant wave height in a matter of seconds. Different structures are used to generate wind waves and swells in the model. Forecasting is not performed iteratively, thereby preventing a cold start and error accumulation in long-term forecasting.

Traditional CNNs extract key information during the convolution and pooling processes to better recognize the receptive field of all regions. As a result, considerable amount of secondary information is lost. To retain the information about the surrounding areas during the convolution process considering their continuous characteristics, OSWave-CNN adapts a CNN network structure and remove the pooling layer. It is constructed with alternating 1*1 and 3*3 convolution kernels (Figure 2), and the model output is the significant wave height field of global sea waves. To effectively predict the characteristics of wind waves and swells simultaneously, a 3*3 convolution kernel (equivalent to using eight points surrounding the central point in the calculation) is selected for the OSWave-CNN design, and the global wind field data

at N (N ≥ 2) moments preceding the current moment were used as the input.

The 2010–2018 global wind field and significant wave height data provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5 (EAR5) were used to train OSWave-CNN, with a time interval of 1 hour and a spatial resolution of 0.5 degrees. The extraction of the input wind field elements in OSWave-CNN is limited by the number of layers, and the receptive field of the model can be expanded by 0.5 degrees through one convolution using a 3*3 convolution kernel. With a maximum wind speed of 70 m/s, ocean surface waves can propagate approximately 735 km (about 7°) in 3 h. Thus, more than 14 grids are required for feature extraction, and the OSWave-CNN depth was set to 15 layers.

Based on the trained OSWave-CNN deep neural network model, the 2019 wind field data were selected as test input data, and the results were compared with the 2019 significant wave heights. The results show that the significant wave height model effectively simulated the seasonal mean significant wave height (Figure 3), typhoon waves (Figure 4), and the distribution characteristics.

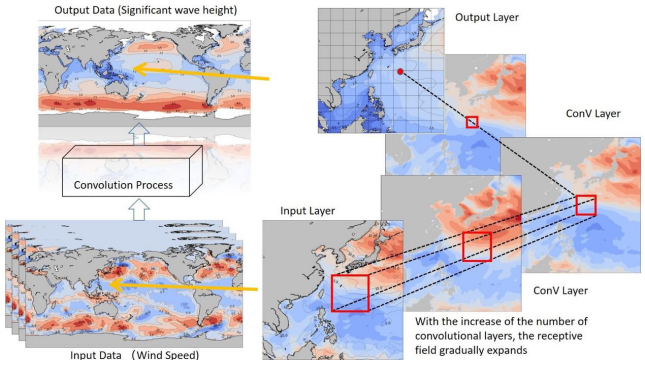

Figure 2: Schematic of the OSWave-CNN neural network model

4. Final Remarks

The emergence of massive data has moved scientific research from experimental induction, model deduction, and simulation to a fourth paradigm of data-driven scientific development. This has created new development opportunities and considerable challenges in marine science and computer science. Machine learning plays an important role in ocean numerical simulation and marine science; however, it cannot replace physical models and only serves to supplement and enhance them. Thus, the trend for integrating machine learning and marine science is to combine machine learning with physical models. As a result, coordinating the design of machine learning models that incorporate physical phenomena and features is a necessity. Feature extraction methods and deep learning neural networks are still in an exploratory stage and need to be further improved and developed to capture the temporally and spatially continuous characteristics of the ocean.

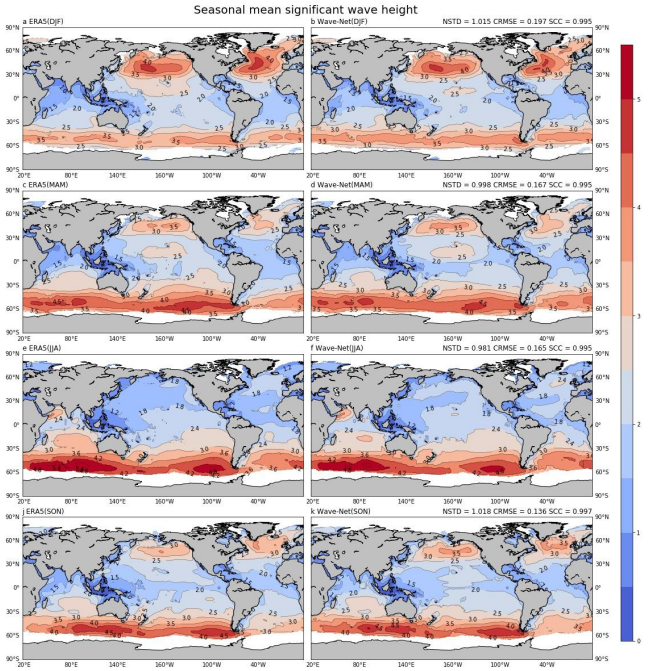

Figure 3: Climatological distributions of the seasonal mean significant wave heights (colored shaded) from 3-hourly ERA5 (left column) and OSWave-CNN (right column) data. The data were calculated for the boreal winter (December-January-February) on a seasonal basis.

