



## RESEARCH ARTICLE

10.1029/2022MS003596

## Special Section:

Machine learning application to  
Earth system modeling

## Key Points:

- Ensemble convolutional neural networks (CNNs) are trained to emulate seasonal tropical cyclone (TC) activity using environmental factors
- The trained CNNs can be utilized to study seasonal TC variability, and their changes in the past, current and future climates
- Skillful seasonal TC predictions can be made using CNN-based statistical-dynamical hybrid framework

## Supporting Information:

Supporting Information may be found in  
the online version of this article.

## Correspondence to:

D. Fu,  
fudan1991@tamu.edu

## Citation:

Fu, D., Chang, P., & Liu, X. (2023).  
Using convolutional neural network to  
emulate seasonal tropical cyclone activity.  
*Journal of Advances in Modeling Earth  
Systems*, 15, e2022MS003596. [https://doi.  
org/10.1029/2022MS003596](https://doi.org/10.1029/2022MS003596)

Received 22 DEC 2022

Accepted 5 SEP 2023

Using Convolutional Neural Network to Emulate Seasonal  
Tropical Cyclone Activity

Dan Fu<sup>1</sup> , Ping Chang<sup>1,2</sup> , and Xue Liu<sup>1</sup>
<sup>1</sup>Department of Oceanography, Texas A&M University, College Station, TX, USA, <sup>2</sup>Department of Atmospheric Sciences,  
Texas A&M University, College Station, TX, USA

**Abstract** It has been widely recognized that tropical cyclone (TC) genesis requires favorable large-scale environmental conditions. Based on these linkages, numerous efforts have been made to establish an empirical relationship between seasonal TC activities and large-scale environmental favorability in a quantitative way, which lead to conceptual functions such as the TC genesis index. However, due to the limited amount of reliable TC observations and complexity of the climate system, a simple analytic function may not be an accurate portrait of the empirical relationship between TCs and their ambiances. In this research, we use convolution neural networks (CNNs) to disentangle this complex relationship. To circumvent the limited amount of seasonal TC observation records, we implement transfer-learning technique to train ensemble of CNNs first on suites of high-resolution climate model simulations with realistic seasonal TC activities and large-scale environmental conditions, and then on a state-of-the-art reanalysis from 1950 to 2019. The trained CNNs can well reproduce the historical TC records and yields significant seasonal prediction skills when the large-scale environmental inputs are provided by operational climate forecasts. Furthermore, by inputting the ensemble CNNs with 20th century reanalysis products and Phase 6 of the Coupled Model Intercomparison Project (CMIP6) simulations, we investigated TC variability and its changes in the past and future climates. Specifically, our ensemble CNNs project a decreasing trend of global mean TC activity in the future warming scenario, which is consistent with our future projections using high-resolution climate model.

**Plain Language Summary** Tropical cyclones (TCs) require favorable large-scale environmental conditions to form. Pioneer studies show that these conducive conditions include warm sea surface temperatures (SSTs), sufficient low-level vorticities and mid-level humidities, as well as weak-to-moderate vertical wind shears. Several follow-up studies have focused on improving the empirical linkage between number of TC and environmental conditions and developed sets of TC genesis index based on conventional statistical methods. Although these indices can capture climatology of TC genesis spatial distributions and seasonal variation reasonably well, their representations of interannual TC variability are degraded. With the aim to better represent TC interannual variability and long-term trend using large-scale environmental conditions, we trained ensembles of convolution neural networks (CNNs) based on the combination of observations and large sets of high-resolution dynamical climate simulations. The trained CNNs perform significantly well in capturing observed TC interannual-to-multidecadal variability and are broadly applicable to many areas of seasonal TC activities. Using a deep learning technique, this paper introduces a new potential avenue to improve our understanding of TC variability and future changes.

## 1. Introduction

Mounting observational and modeling evidence has suggested that tropical cyclone (TC) genesis requires certain conducive large-scale environmental conditions. Pioneering work of Gray (1968) summarized the geographic distributions and annual cycle climatology of TC genesis, and trailblazed potential linkages between TC genesis and large-scale environmental factors. Gray (1979) proposed a Yearly Genesis Parameter (YGP) and elucidated the constructive environmental factors for TC genesis that include warm sea surface temperature (SST) and deep oceanic mixed layer, preexisting low-level cyclonic vorticity, moist mid-troposphere, weak vertical wind shear, and weak atmospheric stability. Substantial progress has been made since then to improve the quantitative linkage between TC genesis and large-scale environmental favorability, and other well-known TC genesis indices have been developed (Bruyère et al., 2012; Emanuel & Nolan, 2004; Tippett et al., 2011; Wang & Murakami, 2020).

© 2023 The Authors. Journal of  
Advances in Modeling Earth Systems  
published by Wiley Periodicals LLC on  
behalf of American Geophysical Union.  
This is an open access article under  
the terms of the [Creative Commons  
Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License,  
which permits use, distribution and  
reproduction in any medium, provided the  
original work is properly cited and is not  
used for commercial purposes.

These genesis indices are based on the same principal, which assumes that favorable large-scale environmental conditions to TC genesis are products of dynamical potentials alone or a combination of dynamical and thermal potentials with different formulations. For example, the genesis potential index developed by Emanuel and Nolan (2004; hereafter GPI2004) describes a nonlinear multiplication of four major large-scale climate variables in triggering TC genesis, including absolute vorticity at 850 hPa, relative humidity at 600 hPa, vertical wind shear between 850 and 200 hPa, and theoretical TC maximum potential intensity (Bister & Emanuel, 1998). Emanuel (2010) further modified GPI2004 by replacing mid-level relative humidity with the moist entropy deficit and updated the nonlinear multiplication formulation, which has particular importance in addressing issues of TC response to climate change. Following Gray (1968), Emanuel and Nolan (2004), and Emanuel (2010), Tippett et al. (2011) later proposed an improved variant of TC genesis index using similar dynamical and thermal variables but with Poisson regression, and highlighted the need to cap off the contribution from low-level absolute vorticity to TC genesis when it exceeds certain thresholds. All the aforementioned TC genesis indices employed both dynamical and thermal climate variables in their formulations, but the relative contributions of dynamical and thermal variables to TC genesis remain largely elusive. More recently, Wang and Murakami (2020) developed the dynamical genesis potential index (hereafter, DGPI) for recognition of TC genesis potentials purely based on large-scale dynamical factors and showed improved skills in representing TC genesis variations especially under future warmer climate.

Although these empirical TC genesis indices are able to replicate the seasonal cycle and spatial distribution of observed TC genesis over the globe, their skills in representing TC interannual variability in various ocean basins are degraded and highly basin dependent (Menkes et al., 2012). For instance, GPI2004 can capture TC interannual variability reasonably well in the North Atlantic and eastern North Pacific, but not in the western North Pacific (Yu et al., 2018). On the other hand, DGPI can skillfully represent western North Pacific and North Atlantic TC interannual variability but works less well in the eastern North Pacific (Wang & Murakami, 2020). Current TC genesis indices generally have poor skill in representing TC interannual variability in the Indian Ocean. These results may be attributable to a couple of reasons. First, the existing TC genesis indices were derived from climatological mean values, thus information about TC interannual variability was not incorporated. Nonetheless, these indices can reproduce the spatial shift pattern of TC activity modulated by El Niño-Southern Oscillation (ENSO) as in the observations, which is a major driver of TC interannual variability. Second, the derivation of the current TC genesis indices did not consider the spatial variations among different TC-active basins, casting doubt on the applicability of their global usage. For instance, TC genesis mechanisms in the western North Pacific differ from those in the North Atlantic due to strong convection over the Indo-Pacific Warm Pool area. Murakami and Wang (2010) developed a variant of GPI2004 by incorporating 500 hPa vertical velocity as an additional term into the original formulation, leading to a significant improvement in the western North Pacific. Similarly, a more recent study by Wang and Murakami (2020) noted that the selected dynamic potential factors through stepwise regression method are basin-dependent, and it is advantageous to derive DGPIs for different TC basins to optimize their accuracy. Therefore, a single empirical formulation for global TC genesis may have limited accuracy in representing TC interannual variability.

TC genesis indices have also been extensively used to understand future changes of TC activities. However, the relevance of the thermodynamic factors used in TC genesis indices to explain TC changes under global warming have been questioned. Recent studies by Camargo et al. (2014) and Lee et al. (2020) reported two divergent projections of future TC activity due to the different choices of thermodynamic variables in the TC genesis index of Tippett et al. (2011). Projections using saturation deficit indicate a decreasing trend, while those using column relative humidity indicate an increasing trend. There is no established theory that can account for this discrepancy, given that both indices yield similar results for the historical period. We hypothesize that an improved representation of TC genesis indices that can better represent TC year-to-year variations may also lead to uncertainty reduction in the TC future projections. In this study, we attempt to test this hypothesis by incorporating information of TC interannual variability into derivation of a new TC genesis index, based on convolutional neural networks (CNNs), and use them to investigate TC variability and long-term changes under past, present, and future climate conditions.

In recent years, CNN has been widely utilized in the climate and weather studies and recognized as a powerful tool to build linkages between predictor fields and predictand classifications or physical indices (Bolton & Zanna, 2019; Chapman et al., 2019; Chattopadhyay, Hassanzadeh, & Pasha, 2020; Chattopadhyay, Nabizadeh, & Hassanzadeh, 2020; Davenport & Diffenbaugh, 2021; Ham et al., 2019; Liu et al., 2016; McGuire & Moore, 2022;

Weyn et al., 2019, 2021). In the context of climate study, Ham et al. (2019) innovatively applied CNNs for multi-year ENSO predictions and showed the CNN-based forecasts are more skillful than conventional dynamical forecasts of Niño3.4 index at long lead times. Inspired by Ham et al. (2019) and previous studies on TC genesis potential indices, in the present study, we aim to develop CNN-based frameworks to emulate basin-integrated seasonal TC activity using the concurrent large-scale environmental factors. Taking one step further from existing TC genesis indices that estimate TC genesis possibility our CNN models use seasonal-averaged thermal and dynamical large-scale environmental factors as predictors, and basin-integrated seasonal mean TC metrics, such as number of TCs (NTC), number of hurricanes (NHU; TCs with lifetime maximum intensity  $\geq 65$  knots), and Accumulated Cyclone Energy (ACE; Bell et al., 2000), as predictands to depict seasonal TC activities. As such, our framework can emulate not only TC numbers, but also averaged intensity for a given season. We will show that our CNN framework can significantly improve the representation of TC interannual-to-multidecadal variability. We will further explore the applicability of the CNN frameworks derived from the present-day climate to seasonal TC predictions, reconstructions of past historical TC records, and projections of future TC changes.

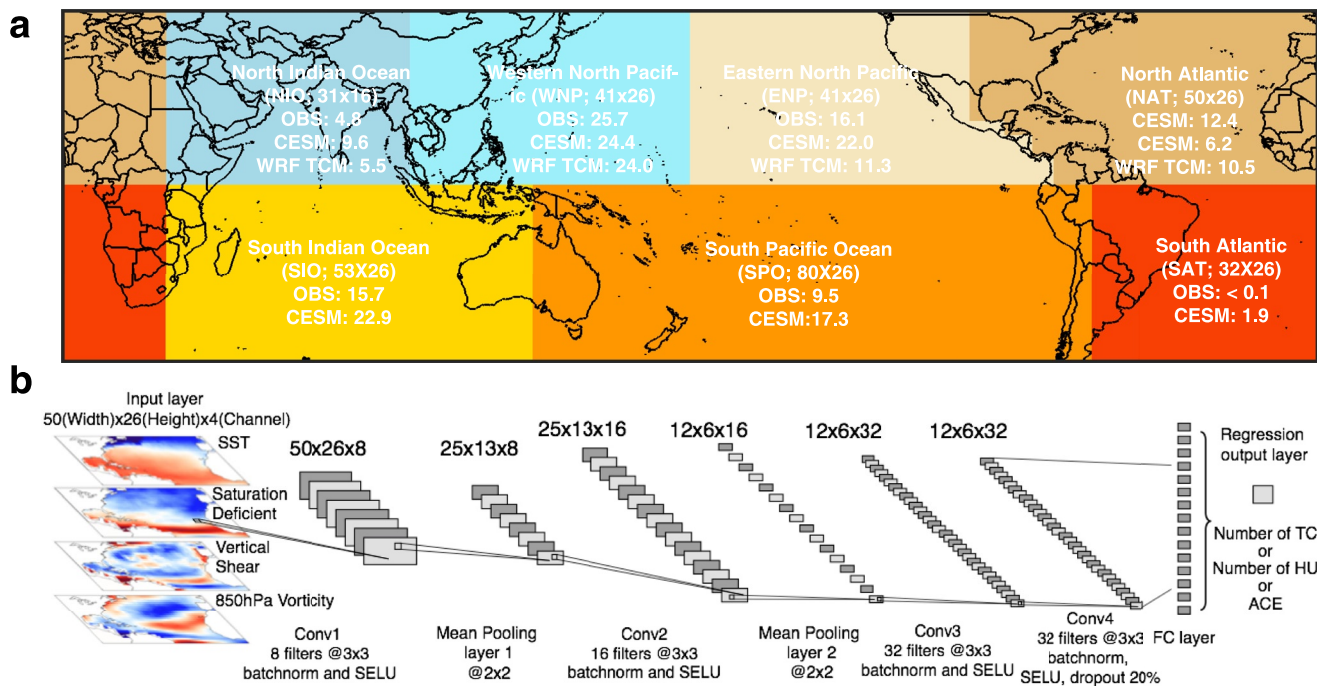
The structure of the paper is as follows. In Section 2, the CNN-based TC emulation framework, including the architecture, training, and interpretation, is discussed. Applications of CNN framework to seasonal TC prediction, 20<sup>th</sup> century historical TC reconstructions, and future projections are presented in Section 3, 4, and 5, respectively. The summary and discussion are given in Section 6.

## 2. CNN Model Development

### 2.1. Data for CNN Training

As discussed earlier in Section 1, large-scale environmental favorability on seasonal TC activity is geographic-dependent. As such, we trained various CNNs targeted for seven different TC active basins, including North Atlantic (NAT), eastern North Pacific (ENP), western North Pacific (WNP), North Indian Ocean (NIO), South Indian Ocean (SIO), South Pacific Ocean (SPO), and South Atlantic (SAT), and for three different seasonal mean TC metrics (i.e., NTC, NHU and ACE; total of 21 CNN models). Figure 1a illustrates the boundary of each basin. For the Northern (Southern) Hemisphere basins, we only focus on June–November (December–May), which is conventionally defined as TC season. Following previous works on TC genesis indices, we used 4 environmental factors as predictors, namely, SST anomalies, saturation deficit (SD; Camargo et al., 2014) anomalies, 850 hPa relative vorticity anomalies, and vertical wind shear anomalies between 850 and 200 hPa. We also tested the sensitivity of using column-integrated relative humidity (CRH; Tippett et al., 2011) anomalies as an alternative humidity environmental factor to SD in our CNN model (details in Section 5). Both SD and CRH are calculated following Bretherton et al. (2004). To derive these environmental factors, we first interpolate the related prognostic monthly mean data (i.e., temperature, humidity and wind velocities) onto a  $2^\circ \times 2^\circ$  grid, then make the diagnostic calculations and seasonal average. Anomalies are finally calculated as departures from the long-term seasonal mean. These four seasonal mean predictors are then concatenated (i.e., longitude, latitude and variables) as inputs to the CNN (Figure 1b). We note that, actual physical data, rather than images, were implemented as the CNN inputs in our framework. The dimension of the output is one, which depicts the seasonal mean NTC, NHU or ACE, respectively.

One major issue in implementing deep learning for climate studies is the limited sample size of available observational data for proper model training (Ham et al., 2019). Due to relative short period of reliable TC observation records, we used transfer-learning technique (Yosinski et al., 2014) to circumvent the limited amount of training data sets and increased the number of training data, in a similar manner as in Ham et al. (2019). To increase the training samples, we collected suites of high-resolution Community Earth System Model version 1.3 (CESM1.3;  $\sim 0.25^\circ$  resolution for atmosphere/land and  $\sim 0.1^\circ$  resolution for ocean/sea-ice) simulations, including 1850 preindustrial control and historical transient climate simulations (Chang et al., 2020), High Resolution Model Intercomparison Project (HighResMIP) 1950 control and historical transient climate simulations (Roberts et al., 2020), and decadal prediction simulations (Yeager et al., 2023), together with Weather Research and Forecast (WRF) based tropical channel model (TCM; 27 km) large ensemble hindcast simulations (Fu et al., 2019, 2021). Both the high-resolution CESM1.3 and WRF TCM demonstrate high fidelity in simulating observed climatology and interannual variability of seasonal mean TC activity and associated large-scale ambient environments (Figure 1a; Figure S1 in Supporting Information S1 present seasonal mean TC activities in each TC active basins; refer Chang et al., 2020; Fu et al., 2019 for more details). We combined these high-resolution climate simulations with



**Figure 1.** (a) Illustration of TC active ocean basin boundaries, including North Atlantic (NAT), eastern North Pacific (ENP), western North Pacific (WNP), north Indian Ocean (NIO), south Indian Ocean (SIO), south Pacific Ocean (SPO) and south Atlantic (SAT). The number in each TC basin shows the annual mean TC member in observation (OBS), high-resolution CESM and WRF tropical channel model (TCM) simulation results, and horizontal dimensions of input layer. (b) Architecture of the CNNs for emulating seasonal TC activity. The CNN model consists of one input layer (predictor; taking NAT as an example), four convolutional (Conv) layers, two mean pooling layers, one fully connected (FC) layer, and one output layer (the predictant). The variables of the input layer contain the anomalies of SST (unit: K), saturation deficit (unit: mm), 200–850 hPa vertical wind shear (unit: m/s), and 850 hPa relative vorticity ( $10^{-5}$  1/s) averaged over June–November (December–May) for Northern (Southern) Hemisphere ocean basins. The input variables are interpolated to  $2^{\circ} \times 2^{\circ}$  resolution grid, and the actual size of input layer varies with different ocean basins. The seasonal mean number of TC (NTC), number of hurricane (NHU), or Accumulated Cyclone Energy (ACE) is output (i.e., predictant).

the state-of-the-art European Center for Medium Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5; Hersbach et al., 2020) and observed TC records from the International Best Tracks Archive for Climate Stewardship (IBTrACS) data set v04 (Knapp et al., 2010) to optimally train the CNN. It is necessary to point out that we only used the high-resolution CESM1.3 historical simulations, including both the preindustrial-control and 1950-control, as well as the transient climate simulations before 1979, to train our CNN model. We implement our future climate simulations as additional segments for testing processes, aiming to explore the feasibility of applying our CNN model to the TC future projections. Table 1 summarizes the sample size of each data set utilized in CNN training. We note that WRF TCM simulations only cover the Northern Hemisphere, and thus the sample size used to train the CNN is larger for the Northern (2906) than Southern (1506) Hemisphere TC active basins.

