Using Convolutional Neural Networks to Emulate Seasonal Tropical Cyclone Activity

Dan Fu[1*], Ping Chang[1,2], Xue Liu[1]

1. Department of Oceanography, Texas A&M University, College Station, TX, USA
2. Department of Atmospheric Sciences, Texas A&M University, College Station, TX, USA

[*]Corresponding author Email: fudan1991@tamu.edu; ORCID: 0000-0001-6423-6117

Key points:

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

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 favorabilities 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 relation between TCs and their ambiences. 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 ensembles of CNNs first on suites of high-resolution climate simulations with realistic seasonal TC activities and large-scale environmental conditions, and then subsequently on the state-of-the-art reanalysis from 1950 to 2019. Our CNNs can remarkably 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 forcing the ensemble CNNs with 20th century reanalysis products and phase 6 of the Coupled Model Intercomparison Project (CMIP6) experiments, we attempted to investigate TC variabilities and their 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 dynamic projections using TC-permitting high-resolution coupled climate model.

Plain Language Summary

Tropical cyclone (TC) requires favorable large-scale environmental conditions to generate. Pioneer studies suggested that these constructive conditions include but not limit to warm sea surface temperatures (SSTs), sufficient low-level vorticities, mid-level humilities, and weak-to-moderate vertical wind shears. Several follow-up studies have devoted to 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 spatial distributions and seasonal cycle reasonably well, their performances in representing the interannual TC variability are significantly decreased. 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 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 evidences have 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 also trailblazed the potential linkages between TC genesis events and large-scale environmental factors. Gray (1979) proposed Yearly Genesis Parameter (YGP) and elucidated the constructive environmental factors for TC genesis 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 progresses have been made since then to improve the quantitative linkage between TC genesis and large-scale environmental favorabilities, and other well-known TC genesis indices have been developed (Emanuel and Nolan 2004, Tippett et al. 2011, Bruyère et al. 2012, Wang and Murakami 2020).

These genesis indices are based on the same principal, which assumes that the favorable large-scale environmental conditions to TC genesis are products of dynamical potentials alone or the combination of dynamical and thermal potentials with different formulations. For example, 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 and Emanuel 1998). Moreover, Emanuel (2010) modified GPI2004 by replacing mid-level relative humidity with the moist entropy deficient and updated the nonlinear multiplication formulation, which has particular importance in relation to issues of climate change. Broadly following Gray (1978), 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 contributions from low-level absolute vorticity to TC genesis are capped off when they exceed certain thresholds. All aforementioned TC genesis indices employed both dynamical and thermal climate variables

85 in their formulations, but the relative roles of thermal versus dynamical potentials in TC
86 genesis remain elusive. More recently, Wang and Murakami (2020) developed dynamical
87 genesis potential index (hereafter, DGPI) for recognition of TC genesis potentials, which is
88 purely based on the large-scale dynamical factors, and showed improved skills in
89 representing TC genesis variations especially under future warmer climate.
90 Although these empirical TC genesis indices were able to replicate the seasonal cycle
91 and spatial distribution of observed TC genesis over the globe, their skills in representing the
92 TC interannual variability in various ocean basins are degraded and highly basin-dependent
93 (Menkes et al. 2012). For instance, GPI2004 can capture TC interannual variability
94 reasonably well in the North Atlantic and eastern North Pacific, but fail in the western North
95 Pacific (Yu et al. 2018). On the other hand, DGPI can skillfully represent the western North
96 Pacific and North Atlantic TC interannual variability, but it is degraded in the eastern North
97 Pacific (Wang and Murakami 2020). Current TC genesis indices generally have very poor
98 skills in reflecting TC interannual variability in the Indian Ocean. These failures and
99 contrasts in capturing TC interannual variability might be attributable to several reasons. First
100 of all, existing TC genesis indices were derived from the climatological mean values, thus the
101 information of interannual variability were intrinsically not incorporated. Nonetheless, these
102 indices can acceptably reproduce the spatial shift pattern of TC activity modulated by El
103 Niño-Southern Oscillation (ENSO) as in the observations, which is a major driver of TC
104 interannual variability. In addition, the derivation of current TC genesis indices did not
105 consider the spatial variations among different TC-active basins, casting doubt on the
106 applicability of their global usage. For instance, TC genesis mechanisms in the western North
107 Pacific differ from those involved in the North Atlantic due to strong convections over the
108 warm pool area. Murakami and Wang (2010) developed a variant version of GPI2004, by
109 incorporating 500hPa vertical velocity as an additional term into the original formulation,
110 which leads to a significant improvement in the western North Pacific. Similarly, in the more
111 recent paper of Wang and Murakami (2020), they noted that the selected dynamic potential
112 factors through stepwise regression method are basin-dependent, and it is worthwhile to

derive variant versions of DGPIs for different ocean basins to optimize the accuracy. As such, using single formulation to represent TC genesis all over the globe would further confound the TC interannual variability.

TC genesis indices are also extensively investigated to understand future changes of TC activities, however, the relevance of the thermodynamic factors of TC genesis indices in explaining TC changes under global warming has 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. 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 in delineating TC year-to-year variations might facilitate to reduce the uncertainties in TC future projections. Along this line, we attempted to incorporate information of the TC interannual variability into the derivation of new TC genesis index, and use deep learning technique, specifically, convolutional neural network (CNN), to overcome the aforementioned limitations in the current TC genesis indices.

In recent years, CNN has been widely utilized in the climate and weather sciences. Research communities use it as a powerful tool to reveal the linkages between three-dimensional predictor fields and the predictand classifications or physical indices (Liu et al. 2016, Chapman et al. 2019, Ham et al. 2019, Bolton and Zanna 2019, Chattopadhyay et al. 2020a, 2020b, Davenport et al. 2021, Weynet al. 2019, 2021, McGuire and Moore 2022). In the context of climate study, pioneering work by Ham et al. (2019) innovatively applied CNNs for multi-year ENSO predictions. In their work, three consecutive months of SST anomalies and ocean heat content anomalies over 0°–360°E, 55°S–60°N were provided to the CNN as the three-dimension predictors, while three-month-averaged Niño3.4 index in the following 23 months were predicted. 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

141 to emulate basin-integrated seasonal TC activity using the concurrent large-scale
142 environmental factors. Take one step further from classic TC genesis index; our CNN models
143 use seasonal averaged thermal and dynamical large-scale environmental factors as predictors,
144 and the basin-integrated seasonal mean TC metrics, such as number of TC (NTC), number of
145 hurricane (NHU; TCs with lifetime maximum intensity ≥65knots), and Accumulated Cyclone
146 Energy (ACE; Bell et al. 2000), as predictands to depict seasonal TC activities. As such, our
147 framework cannot only emulate TC numbers, but also averaged intensity at a given season.
148 We anticipate that our CNN framework can significantly improve the representation of TC
149 interannual-to-multidecadal variabilities. Of particular interest is to explore the applicability
150 of the new CNN frameworks derived from the present-day climate to the seasonal predictions,
151 reconstructions of past historical records, and future climate projections.
152 The structure of the paper is as follows. In Section 2, the CNN-based TC emulation
153 framework, including the architecture, training and interpretation, are discussed. Applications
154 of CNN framework to seasonal TC prediction, 20th century historical reconstructions, and
155 future projections are presented in Section 3, 4, and 5, respectively. The summary and
156 discussion are given in Section 6.
157
158 **2. CNN model development**
159 2.1 Data for CNN training
160 As discussed earlier in Section 1, large-scale environmental favorabilities on seasonal
161 TC activity are geographic-dependent. As such, we trained various CNNs targeted for seven
162 different TC active basins, including North Atlantic (NAT), eastern North Pacific (ENP),
163 western North Pacific (WNP), North Indian Ocean (NIO), South Indian Ocean (SIO), South
164 Pacific Ocean (SPO), and South Atlantic (SAT), and for three different seasonal mean TC
165 metrics (i.e., NTC, NHU and ACE; total of 21 CNN models). Figure 1a illustrates the
166 boundary of each basin. For the Northern (Southern) Hemisphere basins, we only focus on
167 the June-November (December-May), which is conventionally defined as hurricane season.
168 Following pioneering TC genesis index works, we used 4 environmental factors as the

169 predictors, namely, SST anomalies, saturation deficient (SD; Camargo et al. 2014) anomalies,
170 850 hPa relative vorticity anomalies, and vertical wind shear anomalies between 850 and 200
171 hPa. We also tested the sensitivity of using column-integral relative humidity (CRH; Tippett
172 et al. 2011) anomalies as an alternative humidity environmental factor to replace SD in our
173 CNN model (details in Section 5). Both SD and CRH are calculated following Bretherton et
174 al. (2014). To derive these environmental factors, we firstly interpolate the related prognostic
175 monthly mean data (i.e., temperature, humidity and wind velocities) to 2°×2° resolution grid,
176 then make the diagnostic calculations and seasonal average. Anomalies are finally calculated
177 as the departures from the seasonal mean. These four seasonal averaged predictors are then
178 concatenated as the three-dimensional (i.e., longitude, latitude and variables) inputs to the
179 CNN (Figure 1b). We note that, actual data matrices, rather than images, were implemented
180 as the CNN inputs in our framework. The dimension of the output is one, which depicts the
181 seasonal mean NTC, NHU or ACE, respectively.
182 One big important barrier to implementing deep learning for climate study is the limited
183 sample size of observational data for proper model training (Ham et al. 2019). Due to relative
184 short period of reliable TC observation records, we used transfer-learning technique
185 (Yosinski et al. 2014) to circumvent the limited amount of training datasets and increased the
186 number of training data, in a similar manner as in Ham et al. (2019). We collected suites of
187 high-resolution Community Earth System Model version 1.3 (CESM1.3; ~0.25° resolution
188 for atmosphere/land and ~0.1° resolution for ocean/sea-ice) simulations, including 1850
189 preindustrial control and historical transient climate simulations (Chang et al. 2020), High
190 Resolution Model Intercomparison Project (HighResMIP) 1950s control and historical
191 transient climate simulations (Roberts et al. 2020), and decadal prediction simulations
192 (Yeager et al. 2022), together with Weather Research and Forecast (WRF) based tropical
193 channel model (TCM; 27km) large ensemble hindcast simulations (Fu et al. 2019), to
194 increase the training samples. Both high-resolution CESM1.3 and WRF TCM can faithfully
195 capture the climatology and interannual variability of seasonal mean TC activity and
196 associated large-scale ambient environments (Figure 1a; refer Chang et al. 2020 and Fu et al.