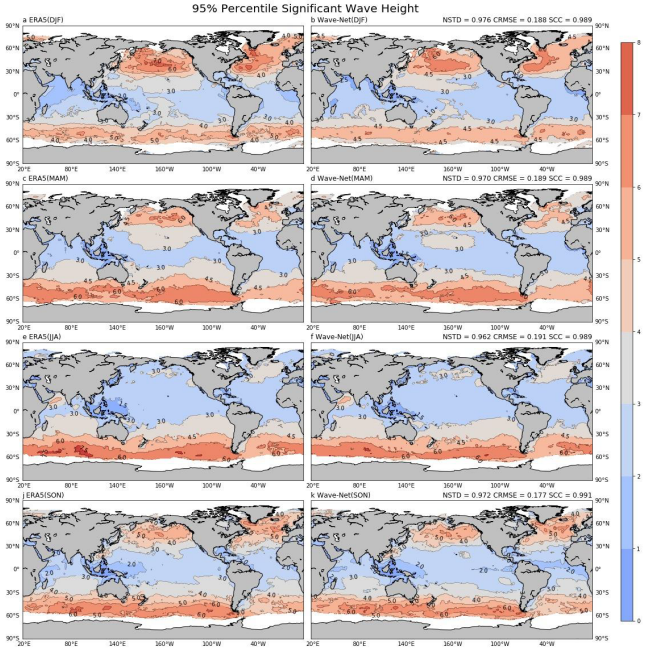

Figure 4: Climatological distributions of the 95th percentile significant wave heights (colored shaded) from 3-hourly data of ERA5 (left column) and OSWave-CNN (right column). The data were calculated for the boreal winter (December-January-February) on a seasonal basis.

Acknowledgments

This research was jointly supported by the National Key Research and Development Program of China (Grant No. 2016YFB0201100), the National Natural Science Foundation of China (Grant No. U1806205, 42022042, and 41821004), and the China-Korea Cooperation Project on Northwestern Pacific Climate Change and its Prediction.

References

Bolton, T., and L. Zanna, 2018: Applications of Deep learning to Ocean Data Inference and Subgrid Parameterization. *Journal of Advances in Modeling Earth Systems*, **11**(1), 376-399. https://doi.org/10.1029/2018MS001472.

Brenwitz, N.D., and C.S. Bretherton, 2018: Prognostic Validation of a Neural Network Unified Physics Parameterization. *Geophysical Research Letter*, **45**(12), 6289-6298. https://doi.org/10.1029/2018GL078510.

Deo, M.C., and C. Sridhar Naidu, 1998: Real time wave forecasting using neural networks. *Ocean Engineering*, **26**(3), 191-203. https://doi.org/10.1016/S0029-8018(97)10025-7.

Gao, S., P. Zhao, B. Pan, Y. Li, M. Zhou, J. Xu, S. Zhong, and Z. Shi, 2018: A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network. *Acta Oceanologica Sinica*, **37**(5), 8-12. https://doi.org/10.1007/s13131-018-1219-z.

Ham, Y.G., J.H. Kim, and J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572. https://doi.org/10.1038/s41586-019-1559-7.

James, S.C., Y. Zhang, and F. O'Donncha, 2018: A machine learning framework to forecast wave conditions. *Coastal Engineering*, **137**, 1-10. https://doi.org/10.1016/j.coastaleng.2018.03.004.

Jiang, G., J. Xu, and J. Wei, 2018: A Deep learning Algorithm of Neural Network for the Parameterization of Typhoon-Ocean Feedback in Typhoon Forecast Models. *Geophysical Research Letters*, **45**(8), 3706-3716. https://doi.org/10.1002/2018GL077004.

Kuang, Z., Z. Song, and C. Dong, 2020: Study on the Future Projection of Global Sea Surface Temperature over 21st Century Using a Biases Correction Model Based on Machine Learning. *Climate Change Research Letters*, **9**(4), 270-284. Doi: 10.12677/CCRL.2020.94031.