## 2.2. CNN Architecture

We first trained the CNN model on our high-resolution climate simulations from scratch, and the trained weights are then transferred to the new training processes as initial weights to formulate fine-tuned CNN model using ERA5 reanalysis and IBTrACS TC observations. Figure 1b illustrates the architecture of our CNN model. The input layer has 4 variables containing seasonal mean SST, SD (or CRH), 850 hPa relative vorticity, and vertical wind shear anomalies. Sizes of input layer vary from basin-to-basin (refer Figure 1a for actual sizes). 4 convolutional processes involve the extraction of local characteristics from the previous layers, with 8, 16, 32 and 32 filters, respectively. Each filter has a kernel size of  $3 \times 3$  with stride of 1. In each convolutional layer, zero padding around the borders of inputs is used to maintain the size before and after applying the filters. Mean-pooling processes reduce the sample size by extracting the mean value from each  $2 \times 2$  grid. The Scaled Exponential Linear Unit (SELU; Klambauer et al., 2017) is used as the activation function to introduce nonlinearity in the extracted features. Compared with Rectified Linear Units (ReLU), SELU solves the problem of off-state as zero



**Table 1**

*The Data Sets for Training the Deep CNN Model*

	Data sets	Sample size	Reference
1	CESM1.3 1850 pre-industrial control simulation	Year 0338–0519 (total: 182)	Chang et al. (2020)
2	CESM1.3 1850–2100 transient climate simulation	Year 1877–1979 and 2-ensemble 1920–1979 (total: 223)	
3	CESM1.3 HighResMIP 1950s control simulation	Year 051–151 (total: 101)	Roberts et al. (2020)
4	CESM1.3 HighResMIP 1950–2100 historical transient climate simulation	Year 1950–1979 (total: 30)	
5	CESM1.3 decadal prediction	10-member ensemble 5-year predictions initialized in every other year from 1982 to 2016 (total: 900)	Yeager et al. (2023)
6	WRF Tropical Channel Model seasonal hindcast simulation	35-member ensemble 1979–2018 (total: 1400)	Fu et al. (2019)
7	ERA5 and IBTrACS	1950–2019 (total: 70)	Hersbach et al. (2020) and Knapp et al. (2010)
	Total: 2906 (1506) for Northern (Southern) Hemisphere basins		

gradients by passing negative values to the next layer. This enables SELU to train deep CNNs effectively because there is no vanishing gradient problem (Moon et al., 2019). The fourth convolutional layer is linked to dropout regularization with dropout probability of 0.2 to reduce the overfitting (Srivastava et al., 2014). The CNN constructs the final prediction through a linear regression output layer, which regresses the desired output onto the intermediate results from the fully connected layer. CNN is trained using a form of stochastic gradient descent, namely, the Adaptive Movement Estimation (Adam) optimization algorithm (Kingma & Ba, 2014), which minimizes the loss function of mean square error (MSE) between the outputs of the CNN and the desired targets. Table 2 lists the key parameters of our CNN model.

To minimize the CNN regression output layer model uncertainty due to random weight initialization and stochastic gradient descent, we conducted the ensemble learning. For each targeted basin and TC metric, we independently trained the CNN for 50 times (50-member ensemble) with the same configuration and different initial random weights on the same data set. Hence, 1050 sets of CNN were trained based on the high-resolution climate model simulations to cover 50-member ensemble of 3 different seasonal mean TC metrics in 7 different ocean basins. These trained weights were then used as initial weights to formulate the final CNN model with the training sets confined to the ERA5 reanalysis and IBTrACS observation for the period 1950 to 2019. We note that TC observations are highly uncertain in pre-satellite era without the support from geostationary satellite imagery (i.e., 1960s) and may significantly underestimate seasonal TC activity (Landsea, 2007; Landsea & Franklin, 2013; Vecchi & Knutson, 2008, 2011; Vecchi et al., 2021). However, we still include these records in our CNN training processes to enlarge the sample size. We have conducted a series of sensitivity analyses using transfer-learning solely with post-satellite era ERA5 data from 1979 to 2019, and the details are discussed in Section 2.3.

Due to the limited size of our training sample (2906 for training-from-scratch and 70 for transfer learning), we did not split our data set into conventional training, validation, and test sets. We used entire observation data as the training and validation sets (70% for training and 30% for validation, shuffled during each training epoch), and we used leave-one-out cross validation method (LOOCV; Elsner & Jagger, 2013) to evaluate the CNN model skills and avoid overfitting due to excessive learning parameters and convolutional layers. LOOCV has been extensively used in the seasonal TC activity training (Li et al., 2013; Murakami et al., 2016). In the LOOCV, we first exclude a single year of observations and predictors; then, we determine the CNN model using the remaining years. Using the model, the seasonal TC metrics for the excluded year are predicted. This is done for 70 years, removing each year's data successively. In addition, we also cross validate CNN using 20th century reanalysis products, instead of ERA5, focusing on the NAT TC activities before 1950, which were excluded in the training processes. The details are discussed in Section 4.

For each of climate model pre-trained CNN, we further deployed 12-member ensemble trainings during the transfer-learning processes on ERA5 and IBTrACS. Although the initial training weights and biases are identical, stochastic Adam optimizer results in deviated CNN regression models due to random choice of mini batch. In total, for each of TC-active ocean basins and each of seasonal mean TC metrics, we trained ensemble of CNNs with 600-members. Each model is then used to make a prediction and the actual prediction skill is evaluated by averaging all ensembles, while the ensemble spread indicates the prediction uncertainties.

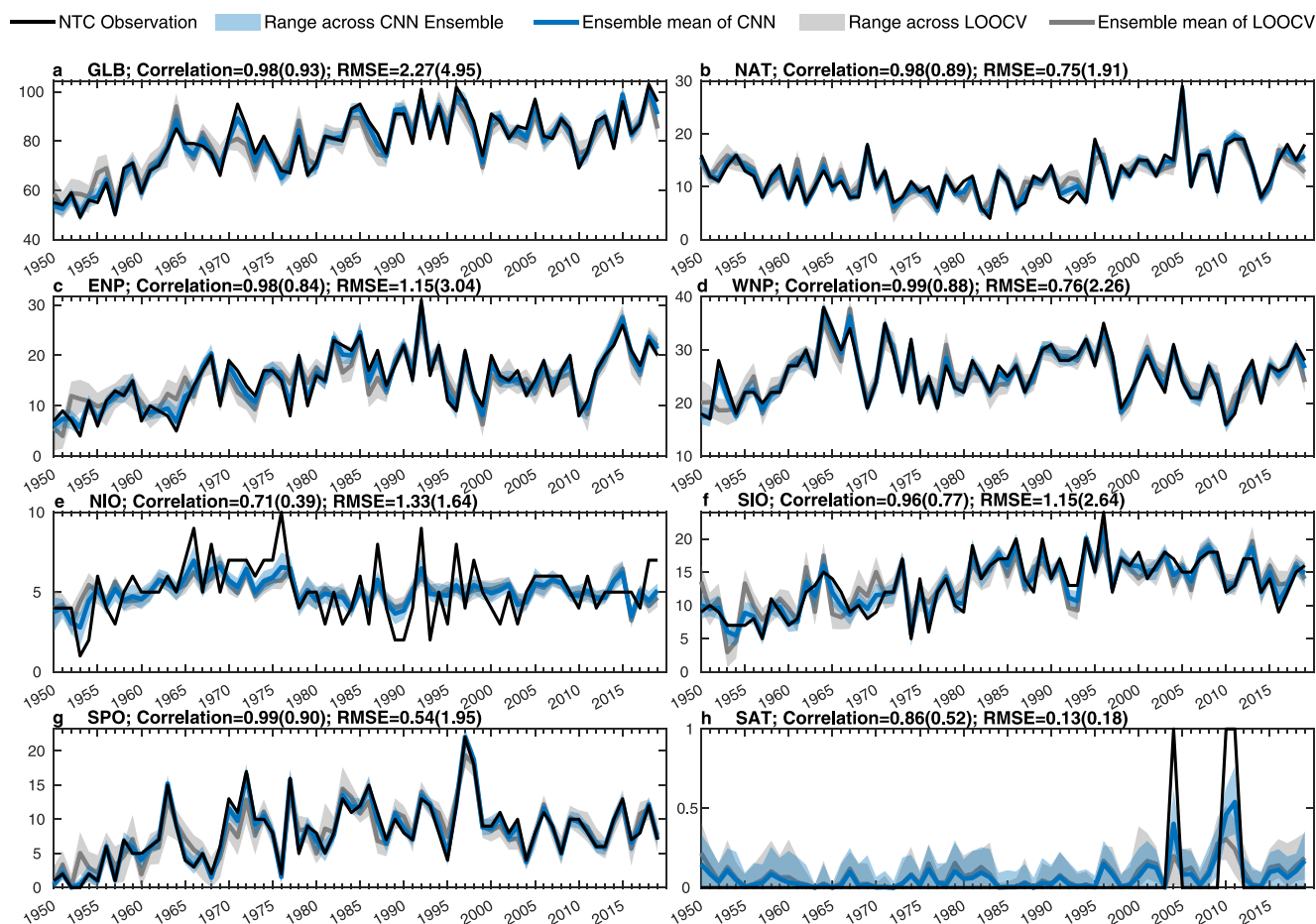
### 2.3. CNN Training, Validation, and Interpretation

In the rest of this paper, for the sake of clarity, CNN emulated NTC, NHU and ACE are always shown in blue, red, and yellow, respectively. Figure 2 illustrates the training and LOOCV skills of seasonal NTC in each TC active ocean basin. The results show high skills in reproducing NTC variability, in terms of high Pearson correlation coefficients, mean-square skill scores (MSSS; Kim et al., 2012; Li et al., 2013), and low root mean square errors (RMSE) over majority basins except for the NIO. For the NIO, the correlation coefficient decreases from 0.71 in the training processes to 0.39 in the LOOCV. Although LOOCV correlation coefficient remains statistically

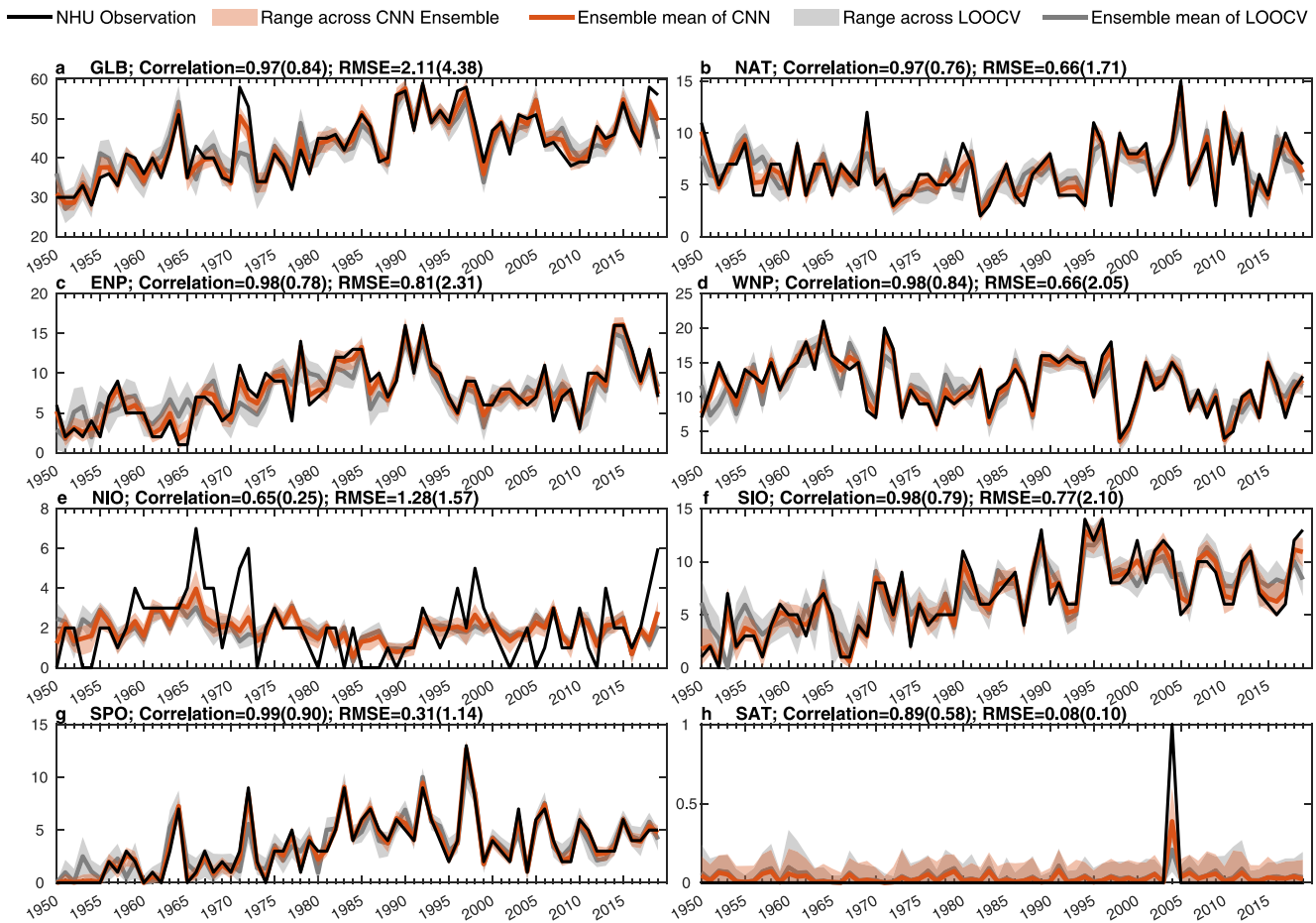
**Table 2**

*Detail of CNN Architecture and the Optimization Parameters*

Number of convolution layers	4
Number of filters for each convolution layers	8, 16, 32, 32
The dimensions of the convolutional filter	$3 \times 3$ , $3 \times 3$ , $3 \times 3$ , $3 \times 3$
Activation function for each convolutional filter	SELU, SELU, SELU, SELU
Mean pooling kernel size	$2 \times 2$
Stride for mean pooling	2
Output size	1
Loss function	Mean square error
Optimizer	Adam
Max training epochs	100
Mini batch size for training	500(climate model)/20(observation fine-tuning)
Initial learning rate	0.001