2019 for details). We combined these TC-permitting high-resolution climate simulation results with state-of-the-science European Centre 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) dataset 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 simulations forced with historical condition (climate 1850s/1950s control simulations and transient simulation years before 1979) to train our CNN model, although our transient climate simulations projected the future climate changes up to 2100 (Representative Concentration Pathway 8.5 for 2006-2100). We implement our future climate simulations as additional segments for testing processes, aiming to explore the feasibility of applying our CNN model to the future climate projections. Table 1 summarizes the sample size of each datasets utilized in CNN training. We note that simulations conducted by WRF TCM only cover Northern Hemisphere; therefore, 2906 (1506) samples are used to train the CNN models for the Northern (Southern) Hemisphere.

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 the fine-tuned CNN model using ERA5 reanalysis and IBTrACS TC observations. Figure 1b illustrated the architecture of our CNN model. 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 3×3. 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 2×2 grid with stride of 1. The Scaled Exponential Linear Unit (SELU; Klambauer et al. 2017) is used as the activation function to introduce

nonlinearity in the extracted features. 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 and 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 dataset. 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 the pre-satellite era without the support from geostationary satellite imagery (i.e. 1960s) and may significantly underestimate seasonal TC activity (Landsea 2007; Landsea et al. 2008, 2014; Vecchi and Knutson 2008, 2010; Vecchi et al. 2021). However, we still include these records in our CNN training processes to enlarge the sample size. 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 and 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

253 are predicted. This is done for 70 years, removing each year's data point successively. In
254 addition, we also cross validate CNN using 20th century reanalysis products, instead of ERA5,
255 focusing on the NAT TC activities before 1950, which were excluded in the training
256 processes. The details are discussed in Section 4.

257 For each of climate model pre-trained CNN, we further deployed 12-member ensemble
258 trainings during the transfer-learning processes on ERA5 and IBTrACS. Although the initial
259 training weights and biases are identical, stochastic Adam optimizer results in deviated CNN
260 regression models due to random choice of mini-batch. In total, for each of TC-active ocean
261 basins and each of seasonal mean TC metrics, we trained ensemble of CNN with
262 600-members. Each model is then used to make a prediction and the actual prediction skill is
263 evaluated by averaging all ensembles, while the ensemble spread indicate the prediction
264 uncertainties.

265

266 2.3 CNN training, validation and interpretation

267 In the rest of this paper, for the sake of clarity, CNN emulated NTC, NHU and ACE
268 were plotted in blue, red and yellow, respectively. Figure 2 illustrates the training and
269 LOOCV skills of seasonal NTC in each TC active ocean basin. It exhibits significant high
270 skills in reproducing NTC variability, in terms of high Pearson correlation coefficient,
271 mean-square skill score (MSSS; Kim et al., 2012; Li et al. 2013) and low root mean square
272 errors (RMSE) over all basins except for NIO. For the NIO, although LOOCV correlation
273 coefficient is still statistically significant at 99% level, it is notably degraded comparing to
274 the trainings and LOOCVs in other TC active basins, which may imply that some nontrivial
275 large-scale environmental factors were not involved in the current CNN predictor variables.
276 In the context of global mean TC activity, the correlation coefficients between observation
277 and ensemble mean CNN training and LOOCV is 0.98, and 0.93, respectively. RMSE
278 between observation and CNN emulated global NTC is 2.27 and 4.95 in training and
279 LOOCV ensemble mean, both of which are significantly smaller than the observed NTC
280 climatology of 79.27 and standard deviation of 12.85. The ensemble spreads of training and

LOOCV, which we define as the entire range across 600-member ensemble CNNs and shown as the shadings in Figure 2, are also considerably small among all TC active basins. Similar high skills can also be obtained in the ensemble CNNs that depict seasonal NHU (Figure 3) and ACE (Figure 4) for the period of 1950-2019. All of these promising assessment results are encouraging for proceeding forward with further applications using our CNN framework.

Although our trained CNN ensemble frameworks illustrate substantial high skills in reproducing seasonal TC variability in the observational records, an often-cited caveat of machine learning is the challenge of physical interpretability compared to the more conventional method. The potential lack of interpretability has implications for the perceived credibility of the model, where machine learning models may achieve promising results for the wrong reasons (Lapuschkin et al. 2019). To decipher these "black boxes" and gain an insight of what predictor features that predominate CNN's regression output, we further conducted occlusion sensitivity analysis (Zeiler and Fergus 2014).

The general concept of our 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 change degree as a function of the mask position. Intuitively, if a portion of input layer is trivial to the prediction accuracy, and one 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 variables at different geographical locations in determining seasonal TC variability. In practice, for each of four predictor variables during 1950-2019, the patch of 6°×6° (i.e. 3×3 grid) box mask is replaced by its own time-independent climatological values, and new predictions are conducted by the original trained CNN with the altered inputs. Therefore, we obtained the new predictions of seasonal TC variability variability as functions of mask geographical locations and predictor variables. We measure the change in prediction skills as the RMSE between new predictions and original training during 1950-2019, through which we highlight the spatial portion of the predictor variables are the most important to the prediction skills for our trained CNN

framework: when that portion of the predictor is occluded to the climatology, the representation of seasonal TC variability will be significantly degraded.

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 Supplementary Figure 1 and 2, respectively. As noted earlier in the Section1, it is not surprising to see that the relative importance of each predictor is different among the various TC active basins. Our occlusion sensitivity analysis broadly suggests that large-scale dynamical factors, especially 850hPa vorticity, contribute more than thermal factors in determining NTC and NHU in each basin, and this is physically consistent with the TC dynamic genesis potential index that recently proposed by Wang and Murakami (2021). Take NAT NTC as an example (Figure 5m-p), 850hPa 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 developed region and Gulf of Mexico, while SST and SD's contributions are relative small. We emphasize that these occlusion sensitivity results does not imply SST are insignificant for regulating seasonal NTC and NHU, because local dynamical factors can be influenced by the remote SST forcing. For example, extensive observational and modeling studies found that Atlantic hurricane 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 (Tang and Neelin, 2004; Smith et al., 2007; Patricola et al., 2016), 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 subcutaneously in the CNN ACE emulation. We hypothesized 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 has been partially examined by the Murakami et al. (2018), who found that the hyperactive six major hurricanes during 2017 NAT hurricane season is

primarily attributed to the local record high SST anomaly confined to the tropical Atlantic, while the remote influences caused by the moderate La Niña condition in the Pacific played the minor role. More rigorous analysis is planed in the future study. Nonetheless, Occlusion sensitivity analysis provides interpretability to the presented high CNN emulation fidelity, which raises our confidence that our trained CNN frameworks are capable of learning physically plausible mechanism, 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 to emulate seasonal TC activity, 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 TC studies. We will start with the application of ensemble CNNs in the seasonal TC prediction.

Seasonal 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 designed to depict the concurrent empirical relationship between seasonal averaged large-scale environmental factors and TC activities, we can apply the trained ensemble CNNs to the operational seasonal forecast systems that are capable to forecast CNN predictors at various lead times to predict the subsequent seasonal TC activities. 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.

Both NCEP CFSv2 and ECMWF SEAS5 are atmosphere–ocean–land fully coupled models that aimed for operational seasonal forecasting. But due to the relative coarse resolutions and dynamical model intrinsic biases, CFSv2 and SEAS5 are not capable to faithfully resolve seasonal TC activities. There are 16 CFSv2 runs per day in operations; four out to ~9 months, three out to 3 seasons, and nine out to 45 days. Focusing on our seasonal

prediction scope, we only implement those ~9 months long forecasts. CFSv2 hindcast simulations were ran from December 1981 to March 2011, with initial conditions of the 00, 06, 12 and 18 UTC cycles for every 5th day starting 00 UTC 1 January of every year (i.e., 292 forecasts for every year during 1982-2010). Operational real-time forecasts were available since 1 April 2011, and were operated 4 cycles everyday. For the details of CFSv2 system design, we refer to Saha et al. (2014) for further reading. To obtain the dynamical model forecasted CNN predictors for Northern (Southern) Hemisphere TC seasons of June-November (December-May) seasonal mean, the earliest CFSv2 operational predictions should be initialized at 00 UTC 27 January (July). 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.