Mandal, S. and N. Prabaharan, 2006: Ocean wave forecasting using recurrent neural networks. *Ocean Engineering*, **33**(10), 1401-1410. https://doi.org/10.1016/j.oceaneng.2005.08.007.

Men, X., R. Jiao, D. Wang, C. Zhao, Y. Liu, J. Xia, H. Li, Z. Yan, J. Sun, and L. Wang, 2019: A temperature correction method for multi-model ensemble forecast in North China based on machine learning. *Climatic and Environmental Research*, **24**(1), 116-124. https://doi.org/10.3878/j.issn.1006-9585.2018.18049. (In Chinese)

Park, M.-S., M. Kim, M.-I. Lee, J. Im, and S. Park, 2016: Detection of tropical cyclone genesis via quantitative satellite ocean surface wind pattern and intensity analyses using decision trees. *Remote Sensing of Environment*, **183**, 205-214. https://doi.org/10.1016/j.rse.2016.06.006.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, 2019: Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**, 195-204. https://doi.org/10.1038/s41586-019-0912-1.

Willis, M.J., and M. von Stosch, 2017: Simultaneous parameter identification and discrimination of the nonparametric structure of hybrid semi-parametric models. *Computers & Chemical Engineering*, **104**, 366-376. https://doi.org/10.1016/j.compchemeng.2017.05.005.

Zhang, T., F. Xie, W. Xue, L. Li, H. Xu, and B. Wang, 2016: Quantification and optimization of parameter uncertainty in the grid-point atmospheric model GAMIL2. *Chinese Journal of Geophysics*, **59**(2), 465-475. https://doi.org/10.6038/cjg20160206. (In Chinese)

Zhang, W., Y. Leung, and J.C. Chan, 2013: The Analysis of Tropical Cyclone Tracks in the Western North Pacific through Data Mining. Part I; Tropical Cyclone Recurvature. *Journal of Applied Meteorology and Climatology*, **52**, 1394-1416. https://doi.org/10.1175/JAMC-D-12-045.1.

Improving machine learning-based weather forecast post-processing with clustering and transfer learning

Xiaomeng Huang, Lin Lin, Yuwen Chen, Yue Chen, Xing Huang, Mingqing Wang, Jonathon S. Wright

Department of Earth System Science, Ministry of Education Key Laboratory for Earth System Modeling, Institute for Global Change Studies,Tsinghua University, Beijing 100084, China

1. Introduction

The skill of numerical weather prediction (NWP) has improved significantly in recent decades due to advances in numerical models, data assimilation, and observation systems (Bauer et al. 2015). Nevertheless, the accuracy of NWP is still limited by imperfect model physics, numerical schemes, and initial/boundary conditions (Bauer et al. 2015; Lynch 2008). Following the pioneering work of Harry and Dale (1972), Model Output Statistics (MOS) have been used operationally for over forty years. Raw model forecasts are post-processed using statistical relationships between observations and NWP results. However, the volume and variety of observational and model output data are increasingly overwhelming conventional implementations of these methods (Agapiou and Athos 2017; Overpeck et al. 2011).

The emergence of machine learning (ML) techniques has provided new perspectives in this field (Reichstein et al. 2019). The climate community has increasingly turned to such techniques for applications such as improving subgrid-scale parameterizations in numerical models (Gentine et al. 2018; Jiang et al. 2018; Rasp et al. 2018; Schneider et al. 2017), improving forecasts at very short or very long lead times (Ham et al. 2019; Pan et al. 2019; Shi et al. 2015), detecting extreme weather (Hwang et al. 2019), and identifying complex teleconnection patterns (e.g., temperature maxima and extreme rainfall (Boers et al. 2019)). ML techniques could also substantially improve the accuracy of NWP results (McGovern et al. 2017; Rasp and Lerch 2018; Scher 2018). The success of ML relies heavily on the quality and quantity of training data. Unfortunately, observations are usually sparse, especially for newly-built weather stations. Essential questions therefore arise regarding whether and by what means models trained on data-rich stations can be reliably extended to newly-built stations with limited data records. Clustering techniques are widely used to extract information hidden in complex spatio-temporal data (Bador et al. 2015). Stations classified within the same cluster often share similar meteorological features. This type of feature-based classification provides a natural foundation for transfer learning, a technique by which knowledge gained in completing one task is repurposed for a different but related task (Sinno Jialin; Qiang 2010). These methods may permit models trained for data-rich stations to be rapidly fine-tuned for application to data-poor stations. To take full advantage of these techniques, we propose a new framework that combines three different ML methods: Clustering, Decision trees, and Transfer learning, or CDT for short. We apply CDT to surface air temperature forecasts as an illustrative validation of this framework and its applicability.