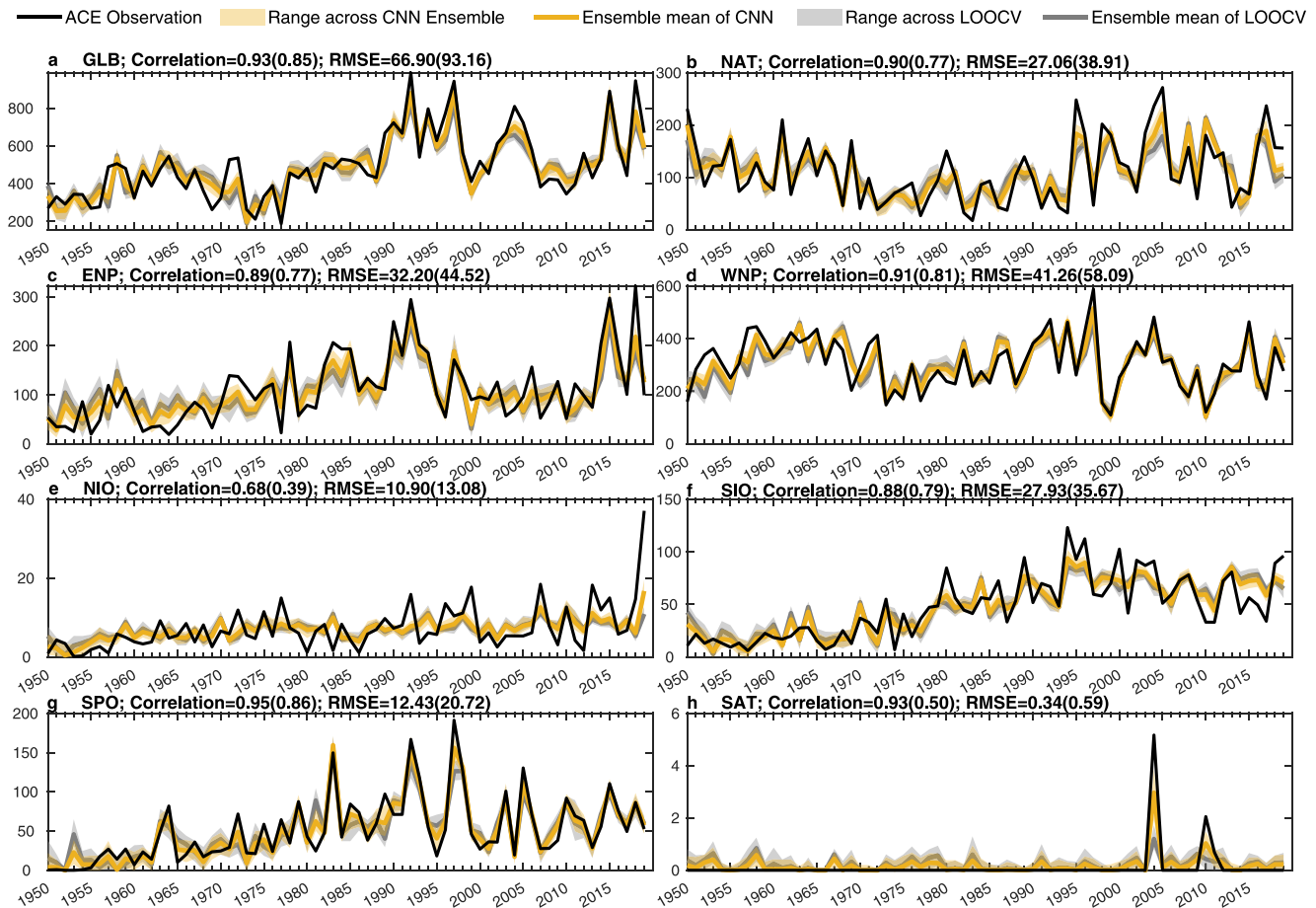
**Figure 2.** Time-series of seasonal mean number of TC (NTC) for (a) global (defined as the sum of individual ocean basins), (b) NAT, (c) ENP, (d), WNP, (e) NIO, (f) SIO, (g) SPO and (h) SAT. Observed NTC is shown in black line. Ensemble mean CNN emulated and LOOCV results are shown in blue and gray lines, respectively. Ranges across CNN ensembles are shown in shadings. Numbers shown in each panel denote the Pearson correlation coefficients and root mean square errors (RMSE) between observation and CNN model results. Leave-one-out cross validation (LOOCV) results are listed in the parentheses.



**Figure 3.** Similar as to Figure 2, but for seasonal mean number of hurricane (NHU; lifetime maximum wind speed  $\geq 33$  m/s).

significant at 99% level, it is notably lower compared to those in the other TC active basins. This discrepancy might indicate that certain crucial large-scale environmental factors were not incorporated in the current CNN predictor variables for the NIO TC activity emulations. The CNN emulations for the SAT NTC show small local maxima during the years with observed TCs (i.e., 2004, 2010, and 2011), with an NTC value of around 0.5. Although correlation coefficient is considerably high, these values do not hold physical significance, as it is not possible to have a “half” TC in a season. Moreover, the emulated SAT TC activity demonstrates weak interannual variation, which is not consistent with the observations. Therefore, we consider the validation in the SAT sector less reliable than in other basins, except for the NIO, due to the scarcity of TCs in this area. In the context of global mean TC activity, the correlation coefficients between observation and ensemble mean CNN training and LOOCV are 0.98, and 0.93, respectively. RMSE between observation and CNN emulated global NTC is 2.27 and 4.95 in training and LOOCV ensemble mean, both of which are significantly smaller than the observed NTC climatology of 79.27 and standard deviation of 12.85. The ensemble spreads of the training and LOOCV, which are defined as the entire range across 600-member ensemble CNNs and shown as the shadings in Figure 2, are also small among all TC active basins. Similar high skills are also found in the ensemble CNNs for seasonal mean NHU (Figure 3) and ACE (Figure 4) during the period of 1950–2019. These promising assessment results encourage us to proceed with further applications using the CNN framework.

Additionally, we performed a series of sensitivity analyses to evaluate the impact of the sample size of observational data on the accuracy of transfer-learning CNNs. In these analyses, we only feed pre-trained CNNs with 1979–2019 ERA5 data. The results of the sensitivity analyses are shown in Figure S2 in Supporting Information S1. In general, we observed that RMSE in NTC, NHU, and ACE roughly doubles when the transfer learning is confined to the period from 1979 to 2019. The degradation in Pearson correlation coefficients is relatively smaller, but there is a notable increase in disparities between the LOOCV skills and training skills within the limited 1979–2019 transfer



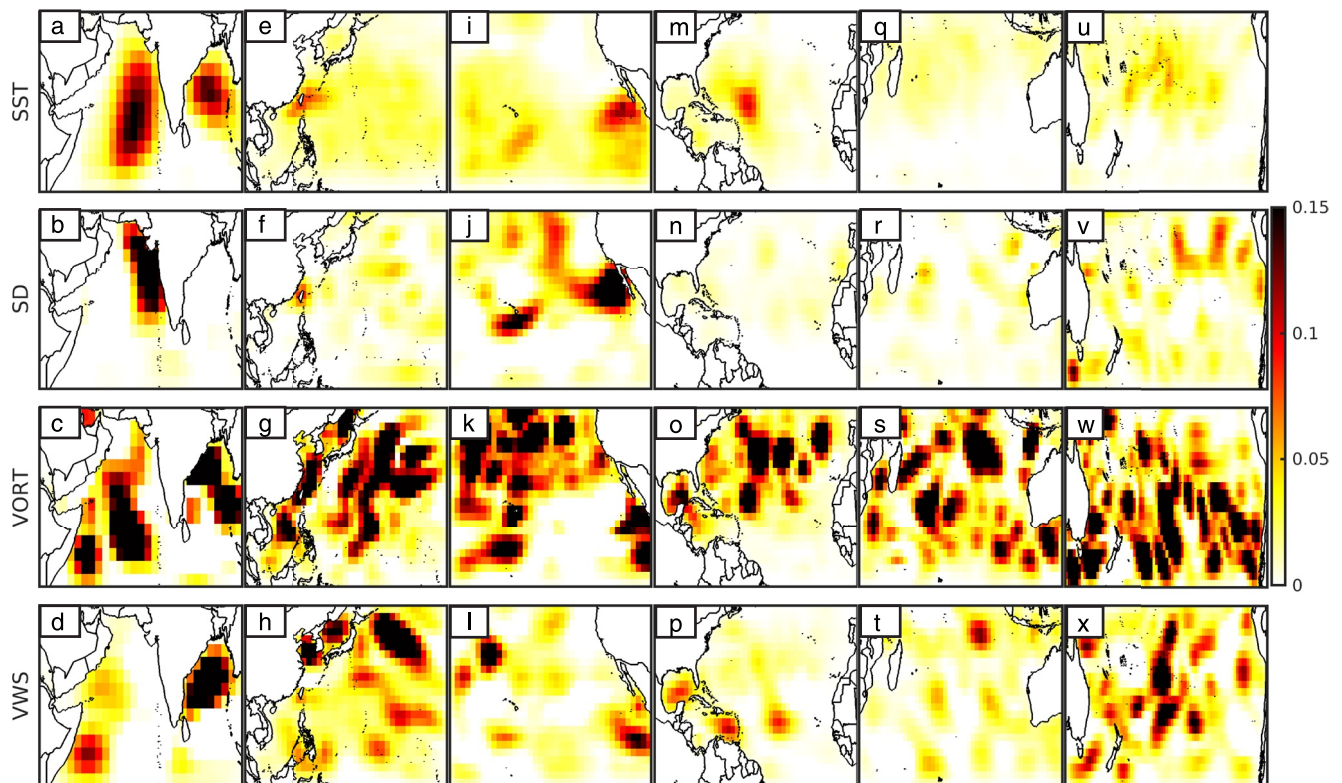
**Figure 4.** Similar as to Figure 2, but for seasonal mean accumulated cyclone energy (ACE).

learning data set (as illustrated in Table 1). These findings suggest that relying exclusively on post-satellite reanalysis data for transfer learning would significantly impair the accuracy of the transfer-learning CNNs in our study.

Although our trained CNN ensemble frameworks demonstrate high skill in reproducing observed seasonal TC variability, an often-cited caveat of machine learning is the challenge of physical interpretability compared to more conventional statistical methods. The lack of physical insights has implications for the perceived credibility of the approach, in which machine learning models may achieve promising results for wrong reasons (Lapuschkin et al., 2019). To decipher these “black boxes” and gain insight into what predictor features that predominate CNN’s regression output, we conducted the occlusion sensitivity analysis (Zeiler & Fergus, 2014).

The general concept of the occlusion sensitivity analysis is to check if the trained CNNs can truly identify the physical meaningful spatial patterns in the input layer by systematically occluding different portions of the predictor variables with an occluding mask and measuring the degree of changes as a function of mask position. Intuitively, if a portion of input layer is trivial to the prediction accuracy, and one was to synthetically mask this portion with “gray patch,” one can expect the degradation of prediction skill is also relatively small, and vice versa. As such, we can explore the relative importance of each predictor variable at different geographical locations in determining seasonal TC variability. In practice, this occlusion sensitivity was conducted by replacing a patch of  $6^\circ \times 6^\circ$  (i.e.,  $3 \times 3$  grid) box mask to each of four predictor variables during 1950–2019 with its own long-term mean climatological values, and new predictions are conducted by the original trained CNN with the altered inputs. The box mask moves across the each of four predictor variables, therefore, we obtained the new predictions of seasonal TC variability as functions of mask geographical locations and predictor variables. We then measure the change in prediction skills as the RMSE between new predictions and original training during 1950–2019 and use it to determine the regions of each predictor variable that are the most important to the prediction skills of our trained CNN framework: when that portion of the predictor is replaced with the climatology, the representation of seasonal TC variability will be significantly degraded.





**Figure 5.** Occlusion sensitivity maps that highlight the relative importance in emulating seasonal NTC in different ocean basins. (a–d): Relative importance of SST, saturation deficit (SD), 850 hPa vorticity and vertical wind shear in NIO, respectively. (e–h), (i–l), (m–p), (q–t), And (u–x) are similar, but for the relative importance of 4 variables in WNP, ENP, NAT, SIO, and SPO, respectively. Areas in the map with higher values correspond to regions of input variables that contribute more significantly to impact the CNN prediction skills. Intuitively, the sensitivity map shows which area most affect the prediction RMSE when changed. Refer main text for details.

Figure 5 shows occlusion sensitivity maps that highlight the relative importance of predictors in emulating the NTC in different ocean basins. Similar occlusion sensitivity maps that decipher the NHU and ACE emulations are illustrated in Figures S3 and S4 in Supporting Information S1, respectively. As noted earlier in the Section 1, it is not surprising to see that the relative importance of each predictor is different among different TC basins. Our occlusion sensitivity analysis broadly suggests that large-scale dynamical factors, especially 850 hPa vorticity, contribute more than thermal factors in determining NTC and NHU in each basin, which is consistent with the TC DGPI recently proposed by Wang and Murakami (2020). Taking NAT NTC as an example (Figures 5m–5p), 850 hPa vorticity at North Atlantic subtropical gyre and Gulf of Mexico region contributes most to the NAT NTC interannual variability, followed by vertical wind shear in the TC main development region and Gulf of Mexico, while SST's and SD's contributions are relatively small. We emphasize that these occlusion sensitivity results do not imply that SST are not important for regulating seasonal NTC and NHU, because local dynamical factors can be influenced by remote SST forcing. For example, extensive observational and modeling studies showed that Atlantic TC seasons are remotely influenced by tropical Pacific SST variability via ENSO teleconnection patterns through both a Walker Circulation type response and an upper tropospheric temperature response (Patricola et al., 2016; Smith et al., 2007; Tang & Neelin, 2004), while Atlantic SST variability can also remotely influence ENP TC activity through a Walker Circulation type response analogous to the ENSO-Atlantic TC teleconnection (Patricola et al., 2017).

Interestingly, the relative importance of thermal factors is increased substantially in the CNN ACE emulation. We hypothesize these enhanced thermal contributions may imply local SST and SD are more effective in influencing seasonal mean TC intensity than the TC occurrence. This hypothesis was partially examined by Murakami et al. (2018), who found that the hyperactive six major hurricanes during 2017 NAT TC season was primarily attributed to the record high SST anomaly confined to the tropical Atlantic, while the remote influence from the moderate La Niña in the Pacific played a minor role. More rigorous analyses need to be conducted in future

studies. Nonetheless, the occlusion sensitivity analysis provides an explanation to the high fidelity of CNN emulations that are consistent with our current understanding of TC dynamics. This gives us the confidence that our trained CNN frameworks are capable of learning physically plausible mechanisms, not only acting just as “black boxes,” to account for seasonal TC activities.

### 3. CNN-Based Statistical-Dynamical Hybrid Seasonal TC Prediction

With the properly developed CNN framework described above, in the following sections we propose a number of different pathways that we anticipate our machine learning model as a starting point to shed light on the future studies. We will start with the application of ensemble CNNs in seasonal TC prediction.

Seasonal TC prediction methods can be broadly categorized into dynamical, statistical, or statistical-dynamical hybrid approach (i.e., dynamical models combined with statistical approaches). As our CNN framework is based on concurrent relationships between seasonal averaged large-scale environmental factors and TC activities, we can forecast seasonal TC activities at various lead times by inputting the trained ensemble CNNs with predicted large-scale environmental variables from operational seasonal forecast systems. In this study, we utilized the National Center for Environmental (NCEP) Prediction Climate Forecast System version 2 (CFSv2; Saha et al., 2014) and ECMWF seasonal forecasting system 5 (SEAS5; Johnson et al., 2019) to drive the proposed CNN-based statistical-dynamical hybrid predictions for seasonal TC activity. We note that, neither NCEP CFSv2 nor ECMWF SEAS5 was included in the CNN training processes. Therefore, the actual hybrid prediction skill depends both on the fidelity of trained statistical CNN framework and the accuracy of the dynamical CFSv2 or SEAS5 predictions.

Both NCEP CFSv2 and ECMWF SEAS5 are atmosphere–ocean–land fully coupled models that aimed for operational seasonal forecasting. However, because of the coarse resolutions, CFSv2 and SEAS5 are not able to explicitly resolve TCs. There are 16 CFSv2 runs per day; four out to ~9 months, three out to 3 seasons, and nine out to 45 days. Focusing on seasonal TC prediction scope, we only implement those ~9 months long forecasts. CFSv2 hindcast simulations were run from December 1981 to March 2011, with initialization at 00, 06, 12 and 18 UTC cycles for every fifth day starting 00 UTC 1 January of every year (i.e., 292 forecasts for every year during 1982–2010). Operational real-time forecasts have been available since 1 April 2011, and have been carried out with 4 cycles every day. For more details about the CFSv2 system design, we refer the readers to Saha et al. (2014). To obtain the dynamical model forecasted CNN predictors for Northern Hemisphere TC seasons (June–November) and Southern Hemisphere TC seasons (December–May), the earliest CFSv2 operational predictions should be initialized at 00 UTC 27 January and 00 UTC 27 July, respectively. Here, we defined hybrid forecasts from January (July), February (August), ..., May (November) CFSv2 initial conditions as the lead month (LM) 5, 4, ..., 1 forecasts for the predictions of Northern (Southern) Hemisphere TC activity in the subsequent TC seasons. To minimize the CFSv2 prediction uncertainties, 20 ensemble members (initial conditions from 4 cycles of the last 5 calendar days of January/July) were taken to derive the CNN predictors for the LM5 predictions, and 40 ensemble members (initial conditions from 4 cycles of the first 10 calendar days) were used to obtain the predictors for the LM4–LM1 CNN hybrid predictions. For the LM0 hybrid prediction, which we defined as the forecasts initialized in each June (December) 1st, we deployed the dynamical prediction system from ECMWF SEAS5. Different from CFSv2, SEAS5 was initialized on the first day of each month and ran for 7 months. During the hindcast period of 1981–2016, 25 ensemble forecast members were constructed, and it increased to 51 ensemble members for the operational forecast period from 2017 up to present.