For all CFSv2 and SEAS5 dynamical ensemble forecasts, we firstly interpolated the monthly mean prognostic variables (i.e., temperature, humidity and wind velocities) from each of ensemble member to the uniform 2°×2° 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 seasonal average and ensemble average to obtain the CNN inputs. We utilized these ensemble averaged predicted large-scale environmental factors to drive each of 600-member ensemble CNN emulations, therefore, the uncertainty of our hybrid seasonal TC predictions were determined by the ensemble CNNs, rather than from dynamical model. The actual hybrid prediction skills of 6-month seasonal TC activity depend on both the fidelity of our CNN machine learning model and the CFSv2/SEAS5 dynamical prediction accuracy.

Table 3 summarizes the hybrid prediction ensemble mean skills in different TC-active basins. Figures 6, 7 and 8 illustrate the prediction uncertainties of NTC, NHU and ACE, respectively. In general, our hybrid prediction framework yields promising seasonal prediction skills in NAT, ENP, WNP and SPO. For the NAT predictions, 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 might be attributable to the different dynamical model skills in the NAT between CFSv2 and SEAS5. May initialization exhibits the highest skills, with the correlation of the NTC, NHU and ACE of 0.72, 0.69, and 0.67 (Table 3).. For reference, the statistical forecasts issued by Colorado State University (CSU) reported that the prediction correlation coefficients of the NAT NTC, NHU and ACE from their June (April) initialized forecasts during 1982–2021 (1995-2021) are 0.71 (0.31), 0.41 (0.00) and 0.41 (0.02), respectively (data is available from https://tropical.colostate.edu/archive.html#verification). Using hybrid technique with the aid from the Geophysical Fluid Dynamical Laboratory (GFDL) Forecast-Oriented Low Ocean Resolution (FLOR) model, Murakami et al. (2016) reported a high correlation coefficient of 0.75 for the seasonal NAT NTC predictions when their hybrid model was initialized in June, while their targeted forecast season is July-November over the 1980-2014 period.. Comparing to the CSU and GFDL, our hybrid predictions are capable of competing with or even slightly outperforming these pioneering

seasonal prediction works in the NAT sector, which demonstrates the feasibility of our hybrid framework in seasonal TC predictions.

Similar robust prediction skills can also be obtained in the ENP and WNP. In contrast to the NAT, June initializations exhibit highest prediction skills in terms of correlation and MSSS in these two basins. We highlight the significant high ACE (NHU) correlations of 0.85 (0.75), and 0.85 (0.76) for the June initialized predictions in the ENP and WNP (Table 3). To the best of our knowledge, these high fidelities have never been reported in the previous studies yet. Skillful June-November ACE predictions can be obtained since January (Figure 8b,c), and skillful NTC and NHU predictions can be made as early as in April (Figure 6b,c and Figure 7b,c). These robust prediction skills make our hybrid prediction model an invaluable tool for North Pacific TC seasonal predictions.

However, skillful seasonal predictions in the SPO can only be obtained just ahead of the TC season, December-May (Table 3), and the hybrid framework performs very poorly in the SIO predictions. For the NIO, although hybrid framework fails to produce skillful NTC predictions, it offers skillful NHU prediction in June initialization and ACE predictions at the lead month of 1-4. As we noted earlier, the skillful hybrid seasonal TC predictions presented here are contributed jointly from the dynamical seasonal predictions and the trained ensemble CNN models. In the future studies, we plan to implement multi-model seasonal forecasts from Copernicus Climate Change Service (C3S) Climate Data Store (CDS), including but not limit to, UK Met Office, Météo France, and JMA, and apply the ensemble CNNs in order to see if we are able to obtain higher prediction skills by using the multi-model ensembles.

4. Reconstruct historical TC records using CNN and 20th century reanalysis

Another novel application of our CNN framework can be the reconstruction of historical TC activity using the 20th century reanalysis products. Although the IBTrACS NAT TC record can trace back to 1851, before routine airborne reconnaissance in the 1940s and geostationary satellite imagery in 1960s, mounting papers suggested that TC climatology is underestimated before 1970s (Landsea 2007; Vecchi and Knutson 2008, 2011; Vecchi et al.,

2021; Emanuel 2010), and numerous efforts have been made to correct for 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 remedy this uncertainty is to leverage state-of-the-art 20th century historical reanalysis to either directly identify TC-like structures (Truchelut et al., 2013; Chand et al. 2022) from the background flow or use monthly mean large-scale environmental factors for dynamical downscaling (Emanuel 2010, 2021). However, these two methods do not provide consensus on the detection of trends, especially in the NAT basin. In this study, we attempt to utilize the 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.

We applied our trained ensemble CNN framework to two historical reanalysis products to derive long-term historical records of TCs globally: NOAA 20th Century Reanalysis Version 3 (1836-2015; Silvinski 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, and uses observed monthly mean SST and sea-ice concentration as boundary conditions to create a comprehensive record of the atmospheric circulation. CERA-20C is a 10-member ensemble coupled reanalysis of the twentieth 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 the trends of the past century extreme weather events and climate. Similar post-processing techniques have been conducted to 20CRv3 and CERA-20C to obtain the CNN predictor variables: we firstly interpolated the monthly mean thermal and dynamical variables to 2°×2° resolution grid, then computed the ensemble average and carried out the diagnostic calculations, and finally derived the 6-month seasonal average as the 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 digitally available ship locations between 1878 and 1965. Combining these statistical adjustments with original IBTrACS observations, we attempt to qualitatively evaluate the fidelity of our historical reconstructions. Our CNN transfer learning is based on the ERA5 1950-2019 data; therefore data before 1950 are excluded in the CNN training processes, which acting as the independent segment for the cross verification of our CNN models. 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, correlation of NTC, NHU and ACE between 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 spreads are increased comparing to the training period of 1950-2019, spreads before 1949 still remain considerably small. These promising cross validations again highlight the fidelity of our CNN framework. It is worthwhile to note that, although our reconstruction approach is very different from Vecchi and Knutson (2008, 2011), Truchelut et al. (2013) and Chand et al. (2022), all our results point to the peak of seasonal TC activity around 1880-1890 (Figure 9a, b; 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 century scale secular trend (Figure 9a-c). If we assume the fidelity of CNN based historical reconstructions are equally well outside of NAT (unfortunately, TC observations outside the Atlantic were insufficient for validation), our results suggest global NTC experienced insignificant upward trend during 1836-1949, but NHU and ACE were significantly decreased (Figure 9d-f).

Supplementary Figure 3 exhibits the historical reconstruction using CERA-20C. Cross-validation with adjusted observations also yield significant high correlation

coefficients and small ensemble spreads. In the context of global TC activity trend, CERA-20C indicated insignificant downward trend of NTC and NHU during 1901-1949, but significant downward trend of ACE. Nonetheless, our ensemble CNN framework provides ancillary and independent method to reconstruct historical TC records in the 19th and 20th century. Although our results do not suggest robust decreasing trend of global TC activity as postulated by the recent study (Chand et al. 2022) that used 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 dominated by the internal climate variability and/or aerosol effects (Vecchi and Knutson 2008, 2011; Vecchi et al. 2021).

5. Explore future TC activity using CNN and CMIP6 projection

Encouraged by the promising fidelities in emulating TC activities in the past historical and present climate conditions, we further explored the feasibility of applying our trained CNN framework to the TC future climate projections.

Although broad consensus has emerged about the increased TC intensity and rainfall rate in the future warmer climates (Patricola and Wehner 2018; Knutson et al. 2020), the uncertainties of future TC frequencies are poorly constrained. For example, vest majority of climate model future projections with TC-permitting horizontal-resolution (≤50km) indicate moderate-to-weak decreasing trend of future TC frequency (IPCC 2013; Zhao et al. 2009, Knutson 2010, 2020; Walsh et al. 2016), but a few others suggest an increased TC frequency (Vecchi et al. 2019). Using the statistical-deterministic method developed by Emanuel et al. (2008), Emanuel et al. (2013) proposed an increasing trend of future TC frequency by downscaling CMIP5 models, while another Poisson regression model (Tippett et al. 2011) based statistical downscaling approach projected two diverging TC frequency climate trends, in which the TC frequency will increase or decrease depending on the selection of the atmosphere moisture variable (CRH or SD) (Camargo et al. 2014; Lee et al. 2020). Specifically, SD-based downscaling leads to a decrease of TC activity, while using CRH results in an increase. The results from the SD choice were consistent with the dynamical

simulated TC reduction from Zhao et al. (2009), but those from the CRH choice were in contrast. Given that our CNN is trained devotedly to improve the representation of TC temporal variability, it is intriguing to examine if our CNN framework, which is also trained on present-day climate condition, is applicable to project future TC activities. In particular, we want to test the sensitivity of moisture predictor variables to the sign of projected TC activity trends, consequently, we repeated the CNN training processes as thoroughly introduced in Section 2, but replaced SD with CRH for the sensitivity test. 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 also 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. As the independent segment for cross validation purposes, we made a comprehensive comparison between the CNN emulated TC activities computed from high-resolution CESM1.3 large-scale outputs with those directly simulated in 1980-2100.