2. Data

NWP data are provided by The International Grand Global Ensemble (TIGGE) project of the European Centre for Medium-Range Weather Forecasts (ECMWF) (Bougeault et al. 2010; Swinbank et al. 2016). The numerical forecasts are initialized twice per day at 00 and 12~UTC with lead times ranging from 6 to 168 hours at 6-hour increments (for a total of 28 lead times). We use data for the period from 1 January 2013 to 31 December 2018. The sample size is therefore 4384 for each weather station and lead time. Five variables are selected: temperature and dew point temperature at 2m height, surface pressure, and the zonal and meridional wind components at 10m height. Observations from weather stations in China are obtained from www.meteomanz.com for the same period (1 January 2013 through 31 December 2018). As too few data are available in Xizang and Qinghai, we omit these areas from the analysis. We select 301 weather stations with data covering at least half of the year 2018 (the testing period as introduced below). Four variables (surface air temperature, surface pressure, surface air relative humidity, and near-surface wind speed) are provided every 3 hours (00, 03, 06, 09, 12, 15, 18, and 21 UTC). Static information for each station is also used, including latitude, longitude, and elevation. Missing values are filled via linear interpolation in the time dimension. The historical observations are processed to generate feature vectors with shapes defined by ($n_{samples}$, n_{steps} , $n_{features}$), where $n_{samples}$ is the number of records for a specified station, n_{steps} is the number of time steps used for temporal pattern mining, and $n_{features}$ is equal to 4 (i.e., the number of measurements to match at each time step). For example, the shape of the input vector for the Beijing station is (4384, 25, 4) when three days of past observations are used. NWP data are interpolated to each station location using an inversion-distance weighting IDW (Myers 1994) applied to forecast data from the four nearest model grid cells. The observational and NWP data are combined for input to the CDT framework.

3. Method

The CDT framework consists of three individual ML modules: clustering, decision-tree, and transfer learning. The clustering module classifies the 301 stations into groups using the traditional K-means technique. Data spanning the six-year period from 2013 to 2018 are divided into three parts: training, validation and testing data. Data from 2013 to 2017 are used for training (80% of the data) and validation (the remaining 20%). All data for 2018 are used for testing. We construct a separate model to post-process ECMWF forecasts at each lead time (28 in total) in each cluster. Separate decision-tree-based post-processing modules are then developed for each cluster and each lead time. Each newly-built station is assigned to the best-fit existing cluster. The transfer learning module is then used to produce the final results.

3.1 Clustering Stations with *K*-means

The traditional *K*-means (Hastie et al. 2009) clustering technique is often used for climate data analysis (Bador et al. 2015; Elsa et al. 2013). Stations with similar features are categorized into *K* individual clusters by calculating the feature distance between them. The features used in this study are the annual mean temperature and daily standard deviations of surface air temperature, surface air relative humidity, near-surface wind speed, surface pressure, latitude, longitude, and elevation. Models are established and trained for each cluster instead of for each station to reduce the computational cost and enlarge the training sample for each model. The clustering result is highly sensitive to the value of *K*. We use the Silhouette Coefficient (Rousseeuw 1987) to identify the optimal value of *K*. This metric measures the consistency of samples within each cluster as the ratio between cluster tightness and cluster dissociation.A larger Silhouette Coefficient indicates an increase in the inter-cluster distance relative to the intra-cluster distance. The maximum coefficient thus marks the optimal clustering result according to this metric. The average Silhouette Coefficient varies with the number of clusters *K*. We use the ASC to reduce the number of candidate K values. Although the ASC is useful for identifying potential optimal values of *K*, a larger ASC does not necessarily translate to a better ML model result. We test clusters based on $K = 2$, $K = 4$, and $K = 8$, which each produce climatologically coherent station groups.

3.2 Temperature post-processing based on LightGBM

After clustering, we apply a decision-tree model to characterize relationships between the NWP forecasts and observations, correct biases, and identify how different features affect the prediction results. Decision trees are tree-like graph models. Information is passed from the root (representing the raw data) and split into branches at each level. The splitting rule is typically set by the variable that best discriminates among the samples along each branch. Decision trees produce naturally explainable outputs and can provide valuable insight into hidden relationships uncovered by the algorithm. This method has been successfully employed in a wide variety of weather applications (McGovern et al. 2017). Gradient Boosting Decision Tree (GBDT) is a popular decision tree approach that involves an ensemble of sequentially-trained decision trees and gains knowledge by fitting negative gradients (Chen and Guestrin 2016).