To derive inputs for our CNN model from CFSv2 and SEAS5 forecasts, we first interpolated the monthly mean prognostic variables (i.e., temperature, humidity, and wind velocities) from each of ensemble member to the uniform  $2^\circ \times 2^\circ$  resolution grid, and then calculated the anomalies based on the 1993–2016 hindcast climatology. These monthly CFSv2 and SEAS5 anomalies were further added onto ERA5 1993–2016 climatology to reconstruct the “bias-corrected” predictions, in a similar manner as in Bruyère et al. (2014). We then computed CNN predictor variables based on the bias-corrected monthly mean CFSv2 and SEAS5 predictions and conducted 6-month TC season mean and ensemble average to obtain the CNN inputs. We utilized the CFSv2 and SEAS5 ensemble mean large-scale environmental factors as the inputs to each of 600-member ensemble CNNs for statistical-dynamical hybrid predictions of seasonal TC activity. As such, the actual hybrid prediction skills depend both on the fidelity of trained statistical CNN framework and the accuracy of the dynamical CFSv2/SEAS5 predictions.

**Table 3**

*Pearson Correlation Coefficients Between Ensemble Mean CNN Based Statistical-Dynamical Hybrid Predictions and Observed TC Activity With Various Forecast Lead Month (LM)*

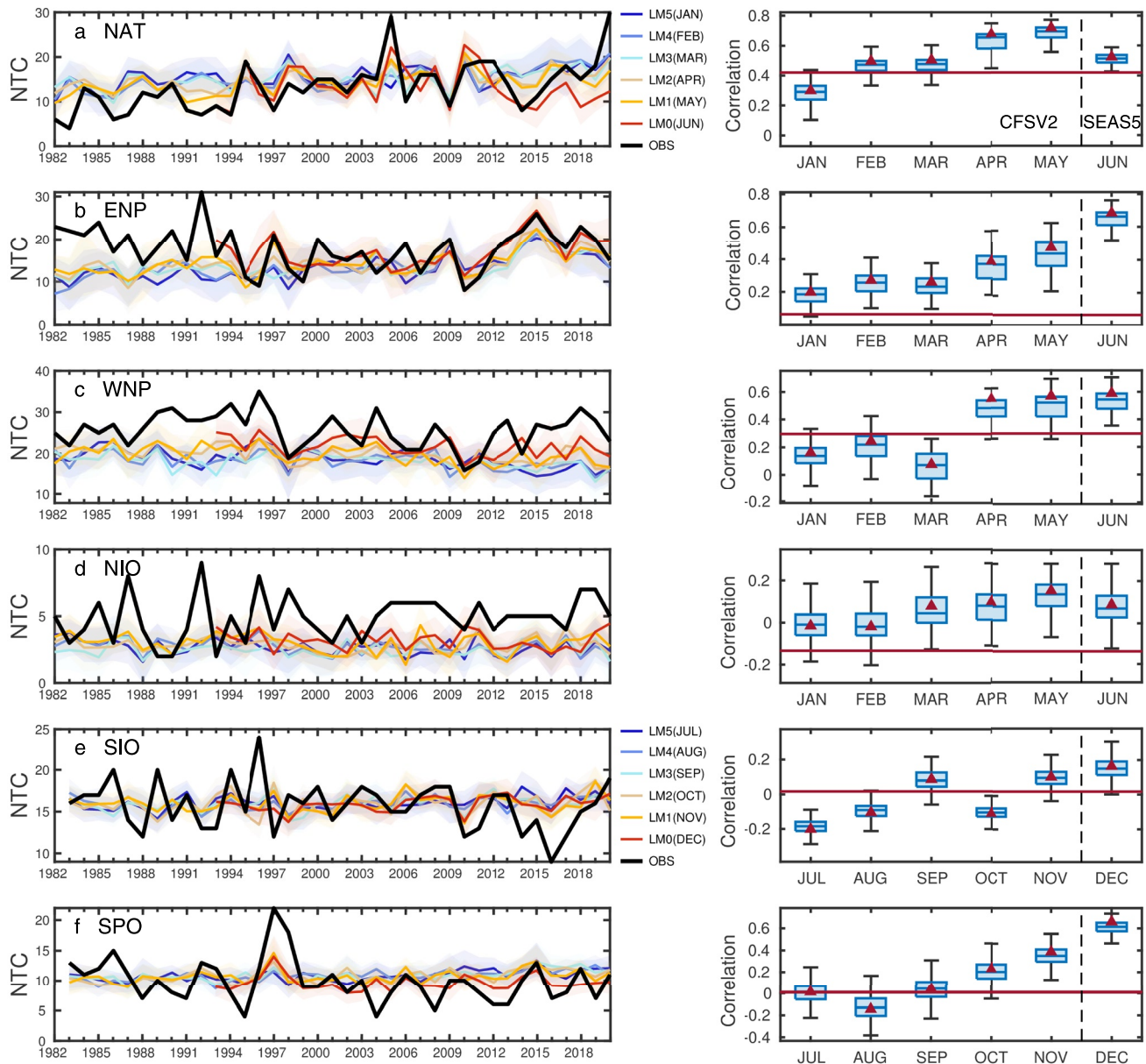
Predictand	LM0	LM1	LM2	LM3	LM4	LM5
NAT NTC	<b>0.52</b> (0.20)	<b>0.72</b> (0.47)	<b>0.68</b> (0.40)	<b>0.50</b> (0.25)	<b>0.50</b> (0.25)	0.30(0.06)
NAT NHU	<b>0.68</b> (0.39)	<b>0.69</b> (0.47)	<b>0.59</b> (0.34)	<b>0.39</b> (0.11)	<b>0.45</b> (0.17)	0.26(−0.12)
NAT ACE	<b>0.65</b> (0.40)	<b>0.67</b> (0.45)	<b>0.55</b> (0.31)	<b>0.39</b> (0.15)	<b>0.47</b> (0.22)	0.30(0.05)
ENP NTC	<b>0.68</b> (0.45)	<b>0.48</b> (0.22)	<b>0.39</b> (0.11)	0.26(−0.04)	0.27(−0.04)	0.20(−0.12)
ENP NHU	<b>0.75</b> (0.54)	<b>0.31</b> (0.05)	0.25(−0.01)	0.22(−0.05)	0.22(−0.05)	0.17(−0.12)
ENP ACE	<b>0.85</b> (0.59)	<b>0.60</b> (0.31)	<b>0.59</b> (0.30)	<b>0.42</b> (0.18)	<b>0.48</b> (0.21)	<b>0.31</b> (0.10)
WNP NTC	<b>0.59</b> (0.34)	<b>0.57</b> (0.32)	<b>0.55</b> (0.30)	0.08(−0.14)	−0.02(−0.01)	−0.01(−0.14)
WNP NHU	<b>0.76</b> (0.56)	<b>0.62</b> (0.38)	<b>0.51</b> (0.25)	0.25(−0.01)	<b>0.31</b> (0.02)	0.29(−0.02)
WNP ACE	<b>0.85</b> (0.66)	<b>0.79</b> (0.59)	<b>0.61</b> (0.38)	<b>0.47</b> (0.22)	<b>0.43</b> (0.18)	<b>0.33</b> (0.09)
NIO NTC	0.09(−0.17)	0.15(−0.06)	0.10(−0.05)	0.08(−0.05)	−0.02(−0.12)	−0.01(−0.14)
NIO NHU	<b>0.48</b> (0.23)	<b>0.37</b> (0.13)	0.28(0.07)	0.15(0.01)	<b>0.36</b> (0.12)	<b>0.34</b> (0.12)
NIO ACE	0.34(0.12)	<b>0.56</b> (0.28)	<b>0.50</b> (0.24)	<b>0.48</b> (0.22)	<b>0.35</b> (0.12)	0.29(0.08)
SIO NTC	0.16(0.01)	0.10(−0.06)	−0.11(−0.30)	0.09(−0.05)	−0.11(−0.13)	−0.20(−0.22)
SIO NHU	0.15(−0.03)	−0.09(−0.14)	−0.11(−0.13)	−0.16(−0.15)	−0.21(−0.20)	−0.10(−0.16)
SIO ACE	−0.07(−0.13)	−0.22(−0.23)	−0.16(−0.18)	−0.20(−0.18)	−0.30(−0.23)	−0.23(−0.25)
SPO NTC	<b>0.66</b> (0.33)	<b>0.39</b> (0.15)	0.23(0.05)	0.04(−0.05)	−0.14(−0.14)	0.01(−0.06)
SPO NHU	<b>0.59</b> (0.31)	0.23(0.04)	0.09(−0.04)	−0.02(−0.06)	−0.19(−0.14)	0.21(0.04)
SPO ACE	<b>0.54</b> (0.28)	0.18(−0.03)	0.06(−0.13)	−0.04(−0.19)	−0.02(−0.22)	0.10(−0.09)

*Note.* Italic bold face highlights statistical significance of the correlations at the 95% level based on two-tailed Student's *t*-test. The mean-square skill scores (MSSSs; Li et al., 2013; Murakami et al., 2016) are listed in the parenthesis just after correlation coefficients. The MSSS is a metric for the skill comparison of the model and climatology-based forecasts, with higher values indicating a good prediction model. We define forecasts from June (December), May (November), . . . , January (July) initial conditions as the LM 0, 1, . . . , 5 forecasts for the predictions of TC activity in the subsequent season in the Northern (Southern) Hemisphere. We use NCEP CFSv2 (1982–2020) and ECMWF SEAS5 (1993–2020) for LM1–5 and LM0 hybrid predictions, respectively.

Table 3 summarizes prediction skill scores in different TC basins. Figures 6–8 illustrate prediction uncertainties of NTC, NHU and ACE, respectively. In general, our CNN-based hybrid prediction framework yields promising seasonal prediction skills in NAT, ENP, WNP, and SPO. For NAT, the hybrid prediction outperforms persistence predictions and provides skillful forecasts with the correlation coefficient *p*-values smaller than 0.05 since February initialization. Correlation generally increases with the decrease of prediction lead month with the exception for the June initialization, which may be attributable to the different dynamical model skills in the NAT between CFSv2 and SEAS5. Forecasts with May initialization exhibit the highest skill scores, with correlation of 0.72, 0.69, and 0.67 for NTC, NHU, and ACE, respectively (Table 3). For reference, the statistical forecasts issued by Colorado State University (CSU) reported prediction correlation coefficients of 0.71 (0.31), 0.41 (0.00), and 0.41 (0.02) for the NAT NTC, NHU, and ACE from their June (April) initialized forecasts during 1982–2021 (1995–2021), respectively. Using a hybrid technique with inputs from the Geophysical Fluid Dynamical Laboratory (GFDL) Forecast-Oriented Low Ocean Resolution (FLOR) model, Murakami et al. (2016) reported a correlation coefficient of 0.75 for NAT NTC when their hybrid model was initialized in June. However, their targeted forecast season is July–November over the 1980–2014 period that is slightly different from our verification period of June–November over the 1982–2020. Nevertheless, these comparisons show that our CNN-based hybrid predictions are capable of competing with or even slightly outperforming the existing seasonal TC prediction works in the NAT sector, particularly at LM1–4.

Similar robust prediction skills are also found in the ENP and WNP. In contrast to the NAT, forecasts with June initialization exhibit highest skills in terms of correlation and MSSS scores. In particular, the correlation coefficient has values of 0.85 (0.75), and 0.85 (0.76) for ACE (NHU), respectively, in the ENP and WNP (Table 3). To the best of our knowledge, these high skill scores have not been previously reported in the published literature.



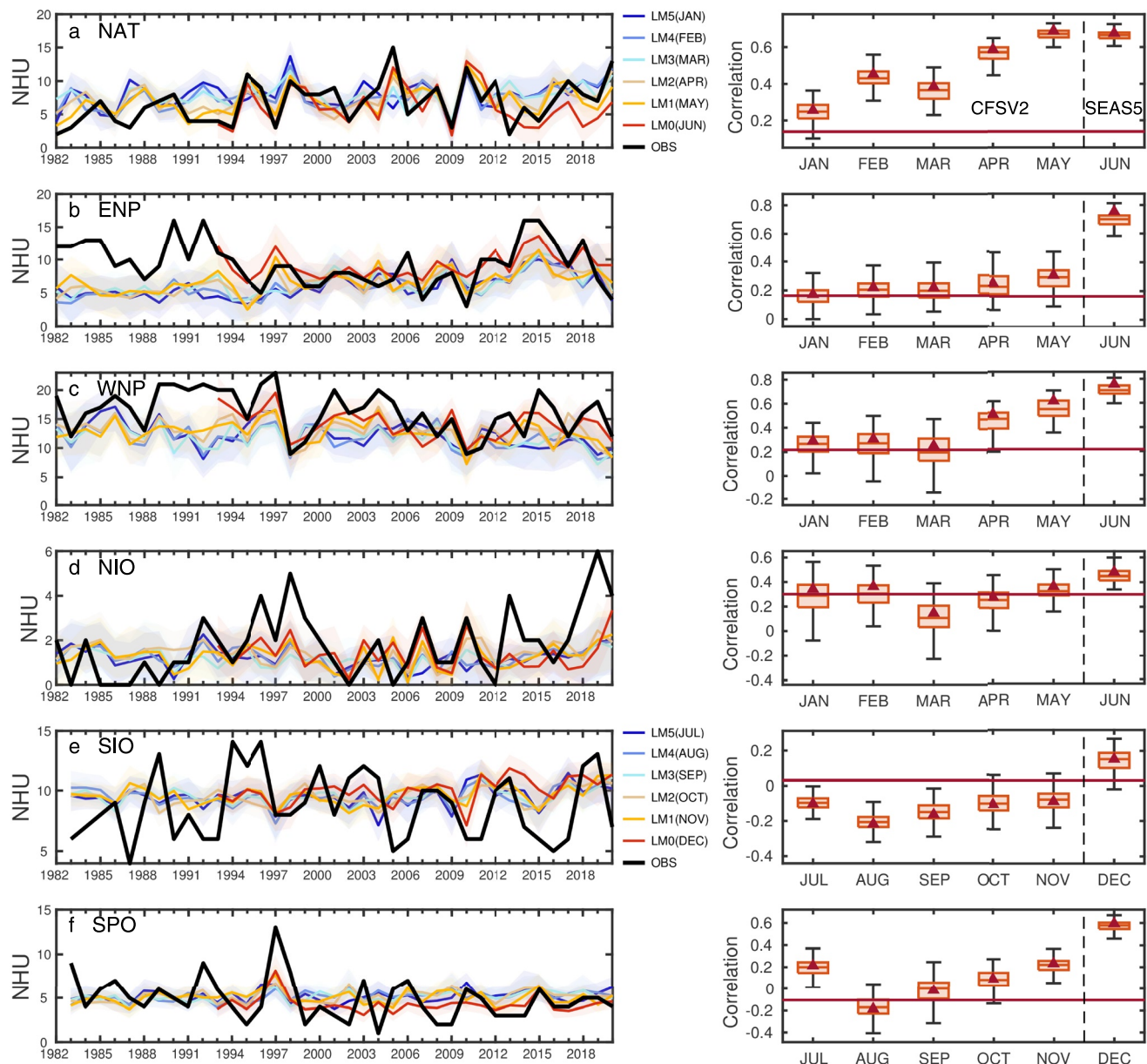


**Figure 6.** CNN based statistical-dynamical hybrid NTC seasonal predictions initialized at different lead month (LM). (a) Left: time-series of NAT NTC in observation (black) from 1982 to 2020, and CNN model predictions with the large-scale environmental conditions predicted from NCEP CFSv2 (LM1-5) and ECMWF SEAS5 (LM0) seasonal forecasts (refer to the figure legend). Note that, we downloaded SEAS5 from Copernicus Climate Change Service (C3S) Climate Data Store (CDS), which only provide seasonal prediction data since 1993. Color shading denotes the range across the CNN ensemble predictions. Right: boxplots of Pearson correlation coefficients between observation and CNN ensemble predictions at different lead months. Boxes show the first quartile, median and third quartile among the 600-member ensemble hybrid prediction, and the dashed vertical lines show the lowest and highest datum still within the 1.5 interquartile ranges. Maroon triangles denote the ensemble mean skill, and maroon horizontal line denotes persistence prediction skills. Persistence prediction is based on trailing 5-year average (WMO, 2008). (b–f) Are similar to (a), but for NTC seasonal predictions in the ENP, WNP, NIO, SIO, and SPO, respectively. Note that, we define SIO and SPO TC season as the period from prior November to following May. We did not show SAT results, given that seasonal mean SAT NTC is less than 0.1 in the observation.

In fact, skillful June–November ACE predictions are obtained at LM5 with January initialization (Figures 8b and 8c), and skillful NTC and NHU predictions are also obtained as early as April initiation (Figures 6b, 6c, 7b, and 7c). These high skills further demonstrate the potential of the CNN-based hybrid prediction model in predicting seasonal TC activity in the North Pacific.