Supplementary Figure 4 exhibits the comparisons between CNN emulated and 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 indicate same sign of change, the CNN emulated NTC increase with time while direct dynamical simulations suggest a significant decreasing trend. Moreover, the ensemble spreads of 600-member ensemble CNN emulation are substantial huge, which is another aspect of significant degradation from the proposed applications in the seasonal prediction and historical reconstruction. These poor performances greatly reduce the potential feasibilities in investigating TC future projections.

However, it is noteworthy to point out that transfer-learning technique was employed to both CNN_SD and CNN_CRH, which indicates that CNN_SD and CNN_CRH are strongly constrained by the recent observations. When these observation-constrained CNNs were applied to climate model projections, they may produce biased results due to intrinsic model biases (e.g., SST cold bias, Wang et al. 2014; double ITCZ, Tian et al. 2020) in the climate

561 model projections. To focus on the climate change effects based on the high-resolution
562 CESM1.3 simulations in this study, therefore, we made another sensitivity test that discard
563 transfer-learning technique in the training processing, and use CNNs trained only on the
564 climate model simulations with historical forcing (again, 1980-2100 transient climate
565 simulations were excluded). Figure 10 illustrates similar comparisons but emulated by the
566 non-transfer-learning CNN_SD and CNN_CRH. Notably, we can detect promising
567 improvements: both CNN_SD and CNN_CRH correctly capture the decreasing trend of
568 future TC activity as revealed by the 25-km resolution CESM1.3 simulations, and exhibit
569 robust correlation coefficients. Furthermore, the uncertainty spreads of CNN emulations are
570 significantly reduced comparing to those from the observation-constrained transfer-learned
571 CNNs. Note that, although both CNN_SD and CNN_CRH are trained on the climate
572 historical forcing, where global warming signal is much weaker than those in the mid and late
573 21th century, they still work reasonably-well to predict TC activity changes in the future
574 climate. However, some inconsistencies still exist. For instance, CNN emulated decreasing
575 trends of NTC and NHU are too strong comparing to those directly resolved in the CESM1.3.
576 Nonetheless, the presented results provide positive insights that properly trained CNNs only
577 using present-day climate conditions can still be applicable to investigate the climate change
578 problem.

579 Besides the cross validation using high-resolution CESM simulations, we also applied
580 the non-transfer-learning CNN_SD and CNN_CRH to different 36 CMIP6 models (see
581 Supplementary Table S1 for a list) that participated to the Scenario Model Intercomparison
582 Project (ScenarioMIP), aiming to gain multi-model mean of future TC projections. Figure 11
583 shows the global integrated, and Northern/Southern Hemisphere mean CNN emulated NTC
584 from 36 CMIP6 models. Large-scale environmental conditions for the period of 1950–2015
585 are taken from the historical simulations, and those for the periods from 2015 to 2100 are
586 from future climate simulations under shared socio-economic pathway 5-8.5 (SSP585)
587 scenario. NHU and ACE emulation are illustrated in Figure 12, and 13, respectively.
588 Supplementary 5 and 6 further compare the CNN emulations in each individual TC-active

589 basin. Consistent with many previous studies based on the dynamical model proejctions
590 (Knutson et al. 2015, 2020; Walsh et al. 2016; Sugi et al. 2017; Yoshida et al. 2017), our
591 CNN emulation frameworks also indicate a decrease of seasonal TC activity in the future
592 warmer climate, either using SD or CRH as moisture predictor variable. Specifically, using
593 CNN_SD (CNN_CRH), 35 (36) of total 36 CMIP6 models project decreasing of global mean
594 NTC, which is primarily driven by the significant reduction of NTC in the Southern
595 Hemisphere (Figure 11). Pinpoint to each basin, both CNN_SD and CNN_CRH emulated
596 robust NTC decrease in the SIO (Supplementary Figure S5), as also seen in CMIP5 and
597 HighResMIP studies (Tory et al., 2013; Bell et al., 2019; Roberts et al., 2020). In addition,
598 NTC will be declined in the WNP and SPO, but increased in the NIO. No significant change
599 is projected in the ENP, but NAT yields a mixed projection: increase of NTC up to 2050s, but
600 decrease afterwards. We will explore the plausible physical mechanism for this unusual NAT
601 NTC projection in the future study.
602 Our CNN framework project weaker decrease trends of the global integrated NHU and
603 ACE (Figure 12 and 13). Majority of CMIP6 models project slight increases of NHU and
604 ACE in the Northern Hemisphere, which compensate the substantial decrease trends in the
605 Southern Hemisphere. It is intriguing to note that, although our CNN framework is purely
606 statistical based and the NTC are trained independently from the NHU or ACE, it can still
607 capture the increasing trend of intense TC fraction (Supplementary Figure 6), which is a
608 general consensus emerged from the TC-permitting climate model simulations (Knutson et
609 al., 2020).
610
611 **6. Conclusions and discussions**
612 The overarching goal of this study is to extend the pioneer works of TC genesis index,
613 and to use machine learning technique to better reveal the empirical linkage between
614 large-scale environmental conditions and temporal TC variability. Leveraging large suites of
615 TC-permitting high-resolution climate model simulations and reanalysis products, we trained
616 ensemble of CNNs that yield high fidelity in capturing observed TC

interannual-to-multidecadal variability. Using occlusion sensitivity analysis, we confirm that physical plausible mechanisms about TC environmental favorability are properly implemented in the trained CNN framework. We further demonstrate that this CNN framework can be applicable to many aspects of TC-climate studies. In particular, our preliminary applications highlight:

Using large-scale CNN predictors forecasted by the operational NCEP CFSv2, the CNN based statistical-dynamical hybrid model can make skillful NAT (WNP and ENP) predictions in early February (late January). In particular, correlation coefficients of the predicted and observed ENP and WNP ACE reach to 0.85 when the hybrid prediction is initialized in June, just at the beginning of conventionally defined hurricane season.

By applying 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 dominated by the internal climate variability and/or aerosol effects, which agrees with the arguments from previous studies that using entirely different methods (Vecchi and Knutson 2008, 2011; Vecchi et al., 2021; Chand et al., 2022).

The observation-constrained (transfer-learning-employed) CNNs illustrate big biases in future climate projections, presumably attribute to the "out-of-sample" error. However, if we discard transfer-learning technique and train CNN entirely on climate model simulations forced by the climate historical condition, our CNN frameworks can promisingly capture the decreasing trend of future TC activity, which is consistent with our direct high-resolution CESM1.3 projections in TC-permitting 25-km resolution. In addition, this projection is independent to the choice of CNN moisture predictor variables. We further applied these non-transfer-learning CNNs to 36 CMIP6 models, and found a robust decrease of global TC activity in the future climate, as documented by the recent TC and climate change assessment (Knutson et al., 2020).

In addition to these promising applications, one particularly intriguing property of CNNs is their extreme computational efficiency. Once the CNN predictors are properly computed, the computational cost of CNN prediction is negligible. Our CNN approach is

complementary to earlier TC genesis index studies. We suggest that machine learning approaches, especially deep learning models, may provide a new avenue to improve our understanding of TC variability and future changes. In the future, it would be desirable to train deep learning models on multiple climate model large-ensemble simulations that are able to faithfully reproduce seasonal TC activities, for example, 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), in order to test the sensitivity of CNN seasonal TC activity emulations on the training sets.

Open Research

The data used in this manuscript are available from the following sources: ECMWF ERA5 (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means), SEAS5 (https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-monthly-single-levels), CERA20C (https://apps.ecmwf.int/datasets/data/cera20c/levtype=sfc/type=an/), NCEP CFSv2 (https://www.ncei.noaa.gov/thredds/model/cfs.html), NOAA 20CRv3 (https://psl.noaa.gov/data/gridded/data.20thC_ReanV3.html), CMIP6 models (https://esgf-node.llnl.gov/search/cmip6/), IBTrACS (https://www.ncdc.noaa.gov/ibtracs/index.php?name=ib-v4-access). High-resolution CESM1.3 simulations are available at https://ihesp.github.io/archive. As part of this paper, we will also release the trained ensemble CNNs for seasonal TC activity to allow future studies once the manuscript is accepted for peer-reviewed publication.

Acknowledgments

This research is supported by the U.S. Department of Energy under Award Number DE-SC0020072.

Conflict of Interest

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

673

674 **Reference**

675 Bell, G. D., Halpert, M. S., Schnell, R. C., Higgins, R. W., Lawrimore, J., Kousky, V. E.,
676 Tinker, R., Thiaw, W., Chelliah, M., and Artusa, A. (2000). Climate assessment for
677 1999. *Bulletin of the American Meteorological Society*, 81 (6), S1-S50.
678 doi:10.1175/1520-0477(2000)81[s1:CAF]2.0.CO;2

679 Bell, S. S., Chand, S. S., Tory, K. J., Dowdy, A. J., Turville, C., & Ye, H. (2019). Projections
680 of southern hemisphere tropical cyclone track density using CMIP5 models. *Climate*
681 *Dynamics*, 52(9-10), 6065–6079. https://doi.org/10.1007/s00382-018-4497-4

682 Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data inference and
683 subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, 11(1),
684 376-399.