In this work we use LightGBM (Ke et al. 2017), a highly efficient and scalable GBDT algorithm, to explore the relationships between NWP forecasts and observations in each cluster. LightGBM has been applied to sorting, classification, and regression tasks in a number of big-data studies (Gentine et al. 2018; Ju et al. 2019). Adopting a leaf-wise growth strategy with depth limitation and gradient-based one-side sampling, LightGBM seldom overfits on small training datasets.

Linear regression (LR), artificial neural network (ANN), and long short-term memory (LSTM) with a fully-connected network (FCN) are three widely used machine learning methods in meteorological applications (Cao and Gui 2018; Gensler et al. 2016; Zaytar and Amrani 2016). Therefore, LR, ANN, and LSTM-FCN are used as control models to predict temperature using identical inputs as LightGBM.

3.3 Transfer Learning for Newly-built Stations

In practice, ML models may malfunction due to data deficiencies or over-fitting. Transfer learning helps to reduce the likelihood of these types of failures by transferring knowledge from a previously trained model. The transferred model is then fine-tuned using newly-added data. This approach has been widely applied, including for the prediction of wind speed (Hu et al. 2016; Qureshia and Khan 2019). The LightGBM model for each cluster is taken as a pre-trained model, transferred and further trained on observations from newly-built stations identified as belonging to that cluster. The cluster to which each new station belongs is determined by static geolocation information along with the estimated annual means and standard deviations of key meteorological features (surface air temperature, pressure, wind speed, and relative humidity). The latter are IDW-interpolated from gridded NWP forecasts to accommodate the limited observational records at these stations. To replicate the operational scenario, we randomly select 20% of the stations to serve as synthetic newly-built stations, using the remaining 80% stations to produce pre-trained models for each of the four clusters. We then fine-tune the pre-trained models using data covering between zero and 24 months at 2-month increments. The use of zero months of data corresponds to applying the pre-trained model directly without fine-tuning. We then evaluate the corrected forecasts for the `new' stations using testing data from the year 2018. To validate the transfer learning results, we select seven lead times

ranging from 24h to 168h at 24h increments.

4. Results

4.1 Effect of clustering and advantages of LightGBM in comparision with LR, ANN and LSTM-FCN

The benefits of post-processing are most significant at short lead times, with RMSE reductions as large as 39.4% (1.02°C for 1-day forecasts (6-24 h lead times). Improvements decrease steadily to 20.0% (0.68°C) for 7-day forecasts (144-168 h lead times). The average RMSE (root mean square error) across all lead times is reduced by 0.81°C, corresponding to a 27.9% increase in accuracy. Clustering improves the effectiveness of the decision tree algorithm, with the greatest error reduction achieved when stations are grouped into four clusters. Compared to models without clustering (i.e., a single model trained on all stations), the RMSE is reduced by 0.54% when two clusters are used (*K*=2), 0.62% when *K*=4, and 0.41% when *K*=8. Since the*K*=4 result produces the smallest RMSE, we adopt this model for all subsequent experiments. Clustering reduces the RMSE at 296 out of 301 individual stations (98.3%) when *K*=4. In addition, clustering (*K*=*4*) can reduce the training time of the LightGBM by approximatly 40% in comparision with K=1.

The overall RMSE is reduced by 0.49°C (16.8%) under LR, 0.71°C (24.7%) under ANN, and 0.71°C (24.7%) under LSTM-FCN in the *K*=4 scenario, including RMSE reductions at 211 stations under LR, 270 stations under ANN, and 272 stations under LSTM-FCN.

LightGBM outperforms all three models, providing a further reduction of the RMSE for surface air temperature forecasts of 14.2% relative to LR, 3.8% relative to ANN, and 2.6% relative to LSTM-FCN, indicating that LightGBM is more effective for this application. LightGBM also takes less time for training (around 10 minutes) than ANN (around 20 minutes) or LSTM-FCN (around 40 minutes).

4.2 The effect of transfer learning relies on data span and lead time

Based on these findings, we conclude that LightGBM in combination with four clusters presents a substantial improvement over both the original operational forecasts and other ML-learning post-processing products. We therefore apply transfer learning to fine-tune the LightGBM model for extension to data-poor stations.