However, skillful seasonal predictions in SPO can only be obtained just ahead of the Southern Hemisphere TC season of December–May (Table 3), while the hybrid framework performs even worse in SIO. For NIO, although



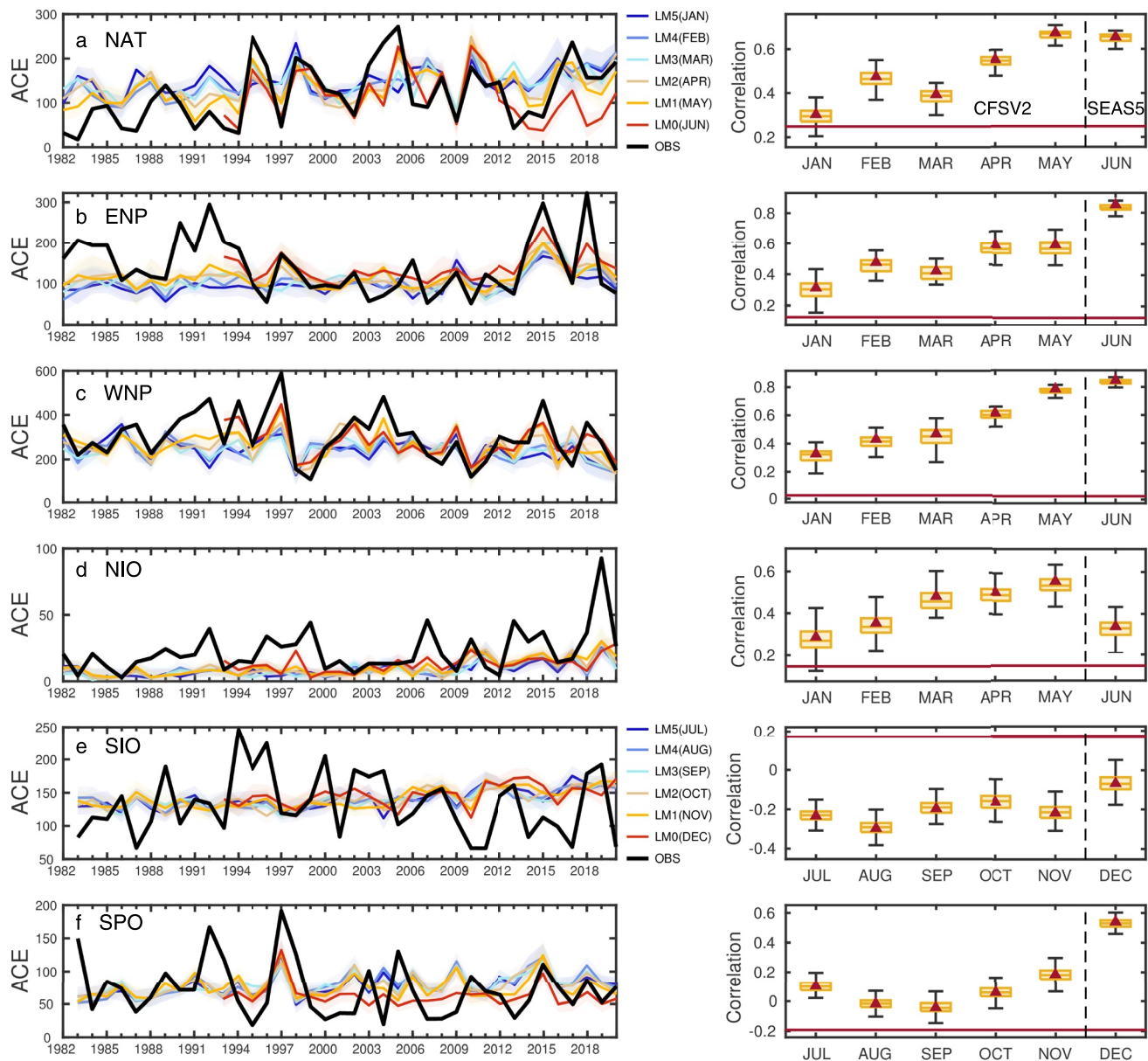


**Figure 7.** Similar to Figure 6, but for the NHU hybrid seasonal prediction skills.

the hybrid framework fails to produce skillful NTC prediction, it shows skillful NHU prediction with June initial-ization and ACE prediction at LM1-4. As noted earlier, the skillful hybrid seasonal TC predictions presented here are determined jointly by the large-scale dynamical seasonal prediction and the statistically CNN model. In future studies, we plan to implement multi-model dynamical seasonal forecasts from Copernicus Climate Change Service (C3S) Climate Data Store (CDS), which includes UK Met Office, Météo France, and JMA model forecasts, to our CNN-based framework in order to see if improved prediction skills can be obtained using the multi-model ensembles.

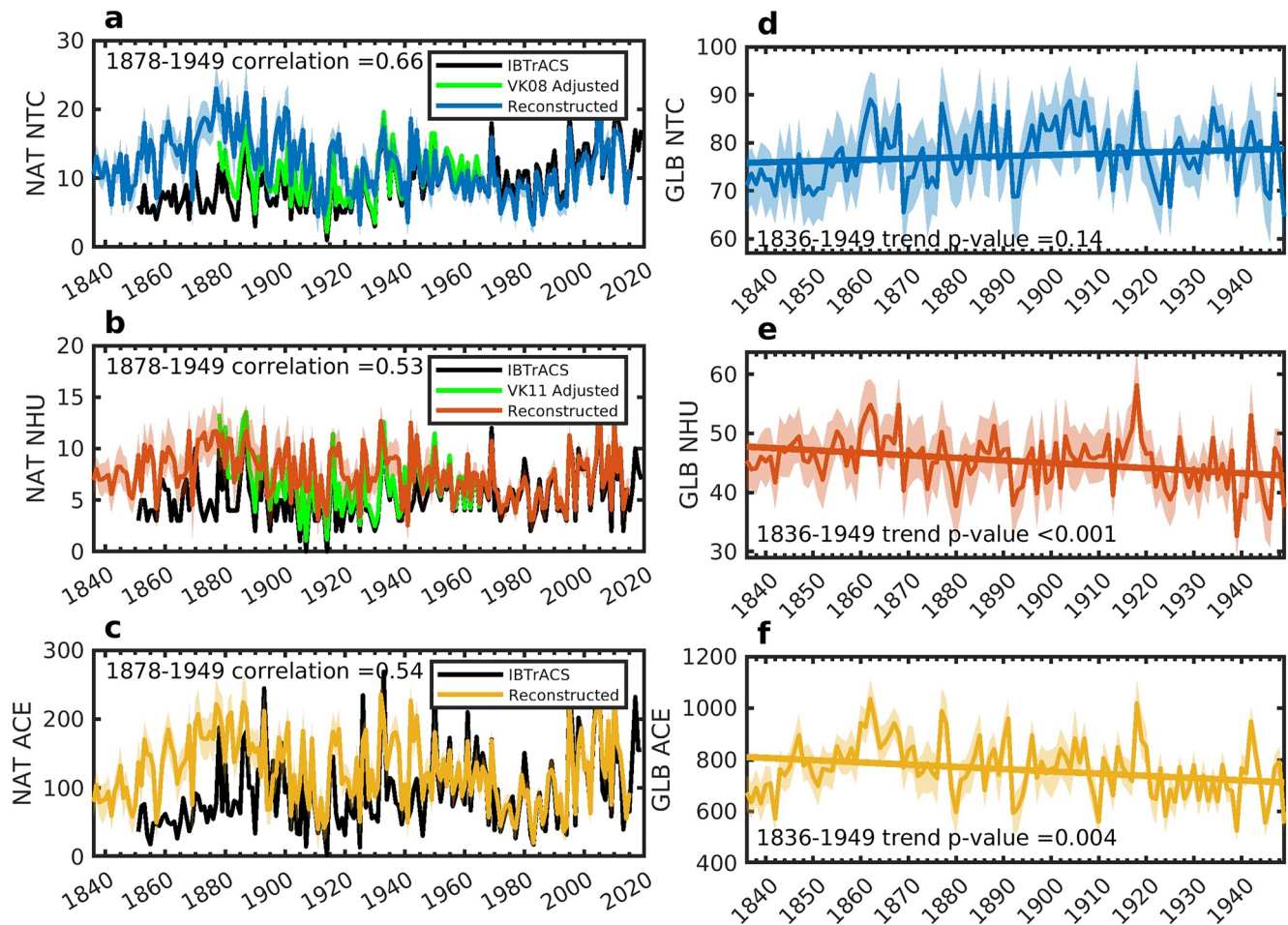
#### 4. Reconstruction of Historical TC Records Using CNN and 20th Century Reanalysis

Our CNN framework can also be applied to the reconstruction of historical TC activity using the 20th century reanalysis product. Although the IBTrACS NAT TC record can be traced back to 1851, before routine airborne reconnaissance in the 1940s and geostationary satellite imagery in 1960s, many previous studies suggested that



**Figure 8.** Similar to Figure 6, but for the ACE hybrid seasonal prediction skills.

the TC climatology is underestimated before 1970s (Emanuel, 2010; Landsea, 2007; Vecchi et al., 2021; Vecchi & Knutson, 2008, 2011), and numerous efforts have been made in attempt to correct for the missing TCs. More importantly, TC observations outside the NAT were generally sparse before the satellite era, which impede the robust detection of climate change trend. An innovative approach to extend the historical TC record beyond the satellite era is to leverage the 20th century historical reanalysis to either directly identify TC-like structures (Chand et al., 2022; Truchelut et al., 2013) from the background flow or use monthly mean large-scale environmental factors for dynamical downscaling (Emanuel, 2010, 2013, 2021). However, these two methods do not show consensus on the detection of TC trends, especially in the NAT basin (Emanuel, 2021; Vecchi et al., 2021). Here, we attempt to utilize the CNN framework trained ensemble CNN framework to the historical reanalysis and offer a complementary insight of historical seasonal TC activities purely from the perspective of large-scale TC favorable environments.



**Figure 9.** Time-series of NAT seasonal mean (a) NTC, (b) NHU, and (c) ACE from the IBTrACS observation in black, adjusted NTC and NHU observation based on Vecchi and Knutson (2008, 2011; VK08 and VK11 for short) in green, and CNN reconstruction based on the NOAA-CIRES-DOE 20th Century Reanalysis (20CRv3) for them period 1836–2015. Pearson correlation coefficient between linear-trend removed adjusted observations and CNN reconstructions during 1878–1965 are listed in each panel. Note that, data before 1950 are not used in the CNN training. (d–f) Are similar, but for global integration results. Linear trend during 1836–1949 and the p-values determined by the Mann-Kendall trend test are listed. We do not show global IBTrACS observation results as other TC active basins outside NAT only provide data since 1949.

We applied our trained ensemble CNN framework to two historical reanalyses to derive long-term historical records of TCs globally: the NOAA 20th Century Reanalysis Version 3 (1836–2015; Silvisinski et al., 2019; hereafter, 20CRv3) and the Coupled ECMWF Reanalysis of the 20th Century (1901–2010; Laloyaux et al., 2018; hereafter, CERA-20C). Using an 80-member ensemble Kalman filter, the 20CRv3 reanalysis assimilates only surface pressure observations with the use of observed monthly mean SST and sea-ice concentration as boundary conditions to produce a historical record of atmospheric circulations. CERA-20C is a 10-member ensemble coupled reanalysis of the 20th century, which assimilates surface pressure and marine wind observations as well as ocean temperature and salinity profiles. Both reanalysis products have been extensively investigated to explore the variability and trends of the past century extreme weather events and climate (Chand et al., 2022; Emanuel, 2021). We applied similar post-processing techniques to 20CRv3 and CERA-20C to obtain the CNN predictor variables by first interpolating the monthly mean thermal and dynamical variables to a  $2^\circ \times 2^\circ$  resolution grid, then computing the ensemble average and performing the diagnostic calculations, and finally deriving the 6-month seasonal average as CNN inputs.

Figure 9 shows the reconstructed TC records back to 1836 using 20CRv3. In the NAT, Vecchi and Knutson (2008, 2011) and Vecchi et al. (2021) documented the statistical adjustments for the missing tropical cyclones before the satellite era based on the estimations of available ship locations between 1878 and 1965. By combining these statistical adjustments with original IBTrACS observations, we can qualitatively evaluate the



fidelity of our historical reconstructions. Recall that our CNN transfer learning is based on the 1950–2019 ERA5 data. This means that the period before 1950 can be used as an independent segment for the cross verification for our CNN model because it is excluded in the CNN training processes. Although CNN reconstruction suggested a general decreasing trend during 1878–1950 (Figure 9a), which is contrast to the increasing trend in IBTrACS and no trend in Vecchi and Knutson (2008) but is consistent with Chand et al. (2022), the interannual variability can be captured remarkably well. After removing the linear trend, correlations of NTC, NHU and ACE between the CNN reconstruction and adjusted observations are 0.66, 0.53, and 0.54, respectively, all of which are significant at 99% confidence level. Although the ensemble spread is increased compared to the training period of 1950–2019, the spread before 1949 still remains relatively small.

It is worth noting that, although our reconstruction approach is very different from those by Vecchi and Knutson (2008, 2011), Truchelut et al. (2013), and Chand et al. (2022), all the results consistently show a peak of seasonal TC activity around 1880–1890 (Figures 9a and 9b; Figure 8a of Truchelut et al., 2013 and Figure 2g of Chand et al., 2022). Furthermore, these independent results clearly suggest that the recent increasing trend of NAT TC activity since 1970s is attributed to the multidecadal variability, probably linked to various internal modes of climate variability and reduced aerosol forcing after the 1970s (Vecchi et al., 2021) instead of the century-scale secular trend (Figures 9a–9c). Unfortunately, TC observations outside the Atlantic were insufficient for validation in other TC basins. However, if we assume the results of the CNN historical reconstructions also hold the other basins, it would indicate that global NTC experienced an insignificant upward trend during 1836–1949, but NHU and ACE experienced a significant decrease (Figures 9d–9f).

In Figure S5 in Supporting Information S1, we show the results of the historical reconstruction using CERA-20C. Cross-validation with the adjusted observations also yields significantly high correlation coefficients and relatively small ensemble spreads. In the context of global TC activity trend, CERA-20C indicated an insignificant downward trend of NTC and NHU during 1901–1949, but a significant downward trend of ACE. Nonetheless, our ensemble CNN framework provides an ancillary and independent method to reconstruct past historical TC records. Although our results do not support a significant decreasing trend of global TC activity as shown by the recent study of Chand et al. (2022) that was based on an older version (v2c) of NOAA 20CR, it does support the findings that the recent increase in NAT TC activity may not be a part of long-term climate trend, but more likely be associated with internal climate variability (Vecchi & Knutson, 2008, 2011; Vecchi et al., 2021).

## 5. Explore Future TC Activity Using CNN and CMIP6 Projection

As a final demonstration of our CNN applications to TC studies in this study, we apply the trained CNN framework to explore future changes in TCs using Coupled Model Intercomparison Project (CMIP6) projection.

Although a broad consensus has emerged that TC intensity and rain rate are likely to increase in response to future warming (Knutson et al., 2020; Patricola & Wehner, 2018), projection uncertainties are poorly constrained in terms of future TC frequency changes. For example, a vast majority of climate model future projections with high-horizontal-resolution ( $\leq 50$  km) indicates a moderate-to-weak decreasing trend in future TC frequency (Knutson et al., 2010, 2020; Walsh et al., 2016; Zhao et al., 2009), but a few studies suggest an increase trend in future TC frequency (Vecchi et al., 2019). Using a statistical-deterministic method developed by Emanuel et al. (2008), Emanuel. (2013) shows an increasing trend of future TC frequency by downscaling CMIP5 models, while using a Poisson regression model developed by Tippett et al. (2011), Camargo et al. (2014), and Lee et al. (2020) show two diverging TC frequency climate trends projections, in which the TC frequency will increase or decrease depending on the selection of the atmosphere moisture variable CRH or SD. The use of SD leads to a decrease of TC activity, while using CRH results in an increase. The SD-based results are consistent with the dynamically simulated TC reduction shown by Zhao et al. (2009), but the CRH-based results are not. Given that our CNN is trained devotedly to improve the representation of TC temporal variability, it is intriguing to see if such a CNN trained on present-day climate data, is applicable to future TC activities. In particular, we want to test if the sign of projected TC activity trends is also sensitive to the choice of atmospheric moisture predictor variables. Therefore, we repeated the CNN training processes as described in Section 2 but replacing SD with CRH, so that we obtained two ensemble CNN frameworks. In the rest of this paper, we refer CNN\_SD as for the original trained CNN using SD as moisture predictor variable, and CNN\_CRH for the newly established framework using CRH. We further note that the transient climate simulations of 1980–2100 conducted by 3-member ensemble of high-resolution CESM1.3 were excluded from training and validation processes, so that we can use them as