685 Bruyère, C. L., Done, J. M., Holland, G. J., & Fredrick, S. (2014). Bias corrections of global
686 models for regional climate simulations of high-impact weather. *Climate Dynamics*,
687 43(7), 1847-1856.

688 Camargo, S. J., Emanuel, K. and Sobel, A. H. (2007a). Use of a genesis potential index to
689 diagnose ENSO effects on tropical cyclone genesis. *J. Clim.*, 20, 4819–4834.

690 Camargo, S. J., Sobel, A. H., Barnston, A. G. and Emanuel, K. (2007b). Tropical cyclone
691 genesis potential index in climate models. *Tellus, Ser. A*, 59, 428–443.

692 Camargo, S.J., Tippett, M. K., Sobel, A. H., Vecchi, G. A., and Zhao, M. (2014). Testing the
693 performance of tropical cyclone genesis indices in future climates using the HiRAM
694 model. *J. Climate*, 27, 9171– 9196. https://doi.org/10.1175/JCLI-D-13-00505.1.

695 Chattopadhyay, A., Hassanzadeh, P. & Pasha, S. (2020a). Predicting clustered weather
696 patterns: A test case for applications of convolutional neural networks to spatio-temporal
697 climate data. *Sci. Rep.*, 10, 1317. https://doi.org/10.1038/s41598-020-57897-9

698 Chattopadhyay, A., Nabizadeh, E., &Hassanzadeh, P. (2020b). Analog forecasting of
699 extreme-causing weather patterns using deep learning. *Journal of Advances in Modeling*
700 *EarthSystems*,12, e2019MS001958. https://doi.org/10.1029/2019MS001958

701 Chand, S.S., Walsh, K.J.E., Camargo, S.J. et al. (2022). Declining tropical cyclone frequency
702 under global warming. *Nat. Clim. Chang.,* 12, 655–661.
703 https://doi.org/10.1038/s41558-022-01388-4

704 Chang, C.-C., & Wang, Z. (2018). Relative impacts of local and remote forcing on tropical
705 cyclone frequency in numerical model simulations. *Geophysical Research Letters*, 45,
706 7843–7850. https://doi.org/10.1029/2018GL078606

707 Chapman, W. E., Subramanian, A. C., Delle Monache, L., Xie, S. P., & Ralph, F. M. (2019).
708 Improving atmospheric river forecasts with machine learning. *Geophysical Research*
709 *Letters*, 46, 10,627–10,635. https://doi.org/10.1029/ 2019GL083662

710 Davenport, F. V., & Diffenbaugh, N. S. (2021). Using machine learning to analyze physical
711 causes of climate change: A case study of U.S. Midwest extreme precipitation.
712 *Geophysical Research Letters*, 48, e2021GL093787.
713 https://doi.org/10.1029/2021GL093787

Delworth, T. L., 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.

Elsner, J. B., and Jagger, T. H. (2013). Frequency models. Hurricane Climatology: A Modern Statistical Guide Using R. *Oxford University Press*, 161–193.

Emanuel, K., Sundararajan, R., and Williams, J. (2008). Hurricanes and global warming: Results from downscaling IPCC AR4 simulations. *Bull. Amer. Meteor. Soc*., 89, 347–368, https:// doi.org/10.1175/BAMS-89-3-347.

Emanuel, K. (2010). Tropical cyclone activity downscaled from NOAA- CIRES Reanalysis, 1908-1958. *J. Adv. Model. Earth Syst.*, 2, 1–12.

Emanuel, K. (2013). Downscaling CMIP5 climate models shows increased tropical cyclone activity over the 21st century. *Proc. Natl. Acad. Sci. USA*, 110, 12 219–12 224, https://doi.org/10.1073/ pnas.1301293110.

Emanuel, K. (2021). Response of global tropical cyclone activity to increasing CO2: results from downscaling CMIP6 models. *Journal of Climate*, 34(1), 57-70.

Gray, W. M. (1968). Global view of the origin of tropical disturbances and storms. *Mon Weather Rev*, 96, 669–700.

Gray, W. M. (1979). Hurricanes: Their formation, structure and likely role in the tropical circulation. *Meteorol. Tropical Oceans*, 77, 155–218.

Ham, Y. G., Kim, J. H. & Luo, J. J. (2019). Deep learning for multi-year ENSO forecasts. *Nature,* 573, 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. (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., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B. M. (2019). SEAS5: the new ECMWF seasonal forecast system. *Geosci. Model Dev*., 12, 1087–1117, https://doi.org/10.5194/gmd-12-1087-2019.

Kim, H. S., Ho, C. H., Kim, J. H., and Chu, P. S. (2012). Track-pattern-based model for seasonal prediction of tropical cyclone activity in the western North Pacific. *J. Climate*, 25, 4660– 4678. doi:10.1175/JCLI-D-11-00236.1.

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

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., and Coauthors (2010). Tropical cyclones and climate change. *Nat. Geosci.*, 3, 157–163, https://doi.org/10.1038/ngeo779.

Knutson, T. R., Sirutis, J. J., Zhao, M., Tuleya, R. E., Bender, M., Vecchi, G. A., Villarini, G. and Chavas, D. (2015). Global projections of in- tense tropical cyclone activity for the

late twenty-first century from dynamical downscaling of CMIP5/RCP4.5 scenarios. *J. Climate*, 28, 7203–7224, https://doi.org/10.1175/JCLI-D-15- 0129.1.

Knutson, T. R., and Coauthors (2020). Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bull. Amer. Meteor. Soc.*, 101, E303–E322, https:// doi.org/10.1175/BAMS-D-18-0194.1.

Laloyaux, P. et al. (2018). CERA-20C: a coupled reanalysis of the twentieth century. *J. Adv. Model. Earth Syst.,* 10, 1172–1195.

Landsea, C. W. (2007). Counting Atlantic tropical cyclones back to 1900. *Eos, Trans. Amer. Geophys. Union*, 88(18), 197-202, doi:10.1029/ 2007EO180001.

Lapuschkin, S., Wäldchen, S., Binder, A. et al. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nat. Commun.*, 10, 1096. https://doi.org/10.1038/s41467-019-08987-4

Lee, C. Y., Camargo, S. J., Sobel, A. H. and Tippett, M. K. (2020). Statistical-dynamical downscaling projections of tropical cyclone activity in a warming climate: Two diverging genesis scenarios. *J. Climate*, 33, 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.

Liu, Y. et al. (2016). Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *arXiv preprint arXiv:1605.01156*.

Menkes, C. E., Lengaigne, M., & Marchesiello, P. (2012). Comparison of tropical cyclogenesis indices on seasonal to interannual timescales. *Climate Dynamics*, 38, 301–321.

McGuire, M. P., & Moore, T. W. (2022). Prediction of tornado days in the United States with deep convolutional neural networks. *Computers & Geosciences*, 159, 104990.

Mizuta, R., Murata, A., Ishii, M., Shiogama, H., Hibino, K., Mori, N., et al. (2017). Over 5,000 years of ensemble future climate simulationsby 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

Murakami, H., and Wang, B. (2010). Future change of North Atlantic Tropical Cyclone tracks: Projection by a 20-km-mesh global atmospheric model. *J. Clim.*, 23, 2699–2721.

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.

Patricola, C. M., Chang, P., and Saravanan, R. (2016). Degree of simulated suppression of Atlantic tropical cyclones modulated by flavour of El Niño. *Nat. Geosci.*, 9, 155–160.

Patricola, C. M., Saravanan, R., and Chang, P. (2017). A teleconnection between Atlantic sea surface temperature and eastern and central North Pacific tropical cyclones. *Geophys. Res. Lett.*, 44, 1167–1174, doi:10.1002/2016GL071965