The pre-trained models outperform the original NWP by 0.56°C (16.8%) even without fine-tuning (Figure 1). The RMSE reduction continues to improve as the data span used for fine-tuning is extended, reaching 36.4% (1.23°C) when 12 months of data are used. Further improvements are negligible, indicating that the fine-tuning benefits plateau once the annual cycle is fully represented.

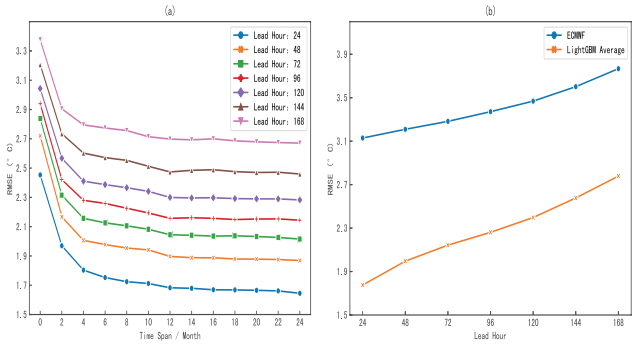

Figure 1: (a) Advantage of pre-cluster: reduction of training time. (b) RMSE values at seven different lead times using pre-trained models based on four clusters for 60 sites randomly selected to serve as synthetic newly-built station. (c) RMSE of the ECMWF forecasts and LightGBM post-processed results at seven different lead times. The LightGBM results reflect average RMSEs for training data time spans ranging from zero to 24 months.

4.3 Comparison of different lead hour and four clusters: importance of features for forecast error correction

LightGBM, as a GBDT variant, is a 'grey box' AI algorithm. Information gain, split times, and coverage rate can be calculated for each feature and used to explain the results (Gilpin et al. 2018). For example, the raw (NWP) surface air temperature forecast contributes the most information for most lead times and cluster members when *K*=4 (Figure 2). Temperature observations are the second most influential feature, but make only marginal contributions in most cases. For clusters where the RMSE of the operational ECMWF forecasts is already relatively small, such as cluster 2, the NWP forecasts account for a larger proportion of the overall influence.

Conversely, observed temperatures play a larger role for clusters with larger RMSEs in the operational forecasts, such as cluster 4. The importance of the operational forecasts also increases as lead time increases, with concomitant reductions in the importance of the direct observations.

5. Conclusions

ML algorithms show great potential for post-processing numerical weather forecasts, but their application is often restricted by the number of available observations. Therefore, we propose the CDT framework, based on clustering, decision tree, and transfer learning, and assess its performance in post-processing ECMWF forecasts of surface air temperature at lead times ranging from 6 to 168 h for 301 weather stations in China. The stations are first divided into two, four, and eight clusters, as these classifications produce climatologically and geographically meaningful station groupings. The CDT framework reduces the average RMSE of temperature forecasts at the 301 stations by up to 0.81°C (27.9%). These benefits are seen for all

clustering scenarios and at all lead times, but the greatest improvements are for the 4-cluster scenario at 6-24 h lead times. Transfer learning aids the extension of models trained on data-rich stations to data-sparse stations within the same cluster. The RMSE at new stations is reduced by 16.8% (0.56°C) relative to the raw ECMWF forecasts even without fine-tuning, rising to 36.4% (1.23°C) once one year of observations is available for fine-tuning the algorithm. These improvements illustrate the great potential of the CDT framework for operational model post-processing, since newly-built sites typically suffer from short data records that restrict the application of AI techniques. An attractive feature of decision tree-based models is that the results can be explained in terms of the contributions from each input feature. Here the main contribution is from the raw ECMWF forecast, especially at longer lead times. However, the station temperature observations are the most important contributor for short lead times at stations in cluster 4, where the operational forecasts are less accurate than in other clusters. Overall, the CDT framework can help to correct prediction biases between NWP and observations, especially for newly-built stations or sites with sparse data records.