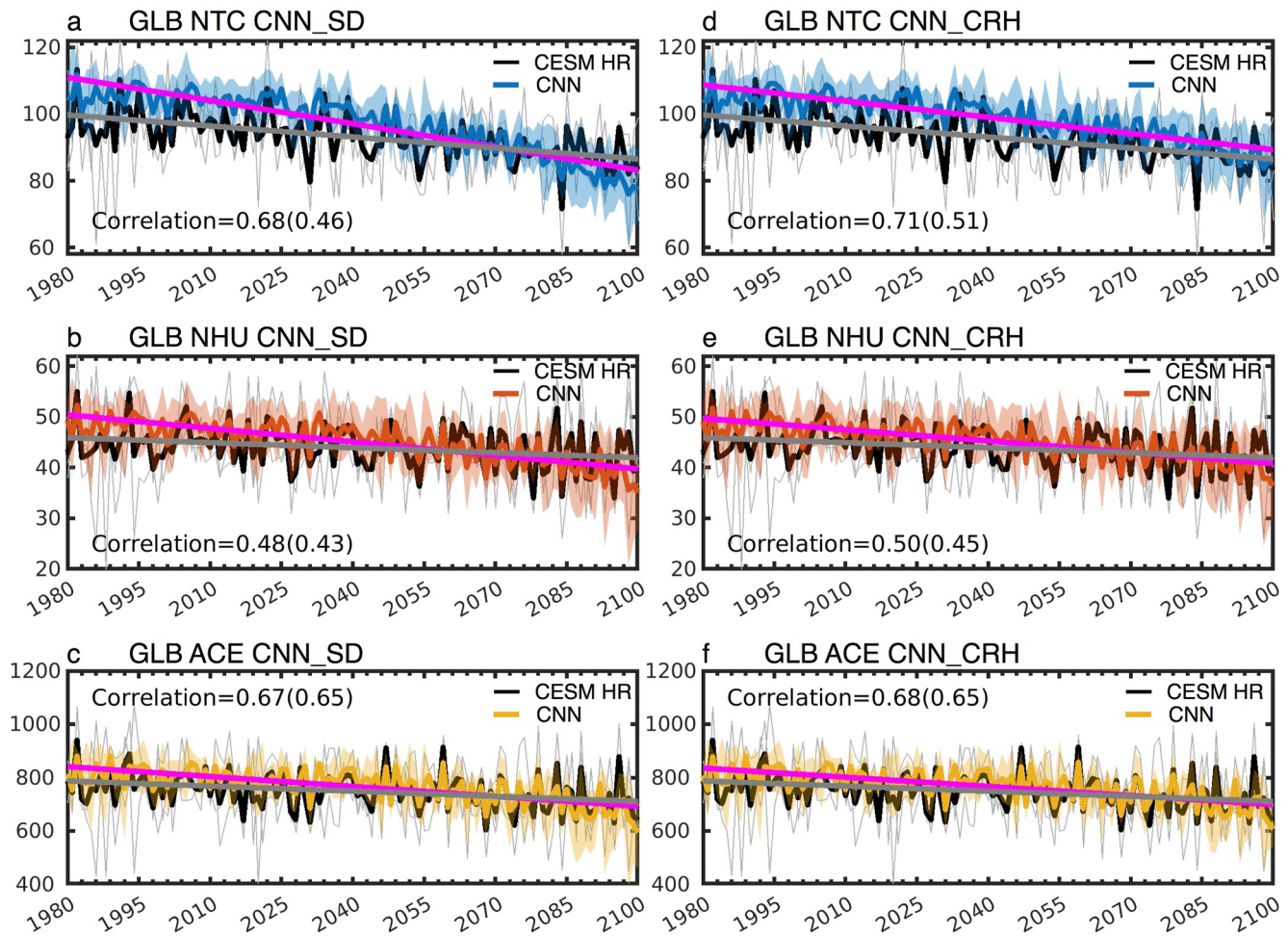
independent cross validation data set and compare CNN emulated TC activities to those directly simulated by high-resolution CESM1.3 from 1980 to 2100.

Supplementary Figure 6 shows a comparison between CNN emulated and high-resolution CESM1.3 direct simulated global NTC, NHU, and ACE for the period 1980–2100. Huge discrepancies can be noticed: although CNN\_SD and CNN\_CRH broadly show same sign of change, the CNN emulated NTC increase with time while directly simulated NTC shows a decreasing trend. Moreover, the ensemble spread of CNN emulations is quite large compared to those in seasonal prediction and historical reconstruction, implying large uncertainties in CNN projection of future TC activities. These results cast some doubts in our CNNs' application to the study of future TC activity changes.

However, it is noteworthy to point out that transfer-learning technique was employed to both CNN\_SD and CNN\_CR after training-from-scratch on high-resolution climate simulations, so that CNN\_SD and CNN\_CRH are both constrained by the recent observations. We note that the empirical relationship between TCs and large-scale environments in observations may not be fully captured by climate model simulations, despite the high fidelity of our high-resolution climate simulations in reproducing observed TC climatology, interannual variability, and associated large-scale ambient environments. These discrepancies can be arisen from various sources, including climate model physics, dynamical core, and resolution (Camargo et al., 2020). While the transfer-learning technique we employed can potentially correct these discrepancies by fine-tuning the CNNs using ERA5 and IBTrACS observations, it is important to acknowledge that transfer-learning technique can also potentially degrades the fidelity of climate model-based TC-environment relationship. This side effect due to transfer-learning may confound future TC projections that solely rely on climate model simulations, given that it is impossible to directly observe the future.

In order to test the impact of the transfer-learning on CNN projections of future TC changes, we made another sensitivity test that discard transfer-learning technique in the training process, and use CNNs trained only on the climate model simulations with historical forcing. Again, 1980–2100 transient climate simulations were excluded in the training. Figure 10 illustrates similar comparisons between the non-transfer-learning CNN\_SD and CNN\_CRH and directly simulated NTC, NHU, and ACE for the period 1980–2100. Now both CNN\_SD and CNN\_CRH display decreasing trends of future TC activities, which are consistent with the directly simulated changes by high-resolution CESM1.3 with statistically significant correlation coefficients. Furthermore, the ensemble spreads of CNN emulations are significantly reduced compared to those from the observation-constrained transfer-learned CNNs. It is interesting to note that, although both CNN\_SD and CNN\_CRH are trained on the historical forcing that contain much weaker warming signals than those in the mid and late 21st century, CNNs are still able to capture long-term changes in TC activities under future warming in high-resolution CESM1.3 future climate simulations. However, some inconsistencies still exist. For instance, CNN emulated decreasing trends of NTC and NHU are too strong compared to those in CESM1.3. Nonetheless, these promising results again cross validate the fidelity our CNN framework and provide positive insights that properly trained CNNs only using present-day climate simulations may still be applicable to investigate the climate model based TC future projections.

Besides the cross validation against the high-resolution CESM simulations, we also applied the non-transfer-learning CNN\_SD and CNN\_CRH to 36 different CMIP6 models participated in the Scenario Model Intercomparison Project (ScenarioMIP) (see Table S2 in Supporting Information S1 for a list). Figure 11 shows the global integrated, as well as Northern and Southern Hemisphere averaged CNN emulated NTC using inputs from the 36 CMIP6 models. CMIP6 historical climate simulations during the period of 1950–2015 and future climate simulations during the period of 2015–2100 under shared socio-economic pathway 5–8.5 (SSP585) scenario were used to derive the large-scale environmental inputs for CNN. CNN emulated NHU and ACE are illustrated in Figures 12 and 13, respectively. Figure S7 in Supporting Information S1 further compares the CNN emulations in each individual TC basin. Consistent with many previous studies of dynamical model projections (Knutson et al., 2015, 2020; Sugi et al., 2017; Walsh et al., 2016; Yoshida et al., 2017), our CNN framework shows a decrease of seasonal TC activity in the future warmer climate, using either SD or CRH as moisture predictor variable. In fact, using CNN\_SD (CNN\_CRH), 35 (36) of total 36 CMIP6 models project a decreasing trend of global mean NTC, which is primarily caused by the significant reduction of NTC in the Southern Hemisphere (Figure 11). Examining changes in each TC basin indicates that both CNN\_SD and CNN\_CRH reveals a robust NTC decrease in the SIO (Figure S7 in Supporting Information S1), consistent with previous CMIP5 and



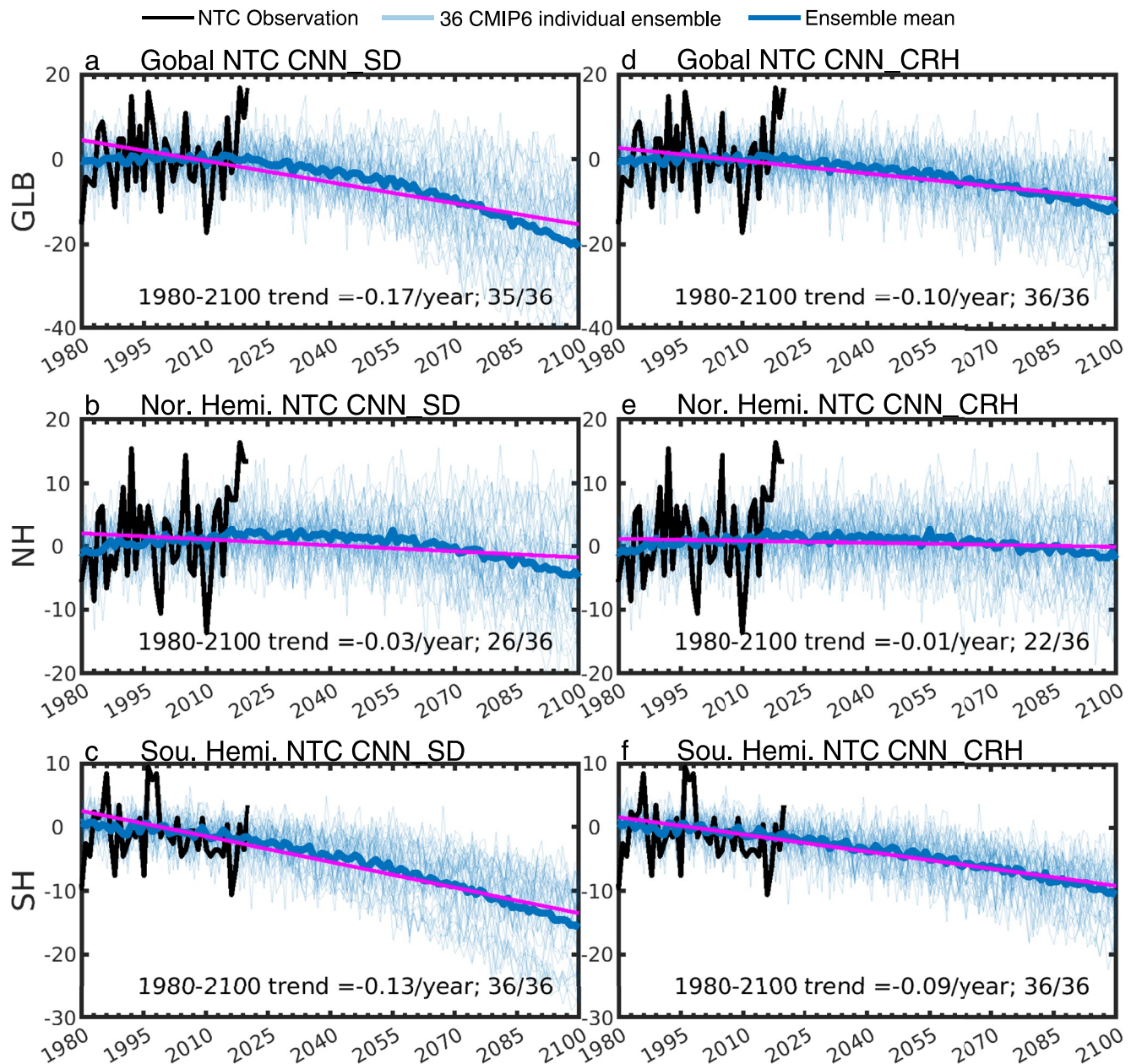
**Figure 10.** Time-series of 3-member ensemble mean CESM HR dynamically resolved TC activity (black) and CNN\_SD (without transfer-learning, trained on climate model only) emulated global (a) NTC (blue), (b) NHU (red) and (c) ACE (yellow). Thin gray lines denote CESM HR three individual ensemble member, and color shadings indicate ranges across CNN ensembles. Pearson correlation coefficients between 1980 and 2100 are listed in each panel. The linear-trend-removed correlation coefficients are listed in the parentheses. All listed correlation coefficients are significant at 95% confidence level. Note that, CESM HR data after 1980 are not used in the CNN training. (d–f) Are similar, but using the alternative trained CNN model with column-integral relative humidity (CNN\_CRH), rather than saturation deficit (CNN\_SD), as the CNN moisture predictor variables.

HighResMIP studies (Bell et al., 2019; Roberts et al., 2020; Tory et al., 2013). NTC also shows a decline in the WNP and SPO, but it shows an increase in the NIO. No significant change is projected in the ENP, but the NAT sees an increase of NTC in the first half of the period up to 2050s and then a decrease in the second half. The cause for this non-monotonic change in NAT NTC deserves future studies.

Weak decreasing trends are projected by our CNN framework for the global integrated NHU and ACE (Figures 12 and 13). A majority of CMIP6 models projects slight increases in Northern Hemisphere NHU and ACE, which compensate for the substantial decreasing trends in the Southern Hemisphere. It is interesting to note that, although our CNN framework is purely empirically based and the NTC training is independent of the NHU or ACE training, it can still capture the increasing trend of intense TC fraction especially in the NAT (Figure 14), which is consistent with the recent high-resolution climate model simulations and observational evidences (Knutson et al., 2020, 2022; Kossin et al., 2020).

## 6. Conclusions and Discussions

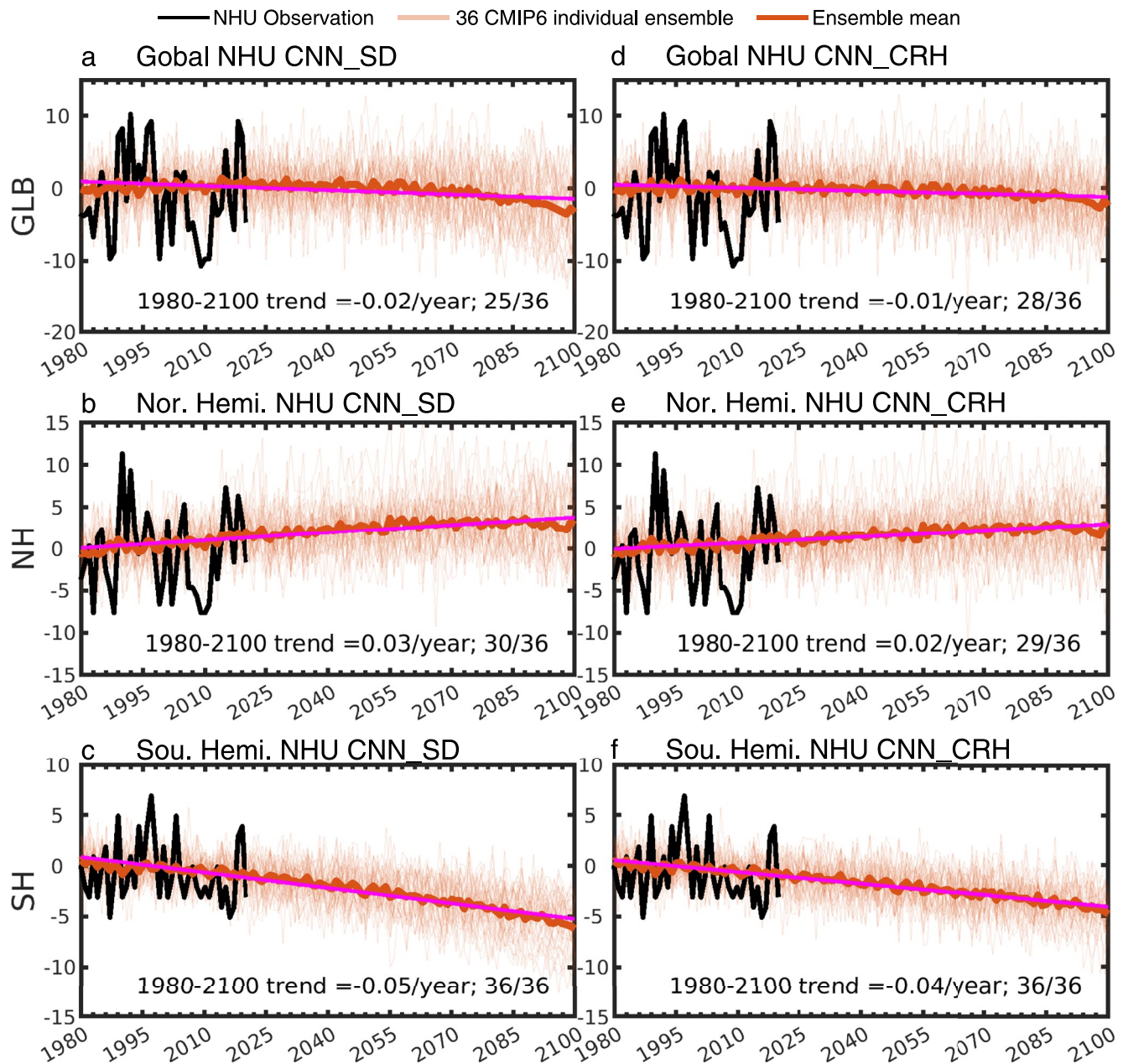
This study extends the previous works on TC genesis index by using a machine learning technique to improve building the empirical linkage between large-scale environmental conditions and TC variability. By leveraging



**Figure 11.** Time-series of CNN\_SD emulated anomalous NTC for (a) global, (b) Northern Hemisphere and (c) Southern Hemisphere integration with the large-scale environmental conditions projected by 36 different Coupled Model Intercomparison Project (CMIP6) model under historical forcing and shared socio-economic pathway 5–8.5 (SSP585). Anomalies are computed as the departures from their 1980–1999 climatology. Black lines denote observation during 1980–2020, thick dark blue lines denote the multi-model mean of the CMIP6 models, and thin light blue lines denote the individual 36 CMIP6 model. Linear trends during 1980–2100 are plotted in magenta and listed in each panel; following by the fraction demonstrating number of individual models showing consistent sign of trend as to the multi-model mean. For example, 35 of 36 CMIP6 model project a decreasing trend with the mean trend of  $-0.17$  per year emulated by the CNN\_SD model. Note that, all multi-model mean linear trends are significant at 95% confidence level based on the Mann-Kendall trend test. (e–f) Are similar, but for the CNN\_CRH emulated NTC projection.

large suites of high-resolution climate model simulations and ERA5 reanalysis product, we developed a framework with ensemble of CNNs that achieve high fidelity in capturing observed TC interannual-to-multidecadal variability. Using occlusion sensitivity analysis, we confirm that physically plausible mechanisms underlying TC environmental favorability are properly implemented in the trained CNN framework. We further demonstrate that this CNN framework is applicable to many aspects of TC-climate studies. A number of preliminary applications are highlighted.