798 Saha, Suranjana and Coauthors (2014). The NCEP Climate Forecast System Version 2. *J.*
799 *Climate*, 27(6), 2185–2208. doi: http://dx.doi.org/10.1175/JCLI-D-12-00823.1.
800 Slivinski, L. C. et al. (2019). Towards a more reliable historical reanalysis: improvements for
801 version 3 of the Twentieth Century Reanalysis system. *Q. J. R. Meteorol. Soc.*, 145,
802 2876–2908.
803 Smith, S. R., Brolley, J., O'Brien, J. J., and Tartaglione, C. A. (2007). ENSO's impact on
804 regional U.S. hurricane activity. *J. Clim.*, 20, 1404–1414.
805 Sugi, M., Yoshimura, J., and Yoshida, K. (2017). Projection of future changes in the
806 frequency of intense tropical cyclones. *Climate Dyn.*, 49, 619–632,
807 https://doi.org/10.1007/s00382-016-3361-7.
808 Tang, B. H., and Neelin, J. D. (2004). ENSO influence on Atlantic hurricanes via
809 tropospheric warming. *Geophys. Res. Lett.*, 31, L24204, doi:10.1029/2004GL021072.
810 Tian, B., & Dong, X. (2020). The double-ITCZ Bias in CMIP3, CMIP5and CMIP6 models
811 based on annual mean precipitation. *Geophysical Research Letters*, 47, e2020GL087232.
812 https://doi.org/10.1029/2020GL087232
813 Tippett, M. K., Camargo, S. J., and Sobel, A. H. (2011). A Poisson regression index for
814 tropical cyclone genesis and the role of large-scale vorticity in genesis. *J. Clim.,* 24,
815 2335–2357.
816 Tory, K. J., Chand, S. S., McBride, J. L., Ye, H., & Dare, R. A. (2013). Projected changes in
817 late-twenty-first-century tropical cyclone frequency in 13 coupled climate models from
818 phase 5 of the Coupled Model Intercomparison Project. *Journal of Climate*, 26(24),
819 9946–9959. https://doi.org/10.1175/JCLI-D-13-00010.1
820 Truchelut, R. E., Hart, R. E., & Luthman, B. (2013). Global identification of previously
821 undetected pre-satellite-era tropical cyclone candidates in NOAA/CIRES
822 Twentieth-Century Reanalysis data. *Journal of applied meteorology and climatology*,
823 52(10), 2243-2259.
824 Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. (2014). How transferable are features in deep
825 neural networks? *Adv. Neural Inf. Process. Syst.,* 27, 3320–3328.
826 Yoshida, K., Sugi, M., Mizuta, R., Murakami, H., and Ishii, M. (2017). Future changes in
827 tropical cyclone activity in high-resolution large-ensemble simulations. *Geophys. Res.*
828 *Lett.*, 44, 9910– 9917. https://doi.org/10.1002/2017GL075058.
829 Vecchi, G. A., and Knutson, T. R. (2008). On estimates of historical North Atlantic tropical
830 cyclone activity. *J. Climate*, 21, 3580-3600.
831 Vecchi, G. A., and Knutson, T. R. (2011). Estimating annual numbers of Atlantic hurricanes
832 missing from the HURDAT database (1878–1965) using ship track density. *J. Climate*,
833 24, 1736–1746.
834 Vecchi, G. A., Landsea, C., Zhang, W., Villarini, G. & Knutson, T. (2021). Changes in
835 Atlantic major hurricane frequency since the late-19th century. *Nat. Commun.,* 12(1),
836 1-9.
837 Walsh, K. J., and Coauthors (2016). Tropical cyclones and climate change. *Wiley Interdiscip.*
838 *Rev.: Climate Change*, 7, 65–89. https://doi.org/10.1002/wcc.371.

Wang, C., Zhang, L., Lee, S. K., Wu, L., & Mechoso, C. R. (2014). A global perspective on CMIP5 climate model biases. *Nature Climate Change*, 4(3), 201-205.

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, 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, e2021MS002502. https://doi.org/10.1029/2021MS002502

WMO (2008). Report from expert meeting to evaluate skill of tropical cyclone seasonal forecasts. *World Meteorological Organization Tech. Doc.,* 1455, Geneva, Switzerland. 27 pp.

Yu, J. Y., Hsiao, L. P. and 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 European conference on computer vision*, 818-833. Springer, Cham.

Zhao, M., Lin, S. J., and Vecchi, G. A. (2009). Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Climate*, 22, 6653–6678, doi:10.1175/2009JCLI3049.1.

Table 1: The datasets for training the deep CNN model

	Datasets	Sample size	Reference
1	CESM1.3 1850 pre-industrial control simulation	Year 0338-0519 (total: 182)	Chang et al. (2021)
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. (2021)
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. (2022)
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			

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	3×3, 3×3, 3×3, 3×3
Activation function for each convolutional filter	SELU, SELU,SELU,SELU
Mean pooling kernel size	2×2
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

Table 3. Pearson correlation coefficients between ensemble mean CNN based statistical-dynamical hybrid predictions and observed TC activity with various forecast lead month (LM). 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.

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

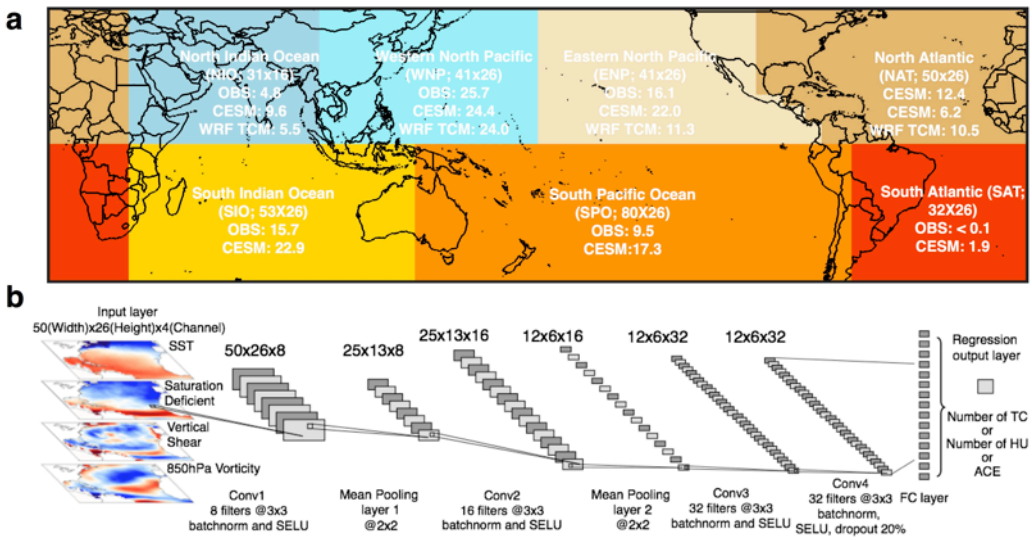

Figure 1: (a) Illustration of TC active ocean basin boundaries, including North Atlantic (NAT), eastern North Pacific (ENP), western North Pacific (WNP), north Indian Ocean (NIO), south Indian Ocean (SIO), south Pacific Ocean (SPO) and south Atlantic (SAT). The number in each TC basin shows the annual 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 850hPa 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°×2° 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. predictand).

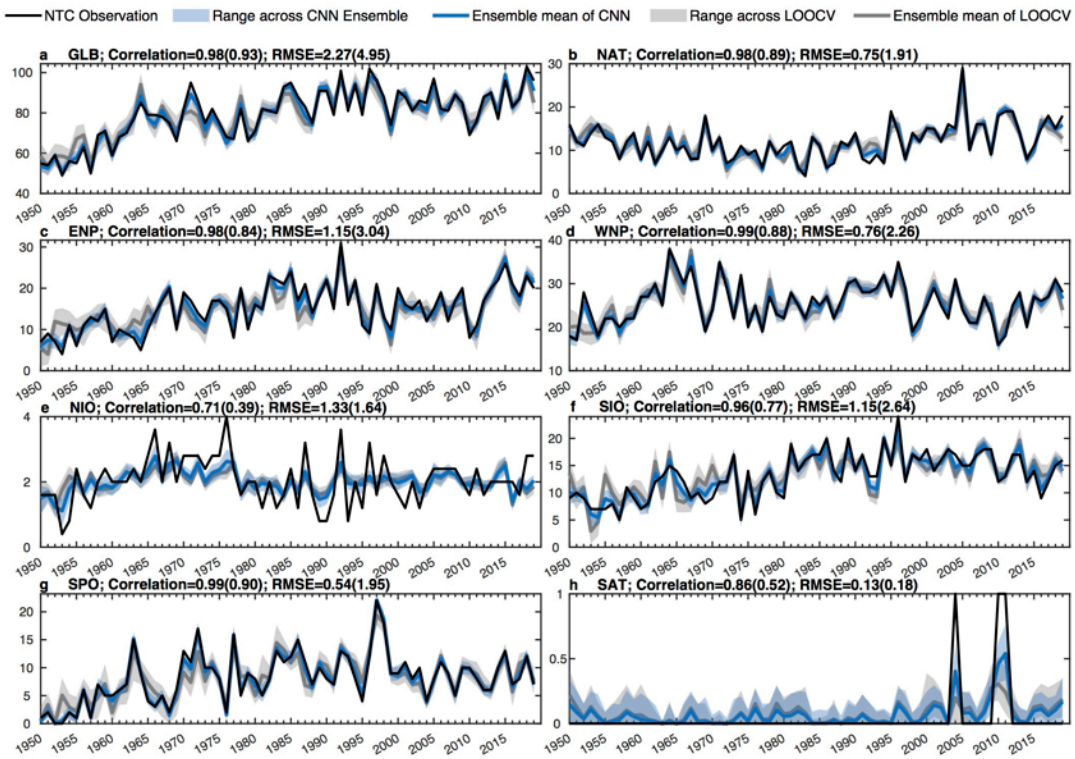

Figure 2: Time-series of seasonal mean number of TC (NTC) for (a) global (defined as the sum of individual ocean basins), (b) NAT, (c) ENP, (d), WNP, (e) NIO, (f) SIO, (g) SPO and (h) SAT. Observed NTC is shown in black line. Ensemble mean CNN emulated and LOOCV results are shown in blue and grey lines, 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.