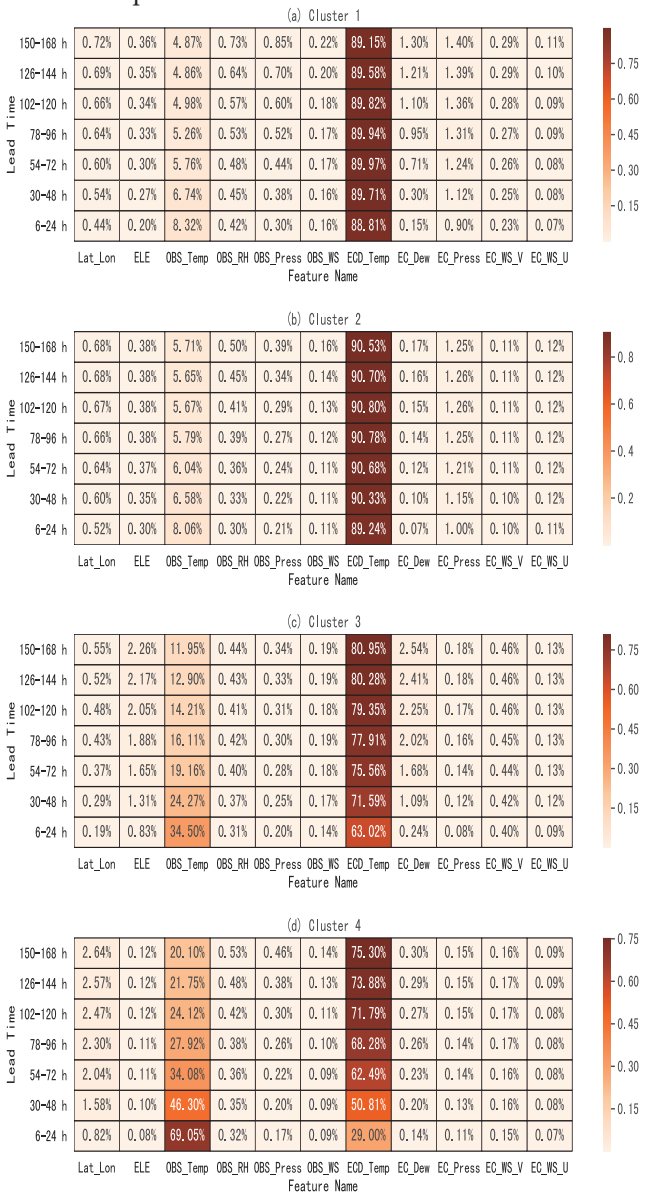

Figure 2: The relative importance of features at different lead times and for different clusters. The 'EC' prefix indicates variables from the original ECMWF forecasts, while the 'OBS' prefix indicates direct observations. Temp stands for temperature; RH for relative humidity; Press for surface pressure; WS for wind speed; dew for dew point temperature; WS_U and WS_V for the zonal and meridional components of wind speed, respectively; Lat_Lon for the latitude and longitude of the station; and ELE for the elevation of the station. The cluster numbers correspond to the K=4 clustering result.

Acknowledgements

This work is supported by the National Key Research and Development Program of China (2016YFB0201100, 2020YFA0607900, 2018YFB0505003) and the National Natural Science Foundation of China (42075137, 41776010).

References

Agapiou, A., 2017: Remote sensing heritage in a petabyte-scale: satellite data and heritage Earth Engine (c) applications. *Int. J. Digit. Earth*.

Bador, M., P. Naveau, E. Gilleland, M. Castella, and T. Arivelo, 2015: Spatial clustering of summer temperature maxima from the CNRM-CM5 climate model ensembles & E-OBS over Europe. *Weather Clim. Extremes*, **9**, 17-24.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47-55.

Boers, N., B. Goswami, A. Rheinwalt, B. Bookhagen, B. Hoskins, and J. Kurths, 2019: Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature*, **566**, 373-377.

Bougeault, P., and Coauthors, 2010: The thorpex interactive grand global ensemble. *Bull. Amer. Meteorol. Soc.*, **91**, 1059-1072.

Cao, Y., and L. Gui, 2018: Multi-Step wind power forecasting model Using LSTM networks, Similar Time Series and LightGBM. *2018 5th International Conference on Systems and Informatics (ICSAI)*, 192-197.

Chen, T., and C. Guestrin, 2016: XGBoost: A Scalable Tree Boosting System. *International Conference on Knowledge Discovery and Data Mining*, **ACM**, 785-794.

Elsa, B., N. Philippe, V. Mathieu, and M. Olivier, 2013: Clustering of Maxima: Spatial Dependencies among Heavy Rainfall in France. *Journey of Climate*, **26**, 7929-7937.

Gensler, A., J. Henze, B. Sick, N. Raabe, and Ieee, 2016: Deep Learning for Solar Power Forecasting - An Approach Using Autoencoder and LSTM Neural Networks. *2016 IEEE International Conference on Systems, Man, and Cybernetics*, 2858-2865.

Gentine, P., M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, 2018: Could Machine Learning Break the

Convection Parameterization Deadlock? *Geophys. Res. Lett.*, **45**, 5742-5751.