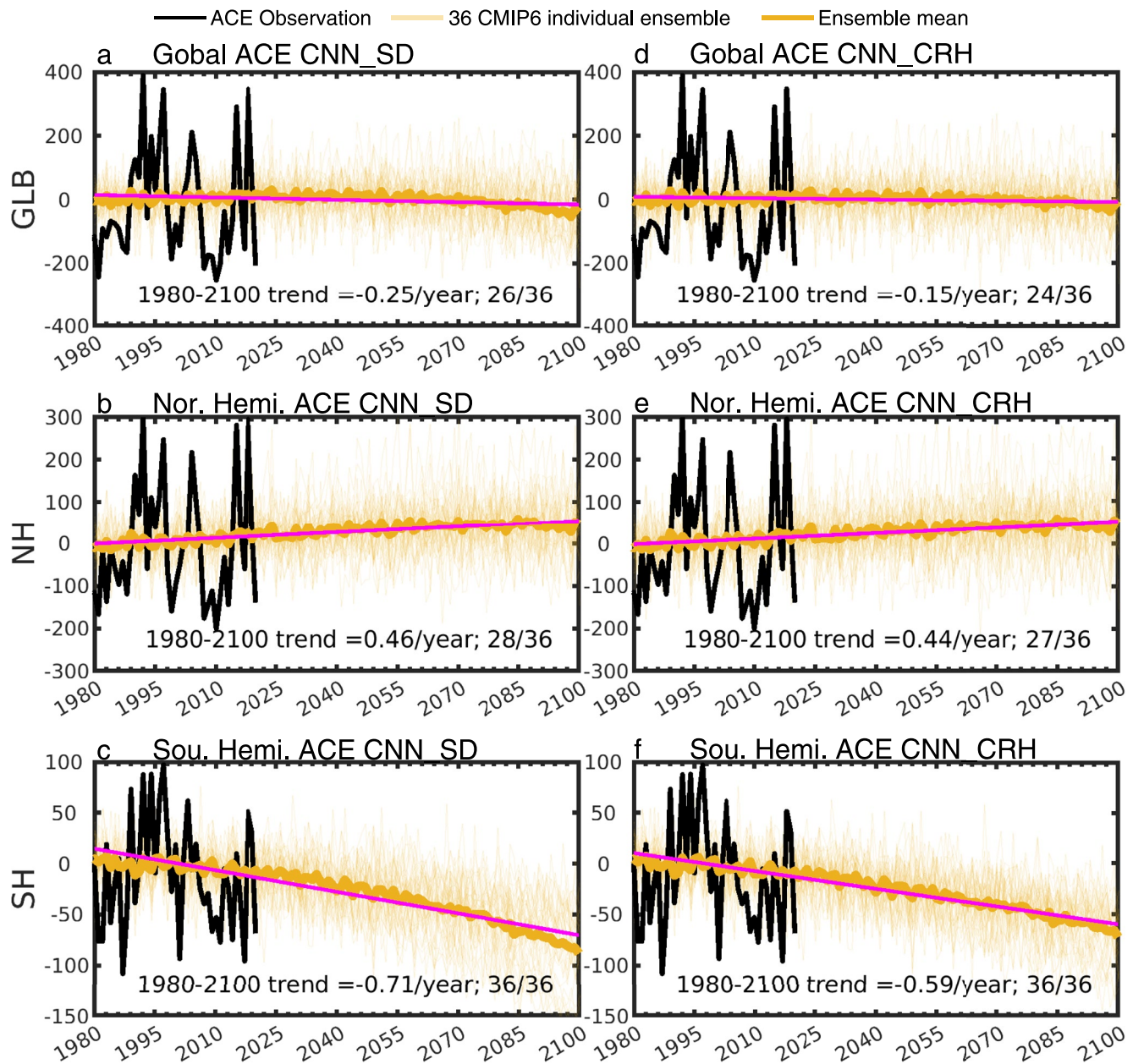


**Figure 12.** Similar to Figure 11, but for the NHU projections.

First, using large-scale CNN predictors forecasted by the operational NCEP CFSv2 and ECMWF SEAS5, we show that the CNN based statistical-dynamical hybrid model can make skillful seasonal TC predictions in NAT (WNP and ENP) from early February (late January). Furthermore, ENP and WNP ACE show high correlation skill values ( $\sim 0.85$ ) when the hybrid prediction is initialized in June, just at the beginning of TC season. In general, our CNN-based hybrid predictions can compete with or even outperform existing seasonal TC predictions particularly at long lead times.

Second, by applying the ensemble CNNs to the NOAA 20CRv3 and ECMWF CERA20C reanalysis products, we found that the recent increase in NAT TC activity since 1970s may not be a part of long-term climate trend, but more likely be dominated by internal climate variability, which agrees with the results of previous studies using entirely different methods (Chand et al., 2022; Vecchi & Knutson, 2008, 2011; Vecchi et al., 2021).

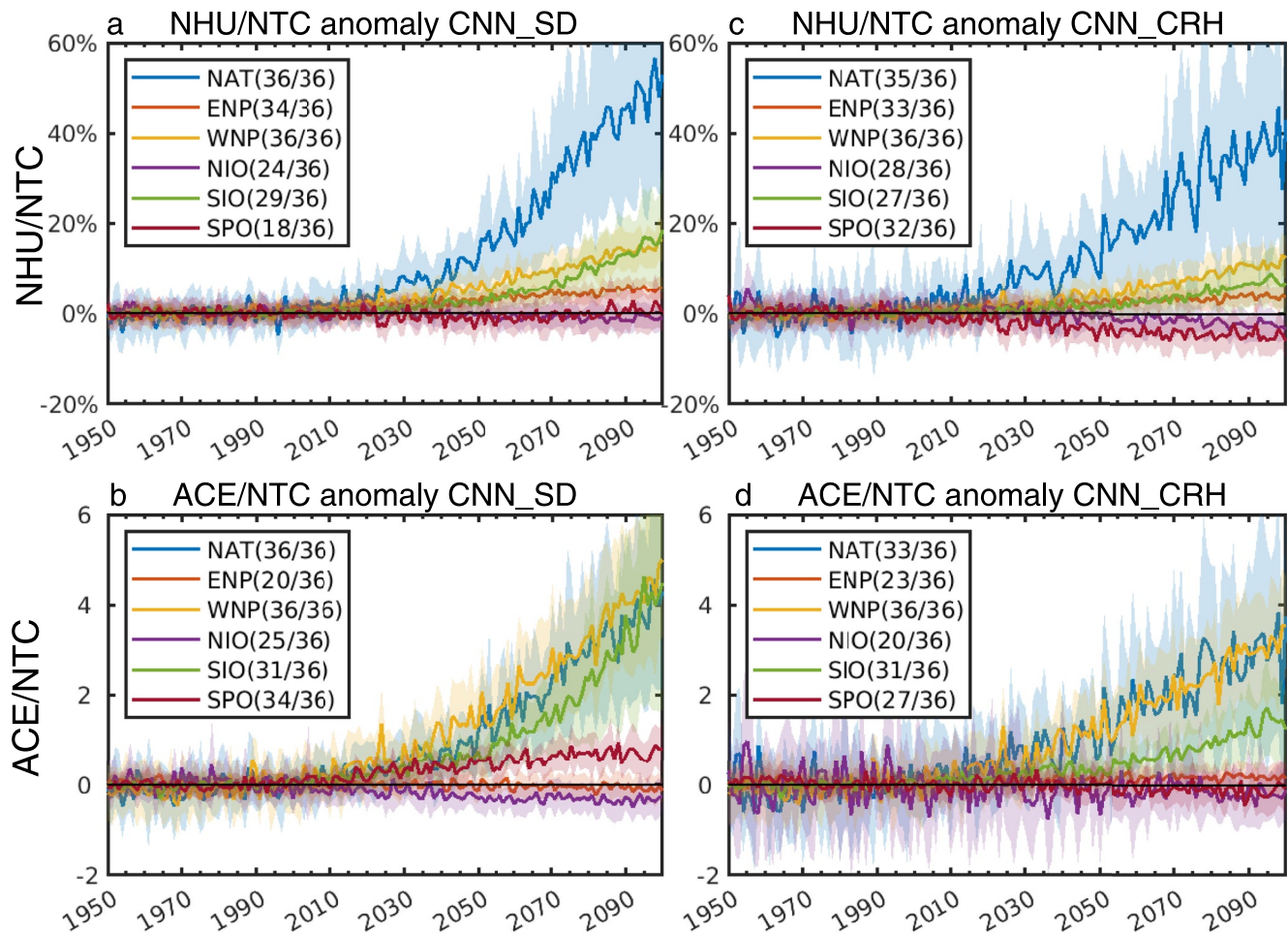




**Figure 13.** Similar to Figure 11, but for the ACE projections.

Third, by validating ensemble CNNs with high-resolution CESM1.3 climate future projection simulations, we found that observation-constrained transfer-learning technique confounds the fidelity of the future TC projection, while non-transfer-learning CNNs that trained exclusively on climate model simulations forced by the historical climate condition show more consistent decreasing trend of future TC activity with the climate model projected TC trend. In addition, such a CNN-based projection is independent of the choice of moisture predictor variables. We further applied these non-transfer-learning CNNs to 36 different CMIP6 models, and multi-model ensemble results show a robust decreasing trend of global TC activity and increasing trend of intense TC fraction in the future warmer climate, which are consistent with the recent assessment of future TC changes (Knutson et al., 2020).

Apart from these promising applications, an especially captivating characteristic of CNNs is their remarkable computational efficiency. Once CNN predictors are properly computed, the computational cost of CNN prediction is negligible. Our CNN approach can be viewed as an extension to previous TC genesis index studies. We



**Figure 14.** Time-series of CNN\_SD emulated anomalous ratio of (a) NHU/NTC, and (b) ACE/NTC in different ocean basins, with the large-scale environmental conditions projected by 36 different Coupled Model Intercomparison Project (CMIP6) model under historical forcing and shared socio-economic pathway 5–8.5 (SSP585). Anomalies are computed as the departures from their 1980–1999 climatology. Standard deviation among 36 CMIP6 models are shown as shadings, while multi-model mean is plotted in lines. The number of individual models with consistent sign of trend as to the multi-model mean is shown in the legend. (c, d) Are similar, but for the CNN\_CRH emulated CMIP6 projection.

suggest that machine learning approaches, especially deep learning models, may provide a new avenue to improve our understanding of TC variability and future changes.

It is important to acknowledge that our CNN framework was trained from scratch using high-resolution CESM1.3 and WRF TCM simulations. Consequently, CNNs without transfer learning at best can perform similarly to CESM1.3 and WRF TCM simulations, and will also inherit their intrinsic biases. In the future, it would be beneficial to train CNNs on multiple climate model large-ensemble simulations that are capable of directly simulating realistic seasonal TC activities, such as GFDL Seamless System for Prediction and Earth System Research (SPEAR, Delworth et al., 2020) or MRI-AGCM3.2 “Database for Policy Decision-Making for Future Climate Change” (d4PDF; Mizuta et al., 2017). This approach would enable us to assess the sensitivity of CNN emulations and CNN-based TC-climate studies across different climate model training sets.

Furthermore, it is worth noting that for the sake of simplicity, we conventionally defined the Northern Hemisphere TC season as June–November and the Southern Hemisphere TC season as December–May in this study. However, this choice may not be optimal for specific basins. For example, the NIO TC activity exhibits peaks during the pre-monsoon (April–June) and post-monsoon (October–December) seasons. Using the June–November average may introduce biases in the CNN training process.

Additionally, an inherent limitation of our framework is that we used the same four environmental factors (SST, SD/CRH, low-level vorticity, and vertical wind shear) as CNN predictors, which may not be the most optimal

combination for emulating TC activities in each TC-active basin. In future studies, we plan to delve deeper into investigating and identifying the optimal large-scale climate variables and TC seasons that can effectively emulate seasonal TC activity, thereby refining our framework and improving its performance.

## Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

## Data Availability Statement

ECMWF ERA5 is publicly available at (Hersbach et al., 2017), SEAS5 is available at (Copernicus Climate Change Service, 2018), and CERA-20C is available at <https://apps.ecmwf.int/datasets>. NCEP CFSv2 is available at <https://www.ncei.noaa.gov/products/weather-climate-models/climate-forecast-system>, and NOAA 20CrV3 is available at [https://psl.noaa.gov/data/gridded/data.20thC\\_ReanV3.html](https://psl.noaa.gov/data/gridded/data.20thC_ReanV3.html). CMIP6 model outputs are available at <https://esgf-node.lnl.gov/search/cmip6/>. IBTrACS is available at (Knapp et al., 2018). High-resolution CESM1.3 simulations are archived at <https://ihesp.github.io/archive>. CSU seasonal TC prediction verification data are available at <https://tropical.colostate.edu/archive.html#verification>. As part of this paper, we are also releasing the trained ensemble CNNs for seasonal TC activity at <https://doi.org/10.5281/zenodo.8299866> (Fu et al., 2023) to allow future studies. The ensemble CNN framework was trained on the Grace Cluster at Texas A&M University. Each of the 21 CNN models required approximately 12 hr of wall-clock time to complete training, including both train-from-scratch and transfer-learning, with the assistance of the NVIDIA A100 40 GB GPU.

## Acknowledgments

Most of the high-resolution CESM1.3 simulation data sets used in this study was generated by the International Laboratory for High-Resolution Earth System Prediction (IHESP), which are publicly available. This research is supported by the National Science Foundation (NSF) under award AGS-2231237 and NOAA's Climate Program Office under award NA20OAR4310408. X.L. was supported by the U.S. Department of Energy (DOE) under Award Number DE-SC0020072. We thank the Texas Advanced Computing Center (TACC) for providing computing resources to conduct high-resolution CESM1.3 and WRF TCM simulations.

## References

- Bell, G. D., Halpert, M. S., Schnell, R. C., Higgins, R. W., Lawrimore, J., Kousky, V. E., et al. (2000). Climate assessment for 1999. *Bulletin of the American Meteorological Society*, 81(6), S1–S50. [https://doi.org/10.1175/1520-0477\(2000\)81\[s1:CAFJ2.0.CO;2](https://doi.org/10.1175/1520-0477(2000)81[s1:CAFJ2.0.CO;2)
- Bell, S. S., Chand, S. S., Tory, K. J., Dowdy, A. J., Turville, C., & Ye, H. (2019). Projections of southern hemisphere tropical cyclone track density using CMIP5 models. *Climate Dynamics*, 52(9–10), 6065–6079. <https://doi.org/10.1007/s00382-018-4497-4>
- Bister, M., & Emanuel, K. A. (1998). Dissipative heating and hurricane intensity. *Meteorology and Atmospheric Physics*, 52(3–4), 233–240. <https://doi.org/10.1007/bf01030791>
- Bolton, T., & Zanna, L. (2019). 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>
- Bretherton, C. S., Peters, M. E., & Back, L. E. (2004). Relationships between water vapor path and precipitation over the tropical oceans. *Journal of Climate*, 17(7), 1517–1528. [https://doi.org/10.1175/1520-0442\(2004\)017<1517:rbwvpa>2.0.co;2](https://doi.org/10.1175/1520-0442(2004)017<1517:rbwvpa>2.0.co;2)
- Brüyère, C. L., Done, J. M., Holland, G. J., & Fredrick, S. (2014). Bias corrections of global models for regional climate simulations of high-impact weather. *Climate Dynamics*, 43(7), 1847–1856. <https://doi.org/10.1007/s00382-013-2011-6>
- Brüyère, C. L., Holland, G. J., & Towler, E. (2012). Investigating the use of a genesis potential index for tropical cyclones in the North Atlantic basin. *Journal of Climate*, 25(24), 8611–8626. <https://doi.org/10.1175/jcli-d-11-00619.1>
- Camargo, S. J., Giulivi, C. F., Sobel, A. H., Wing, A. A., Kim, D., Moon, Y., et al. (2020). Characteristics of model tropical cyclone climatology and the large-scale environment. *Journal of Climate*, 33(11), 4463–4487. <https://doi.org/10.1175/jcli-d-19-0500.1>
- Camargo, S. J., Tippet, M. K., Sobel, A. H., Vecchi, G. A., & Zhao, M. (2014). Testing the performance of tropical cyclone genesis indices in future climates using the HiRAM model. *Journal of Climate*, 27(24), 9171–9196. <https://doi.org/10.1175/JCLI-D-13-00505.1>
- Chand, S. S., Walsh, K. J. E., Camargo, S. J., Kossin, J. P., Tory, K. J., Wehner, M. F., et al. (2022). Declining tropical cyclone frequency under global warming. *Nature Climate Change*, 12(7), 655–661. <https://doi.org/10.1038/s41558-022-01388-4>
- Chang, P., Zhang, S., Danabasoglu, G., Yeager, S. G., Fu, H., Wang, H., et al. (2020). An unprecedented set of high-resolution earth system simulations for understanding multiscale interactions in climate variability and change. *Journal of Advances in Modeling Earth Systems*, 12(12), e2020MS002298. <https://doi.org/10.1029/2020ms002298>
- Chapman, W. E., Subramanian, A. C., Delle Monache, L., Xie, S. P., & Ralph, F. M. (2019). Improving atmospheric river forecasts with machine learning. *Geophysical Research Letters*, 46(17–18), 10627–10635. <https://doi.org/10.1029/2019GL083662>
- Chattopadhyay, A., Hassanzadeh, P., & Pasha, S. (2020). Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. *Scientific Reports*, 10(1), 1317. <https://doi.org/10.1038/s41598-020-57897-9>
- Chattopadhyay, A., Nabizadeh, E., & Hassanzadeh, P. (2020). Analog forecasting of extreme-causing weather patterns using deep learning. *Journal of Advances in Modeling Earth Systems*, 12(2), e2019MS001958. <https://doi.org/10.1029/2019MS001958>
- Copernicus Climate Change Service, Climate Data Store. (2018). Seasonal forecast monthly statistics on pressure levels. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. <https://doi.org/10.24381/cds.0b79e7c5>
- Davenport, F. V., & Diffenbaugh, N. S. (2021). Using machine learning to analyze physical causes of climate change: A case study of U.S. Midwest extreme precipitation. *Geophysical Research Letters*, 48(15), e2021GL093787. <https://doi.org/10.1029/2021GL093787>
- Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J., Dunne, K. A., et al. (2020). SPEAR: The next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *Journal of Advances in Modeling Earth Systems*, 12(3), e2019MS001895. <https://doi.org/10.1029/2019ms001895>
- Elsner, J. B., & Jagger, T. H. (2013). Frequency models. In *Hurricane climatology: A modern statistical guide using R* (pp. 161–193). Oxford University Press.
- Emanuel, K. (2010). Tropical cyclone activity downscaled from NOAA- CIRES reanalysis, 1908–1958. *Journal of Advances in Modeling Earth Systems*, 2, 1–12. <https://doi.org/10.3894/james.2010.2.1>