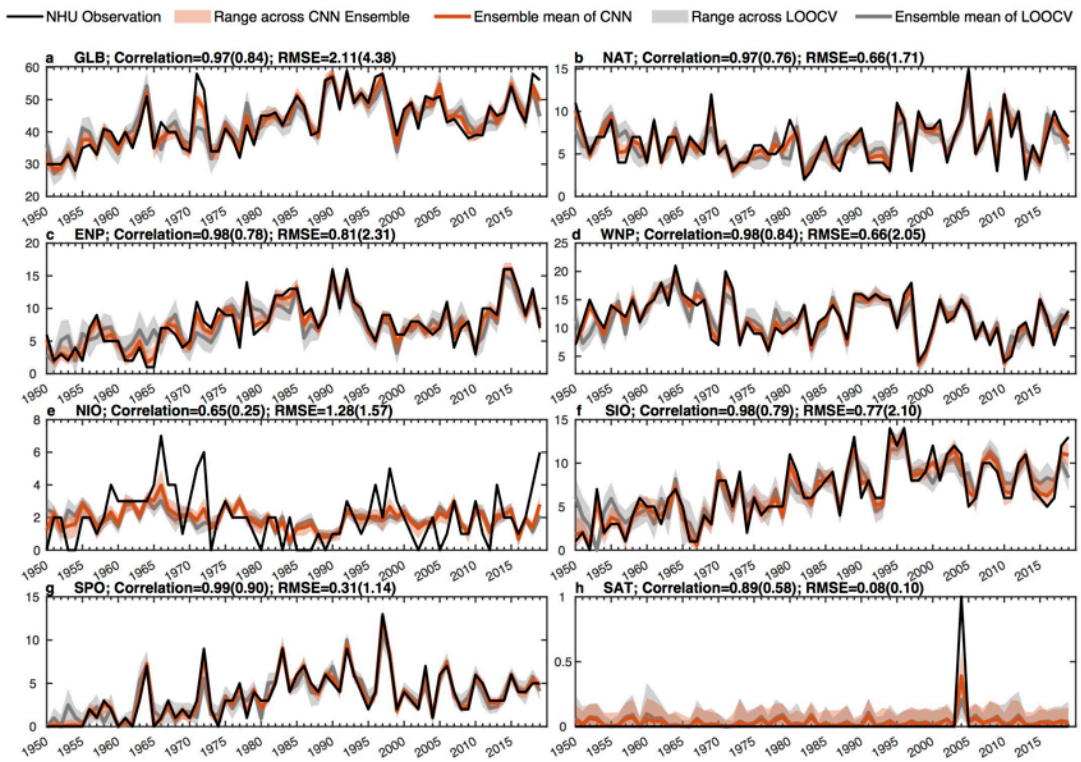

Figure 3: Similar as to Figure 2, but for seasonal mean number of hurricane (NHU; lifetime maximum wind speed ≥33m/s).

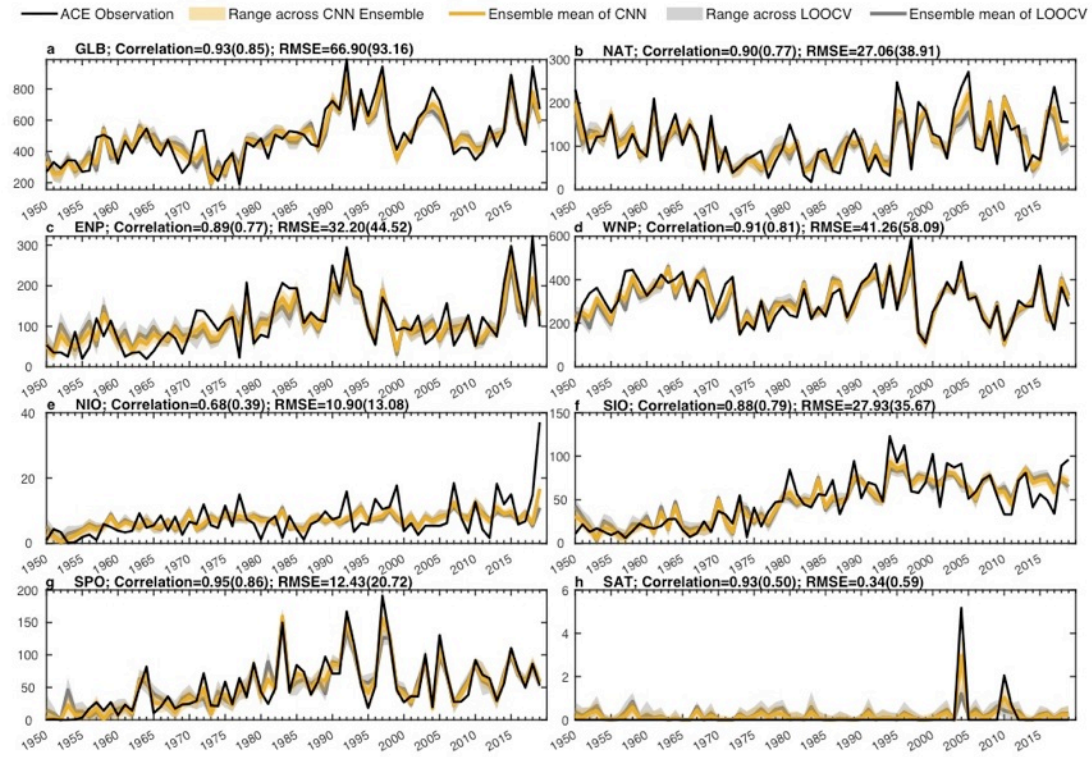

Figure 4: Similar as to Figure 2, but for seasonal mean accumulated cyclone energy (ACE).

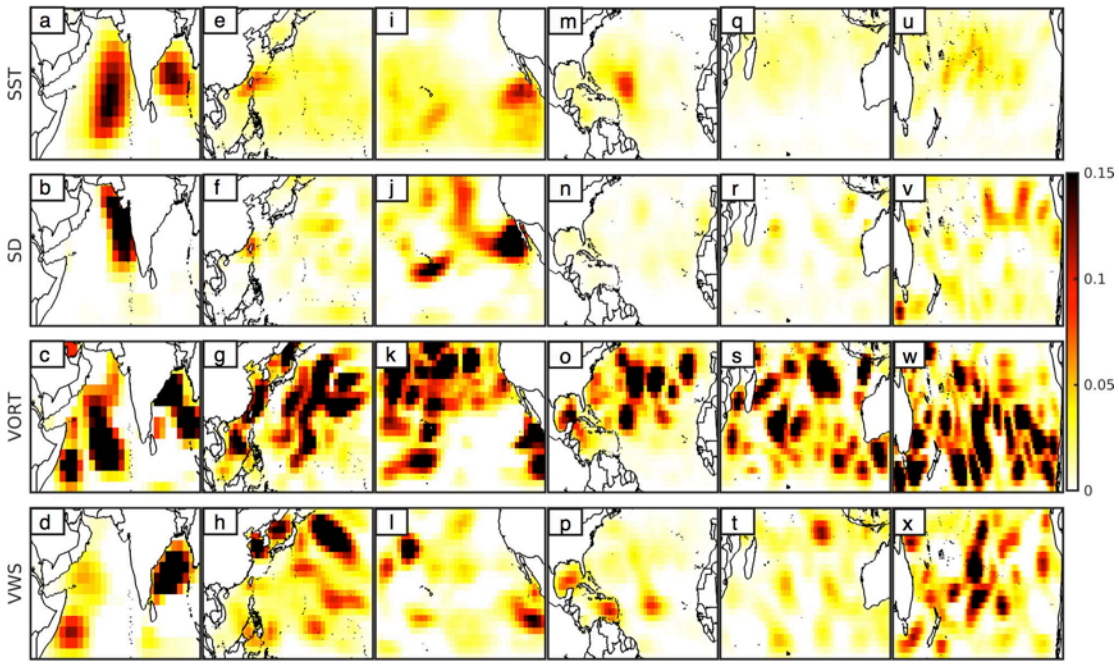

Figure 5: Occlusion sensitivity maps that highlight the relative importance in emulating seasonal NTC in different ocean basins. (a)-(d): Relative importance of SST, saturation deficit (SD), 850hPa vorticity and vertical wind shear in NIO, respectively. (e)-(h), (i)-(l), (m)-(p), (q)-(t), And (u)-(x) 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.