Gilpin, L.H., D. Bau, B.Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, 2018: Explaining Explanations: An Overview of Interpretability of Machine Learning. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 80-89.

Ham, Y.-G., J.-H. Kim, and J.-J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, 573, 568.

Harry, R.G., and A.L. Dale, 1972: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. J*ournal of applied meteorology* (1962), **11,** 1203-1211.

Hastie, T., R. Tibshirani, and J. Friedman, 2009: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition ed. New York, NY: Springer New York.

Hu, Q., R. Zhang, and Y. Zhou, 2016: Transfer learning for short-term wind speed prediction with deep neural networks. *Renewable Energy*, **85**, 83-95.

Hwang, J., P. Orenstein, J. Cohen, K. Pfeiffer, and L. Mackey, 2019: Improving Subseasonal Forecasting in the Western U.S. with Machine Learning. *International Conference on Knowledge Discovery and Data Mining*, **ACM**, 2325-2335.

Jiang, G.Q., J. Xu, and J. Wei, 2018: A Deep Learning Algorithm of Neural Network for the Parameterization of Typhoon-Ocean Feedback in Typhoon Forecast Models. *Geophys. Res. Lett.*, **45**, 3706-3716.

Ju, Y., G. Sun, Q. Chen, M. Zhang, H. Zhu, and M. U. Rehman, 2019: A Model Combining Convolutional Neural Network and LightGBM Algorithm for Ultra-Short-Term Wind Power Forecasting. *IEEE Access*, **7**, 28309-28318.

Ke, G., and Coauthors, 2017: LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds.

Lynch, P., 2008: The origins of computer weather prediction and climate modeling. *J. Comput. Phys.*, **227**, 3431-3444.

McGovern, A., and Coauthors, 2017: Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather. *Bull. Amer. Meteorol. Soc.*, **98**, 2073-2090.

Myers, D.E., 1994: Spatial interpolation: an overview. *Geoderma*, **62**, 17-28.

Overpeck, J.T., G.A. Meehl, S. Bony, and D.R. Easterling, 2011: Climate Data Challenges in the 21st Century. *Science*, **331**, 700-702.

Pan, B., K. Hsu, A. AghaKouchak, and S. Sorooshian, 2019: Improving Precipitation Estimation Using Convolutional Neural Network. *Water Resources Research*, **55**, 2301-2321.

Qureshi, A.S., and A. Khan, 2019: Adaptive transfer learning in deep neural networks: Wind power prediction using knowledge transfer from region to region and between different task domains. *Computational Intelligence*, **35**, 1089-1113.

Rasp, S., and S. Lerch, 2018: Neural Networks for Postprocessing Ensemble Weather Forecasts. *Monthly Weather Review*, **146**, 3885-3900.

Rasp, S., M. S. Pritchard, and P. Gentine, 2018: Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, 9684-9689.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, 2019: Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**, 195-204.

Rousseeuw, P.J., 1987: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, **20**, 53-65.

Scher, S., 2018: Toward Data-Driven Weather and Climate Forecasting: Approximating a Simple General Circulation Model With Deep Learning. *Geophys. Res. Lett.*, **45**, 12616-12622.

Schneider, T., S. Lan, A. Stuart, and J. Teixeira, 2017: Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations. *Geophys. Res. Lett.*, **44**, 12,396-312,417.

Shi, X., Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, 2015: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Nips*, 2-3.

Sinno Jialin, P., and Y. Qiang, 2010: A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, **22**, 1345-1359.

Swinbank, R., and Coauthors, 2016: THE TIGGE PROJECT AND ITS ACHIEVEMENTS. *Bull. Amer. Meteorol. Soc.*, **97**, 49-67.

Zaytar, M.A., and C.E. Amrani, 2016: Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks. *International Journal of Computer Applications*, **143**, 7-11.

No. 81, November 2021

The CLIVAR Exchanges is published by the International CLIVAR Project Office
ISSN No: 1026-0471

Editor: Jose Santos (ICPO)
Guest editor: Jing-Jia Luo (NUIST)**, and Swadhin Behera** (JAMSTEC)

Layout and Designing: Lian Peng (Institute of Oceanology, CAS)

This issue's DOI: 10.36071/clivar.81.2021

WCRP is sponsored by the World Meteorological Organization, the International Science Council and the Intergovernmental Oceanographic Commission of UNESCO.

Contact: Executive Director, ICPO
First Institute of Oceanography, MNR
6 Xianxialing Road, Laoshan District, Qingdao 266061, China.
icpo@clivar.org
http://www.clivar.org