- Emanuel, K. (2013). Downscaling CMIP5 climate models shows increased tropical cyclone activity over the 21st century. *Proceedings of the National Academy of Sciences of the United States of America*, 110(30), 12219–12224. <https://doi.org/10.1073/pnas.1301293110>
- Emanuel, K. (2021). Atlantic tropical cyclones downscaled from climate reanalyses show increasing activity over past 150 years. *Nature Communications*, 12(1), 655–661. <https://doi.org/10.1038/s41467-021-27364-8>
- Emanuel, K., Sundararajan, R., & Williams, J. (2008). Hurricanes and global warming: Results from downscaling IPCC AR4 simulations. *Bulletin America Meteorology Social*, 89(3), 347–368. <https://doi.org/10.1175/BAMS-89-3-347>
- Emanuel, K. A., & Nolan, D. S. (2004). Tropical cyclone activity and global climate. In *Proceedings of 26th conference on hurricanes and tropical meteorology* (pp. 240–241). American Meteorological Society.
- Fu, D., Chang, P., & Liu, X. (2023). Data for “using convolutional neural network to emulate seasonal tropical cyclone activity” [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.8299866>
- Fu, D., Chang, P., Patricola, C. M., & Saravanan, R. (2019). High-resolution tropical channel model simulations of tropical cyclone climatology and intraseasonal-to-interannual variability. *Journal of Climate*, 32(22), 7871–7895. <https://doi.org/10.1175/jcli-d-19-0130.1>
- Fu, D., Chang, P., Patricola, C. M., Saravanan, R., Liu, X., & Beck, H. E. (2021). Central American mountains inhibit eastern North Pacific seasonal tropical cyclone activity. *Nature Communications*, 12(1), 4422. <https://doi.org/10.1038/s41467-021-24657-w>
- Gray, W. M. (1968). Global view of the origin of tropical disturbances and storms. *Monthly Weather Review*, 96(10), 669–700. [https://doi.org/10.1175/1520-0493\(1968\)096<0669:gvotoo>2.0.co;2](https://doi.org/10.1175/1520-0493(1968)096<0669:gvotoo>2.0.co;2)
- Gray, W. M. (1979). Hurricanes: Their formation, structure and likely role in the tropical circulation. *Meteorology over the tropical oceans*. Royal Meteorological Society, 77, 155–218.
- Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, 573(7775), 568–572. <https://doi.org/10.1038/s41586-019-1559-7>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2017). Complete ERA5 from 1940: Fifth generation of ECMWF atmospheric reanalyses of the global climate. *Copernicus Climate Change Service (C3S) Data Store (CDS)*. <https://doi.org/10.24381/cds.143582cf>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al. (2019). SEAS5: The new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12(3), 1087–1117. <https://doi.org/10.5194/gmd-12-1087-2019>
- Kim, H. S., Ho, C. H., Kim, J. H., & Chu, P. S. (2012). Track-pattern-based model for seasonal prediction of tropical cyclone activity in the western North Pacific. *Journal of Climate*, 25(13), 4660–4678. <https://doi.org/10.1175/JCLI-D-11-00236.1>
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017). Self-normalizing neural networks. *Advances in Neural Information Processing Systems*, 30.
- Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, M. C., & Schreck, C. J. (2018). *International best track archive for climate stewardship (IBTrACS) project, version 4*. NOAA National Centers for Environmental Information. <https://doi.org/10.25921/82ty-9e16>
- Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, 91(3), 363–376. <https://doi.org/10.1175/2009BAMS2755.1>
- Knutson, T. R., Camargo, S. J., Chan, J. C. L., Emanuel, K., Ho, C. H., Kossin, J., et al. (2020). Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bulletin America Meteorology Social*, 101(3), E303–E322. <https://doi.org/10.1175/BAMS-D-18-0194.1>
- Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K., Holland, G., Landsea, C., et al. (2010). Tropical cyclones and climate change. *Nature Geoscience*, 3, 157–163. <https://doi.org/10.1038/ngeo779>
- Knutson, T. R., Sirutis, J. J., Bender, M. A., Tuleya, R. E., & Schenkel, B. A. (2022). Dynamical downscaling projections of late twenty-first-century US landfalling hurricane activity. *Climatic Change*, 171(3–4), 28. <https://doi.org/10.1007/s10584-022-03346-7>
- Knutson, T. R., Sirutis, J. J., Zhao, M., Tuleya, R. E., Bender, M., Vecchi, G. A., et al. (2015). Global projections of in- tense tropical cyclone activity for the late twenty-first century from dynamical downscaling of CMIP5/RCP4.5 scenarios. *Journal of Climate*, 28(18), 7203–7224. <https://doi.org/10.1175/JCLI-D-15-0129.1>
- Kossin, J. P., Knapp, K. R., Olander, T. L., & Velden, C. S. (2020). Global increase in major tropical cyclone exceedance probability over the past four decades. *Proceedings of the National Academy of Sciences of the United States of America*, 117(22), 11975–11980. <https://doi.org/10.1073/pnas.1920849117>
- Laloux, P., de Boisseson, E., Balmaseda, M., Bidlot, J., Broennimann, S., Buizza, R., et al. (2018). CERA-20C: A coupled reanalysis of the twentieth century. *Journal of Advances in Modeling Earth Systems*, 10(5), 1172–1195. <https://doi.org/10.1029/2018ms001273>
- Landsea, C. W. (2007). Counting Atlantic tropical cyclones back to 1900. *Eos, Transactions, American Geophysical Union*, 88(18), 197–202. <https://doi.org/10.1029/2007EO180001>
- Landsea, C. W., & Franklin, J. L. (2013). Atlantic hurricane database uncertainty and presentation of a new database format. *Monthly Weather Review*, 141(10), 3576–3592. <https://doi.org/10.1175/mwr-d-12-00254.1>
- Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K. R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications*, 10(1), 1096. <https://doi.org/10.1038/s41467-019-08987-4>
- Lee, C. Y., Camargo, S. J., Sobel, A. H., & Tippet, M. K. (2020). Statistical-dynamical downscaling projections of tropical cyclone activity in a warming climate: Two diverging genesis scenarios. *Journal of Climate*, 33(11), 4815–4834. <https://doi.org/10.1175/JCLI-D-19-0452.1>
- Li, X., Yang, S., Wang, H., Jia, X., & Kumar, A. (2013). A dynamical-statistical forecast model for the annual frequency of western Pacific tropical cyclones based on the NCEP Climate Forecast System version 2. *Journal of Geophysical Research: Atmospheres*, 118(21), 12–061. <https://doi.org/10.1002/2013jd020708>
- Liu, Y., Racah, E., Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., et al. (2016). Application of deep convolutional neural networks for detecting extreme weather in climate dataset. arXiv preprint arXiv:1605.01156.
- McGuire, M. P., & Moore, T. W. (2022). Prediction of tornado days in the United States with deep convolutional neural networks. *Computers & Geosciences*, 159, 104990. <https://doi.org/10.1016/j.cageo.2021.104990>
- Menkes, C. E., Lengaigne, M., Marchesio, P., Jourdain, N. C., Vincent, E. M., Lefèvre, J., et al. (2012). Comparison of tropical cyclogenesis indices on seasonal to interannual timescales. *Climate Dynamics*, 38(1–2), 301–321. <https://doi.org/10.1007/s00382-011-1126-x>
- Mizuta, R., Murata, A., Ishii, M., Shiogama, H., Hibino, K., Mori, N., et al. (2017). Over 5,000 years of ensemble future climate simulations by 60-km global and 20-km regional atmospheric models. *Bulletin of the American Meteorological Society*, 98(7), 1383–1398. <https://doi.org/10.1175/BAMS-D-16-0099.1>



- Moon, J., Park, S., Rho, S., & Hwang, E. (2019). A comparative analysis of artificial neural network architectures for building energy consumption forecasting. *International Journal of Distributed Sensor Networks*, 15(9), 1550147719877616. <https://doi.org/10.1177/1550147719877616>
- Murakami, H., Levin, E., Delworth, T. L., Gudgel, R., & Hsu, P. C. (2018). Dominant effect of relative tropical Atlantic warming on major hurricane occurrence. *Science*, 362(6416), 794–799. <https://doi.org/10.1126/science.aat6711>
- Murakami, H., Villarini, G., Vecchi, G. A., Zhang, W., & Gudgel, R. (2016). Statistical-dynamical seasonal forecast of North Atlantic and US landfalling tropical cyclones using the high-resolution GFDL FLOR coupled model. *Monthly Weather Review*, 144(6), 2101–2123. <https://doi.org/10.1175/mwr-d-15-0308.1>
- Murakami, H., & Wang, B. (2010). Future change of North Atlantic tropical cyclone tracks: Projection by a 20-km-mesh global atmospheric model. *Journal of Climate*, 23(10), 2699–2721. <https://doi.org/10.1175/2010jcli3338.1>
- Patricola, C. M., Chang, P., & Saravanan, R. (2016). Degree of simulated suppression of Atlantic tropical cyclones modulated by flavour of El Niño. *Nature Geoscience*, 9(2), 155–160. <https://doi.org/10.1038/ngeo2624>
- Patricola, C. M., Saravanan, R., & Chang, P. (2017). A teleconnection between Atlantic sea surface temperature and eastern and central North Pacific tropical cyclones. *Geophysical Research Letters*, 44(2), 1167–1174. <https://doi.org/10.1002/2016GL071965>
- Patricola, C. M., & Wehner, M. F. (2018). Anthropogenic influences on major tropical cyclone events. *Nature*, 563(7731), 339–346. <https://doi.org/10.1038/s41586-018-0673-2>
- Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vannière, B., et al. (2020). Projected future changes in tropical cyclones using the CMIP6 HighResMIP multimodel ensemble. *Geophysical Research Letters*, 47(14), e2020GL088662. <https://doi.org/10.1029/2020gl088662>
- Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, 27(6), 2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>
- Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., et al. (2019). Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, 145(724), 2876–2908. <https://doi.org/10.1002/qj.3598>
- Smith, S. R., Brolley, J., O'Brien, J. J., & Tartaglione, C. A. (2007). ENSO's impact on regional U.S. hurricane activity. *Journal of Climate*, 20(7), 1404–1414. <https://doi.org/10.1175/jcli4063.1>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929–1958.
- Sugi, M., Yoshimura, J., & Yoshida, K. (2017). Projection of future changes in the frequency of intense tropical cyclones. *Climate Dynamics*, 49(1–2), 619–632. <https://doi.org/10.1007/s00382-016-3361-7>
- Tang, B. H., & Neelin, J. D. (2004). ENSO influence on Atlantic hurricanes via tropospheric warming. *Geophysical Research Letters*, 31(24), L24204. <https://doi.org/10.1029/2004GL021072>
- Tippett, M. K., Camargo, S. J., & Sobel, A. H. (2011). A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *Journal of Climate*, 24(9), 2335–2357. <https://doi.org/10.1175/2010jcli3811.1>
- Tory, K. J., Chand, S. S., McBride, J. L., Ye, H., & Dare, R. A. (2013). Projected changes in late-twenty-first-century tropical cyclone frequency in 13 coupled climate models from phase 5 of the Coupled Model Intercomparison Project. *Journal of Climate*, 26(24), 9946–9959. <https://doi.org/10.1175/JCLI-D-13-00010.1>
- Truchelut, R. E., Hart, R. E., & Luthman, B. (2013). Global identification of previously undetected pre-satellite-era tropical cyclone candidates in NOAA/CIRES Twentieth-Century Reanalysis data. *Journal of Applied Meteorology and Climatology*, 52(10), 2243–2259. <https://doi.org/10.1175/jamc-d-12-0276.1>
- Vecchi, G. A., Delworth, T. L., Murakami, H., Underwood, S. D., Wittenberg, A. T., Zeng, F., et al. (2019). Tropical cyclone sensitivities to CO<sub>2</sub> doubling: Roles of atmospheric resolution, synoptic variability and background climate changes. *Climate Dynamics*, 53(9–10), 5999–6033. <https://doi.org/10.1007/s00382-019-04913-y>
- Vecchi, G. A., & Knutson, T. R. (2008). On estimates of historical North Atlantic tropical cyclone activity. *Journal of Climate*, 21(14), 3580–3600. <https://doi.org/10.1175/2008jcli2178.1>
- Vecchi, G. A., & Knutson, T. R. (2011). Estimating annual numbers of Atlantic hurricanes missing from the HURDAT database (1878–1965) using ship track density. *Journal of Climate*, 24(6), 1736–1746. <https://doi.org/10.1175/2010jcli3810.1>
- Vecchi, G. A., Landsea, C., Zhang, W., Villarini, G., & Knutson, T. (2021). Changes in Atlantic major hurricane frequency since the late-19th century. *Nature Communications*, 12(1), 1–9. <https://doi.org/10.1038/s41467-021-24268-5>
- Walsh, K. J., McBride, J. L., Klotzbach, P. J., Balachandran, S., Camargo, S. J., Holland, G., et al. (2016). Tropical cyclones and climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 7(1), 65–89. <https://doi.org/10.1002/wcc.371>
- Wang, B., & Murakami, H. (2020). Dynamic genesis potential index for diagnosing present-day and future global tropical cyclone genesis. *Environmental Research Letters*, 15(11), 114008. <https://doi.org/10.1088/1748-9326/abb01>
- Weyn, J. A., Durran, D. R., & Caruana, R. (2019). Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems*, 11(8), 2680–2693. <https://doi.org/10.1029/2019MS001705>
- Weyn, J. A., Durran, D. R., Caruana, R., & Cresswell-Clay, N. (2021). Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *Journal of Advances in Modeling Earth Systems*, 13(7), e2021MS002502. <https://doi.org/10.1029/2021MS002502>
- WMO. (2008). *Report from expert meeting to evaluate skill of tropical cyclone seasonal forecasts* (Vol. 1455, p. 27). World Meteorological Organization Technical Document.
- Yeager, S. G., Chang, P., Danabasoglu, G., Rosenbloom, N., Zhang, Q., Castruccio, F. S., et al. (2023). Reduced Southern Ocean warming enhances global skill and signal-to-noise in an eddy-resolving decadal prediction system. *Npj Climate and Atmospheric Science*, 6(1), 107. <https://doi.org/10.1038/s41612-023-00434-y>
- Yoshida, K., Sugi, M., Mizuta, R., Murakami, H., & Ishii, M. (2017). Future changes in tropical cyclone activity in high-resolution large-ensemble simulations. *Geophysical Research Letters*, 44(19), 9910–9917. <https://doi.org/10.1002/2017gl075058>
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, 27.
- Yu, J. Y., Hsiao, L. P., & Chiu, P. G. (2018). Evaluating the Emanuel–Nolan genesis potential index: Contrast between North Atlantic and western North Pacific. *Terrestrial, Atmospheric and Oceanic Sciences*, 29(2), 201–214. <https://doi.org/10.3319/TAO.2017.09.27.01>
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *In European conference on computer vision* (pp. 818–833). Springer.
- Zhao, M., Lin, S. J., & Vecchi, G. A. (2009). Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *Journal of Climate*, 22(24), 6653–6678. <https://doi.org/10.1175/2009JCLI3049.1>