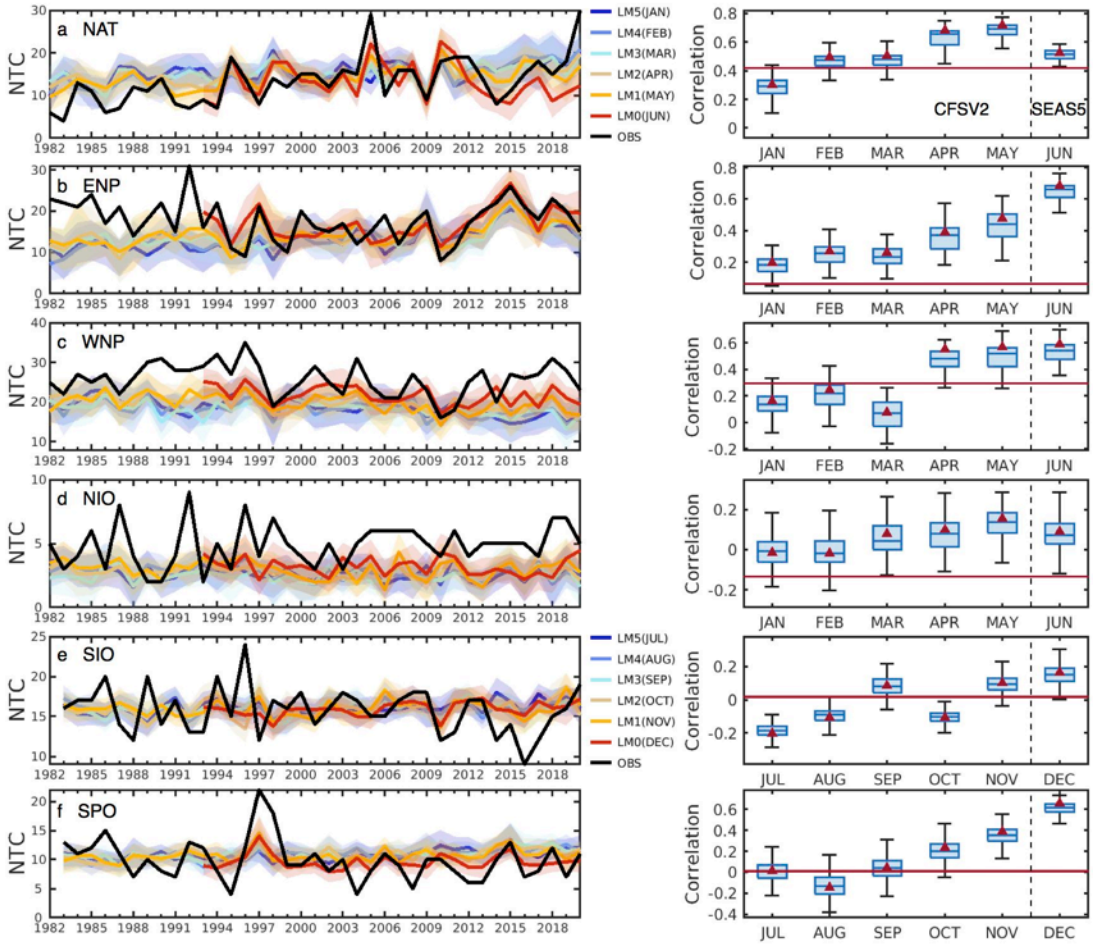

Figure 6: CNN based statistical-dynamical hybrid NTC seasonal predictions initialized at different lead month (LM). (a) Left: time-series of NAT NTC in observation (black) from 1982 to 2020, and CNN model predictions with the large-scale environmental conditions predicted form NCEP CFSv2 (LM1-5) and 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.

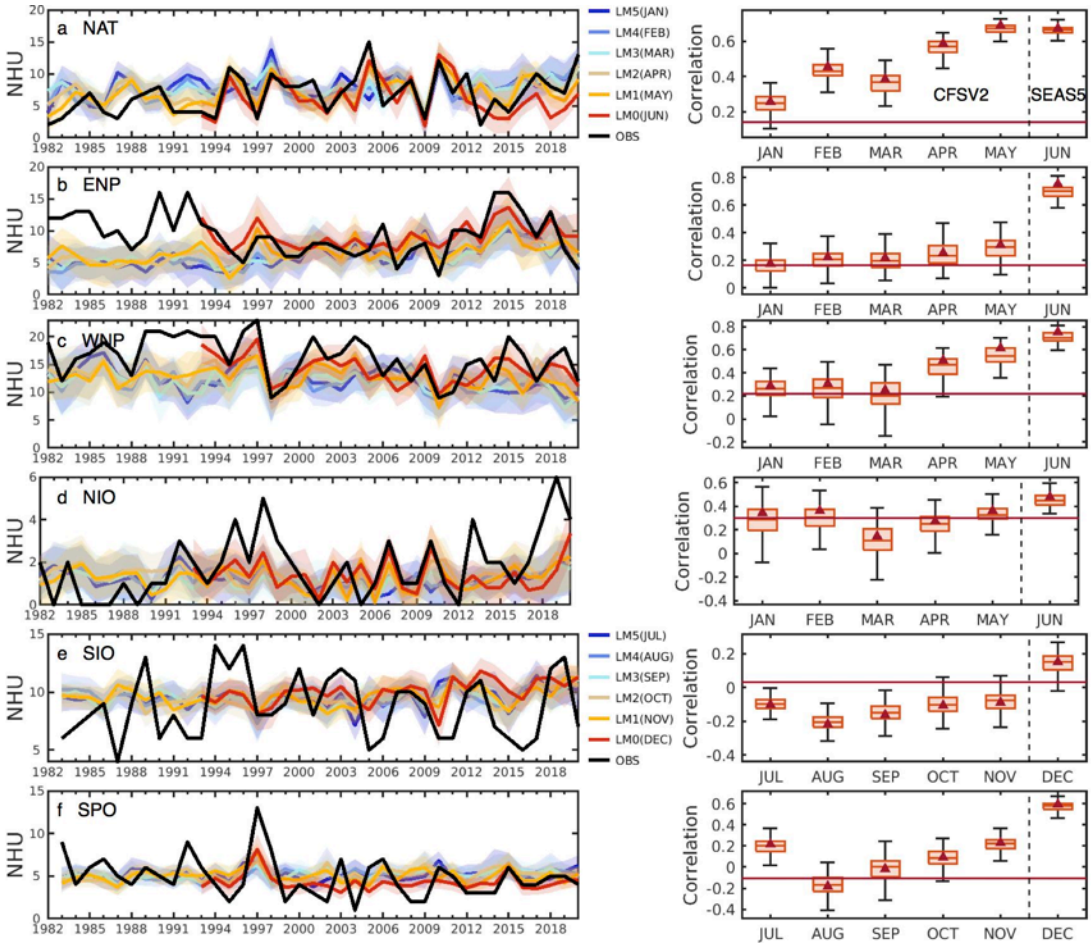

Figure 7: Similar to Figure 6, but for the NHU hybrid seasonal prediction skills.

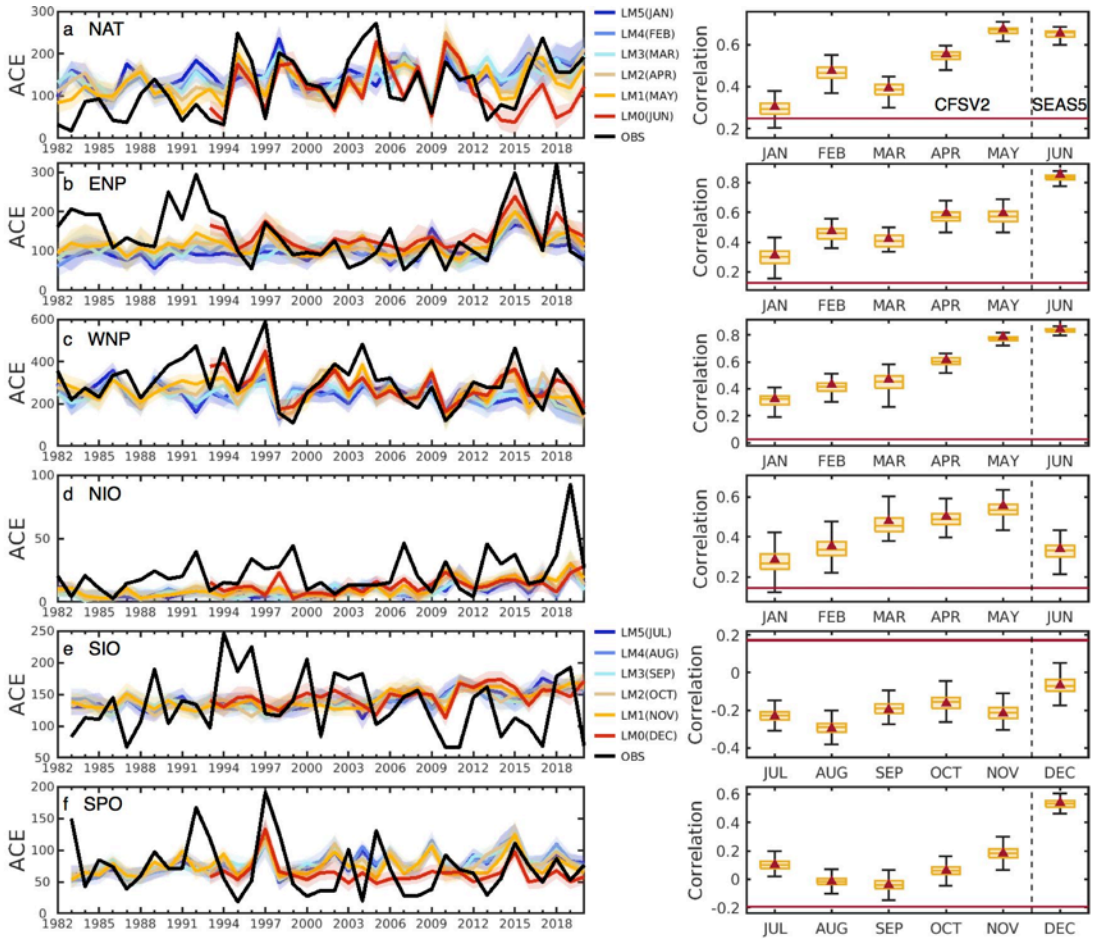

Figure 8: Similar to Figure 6, but for the ACE hybrid seasonal prediction skills.

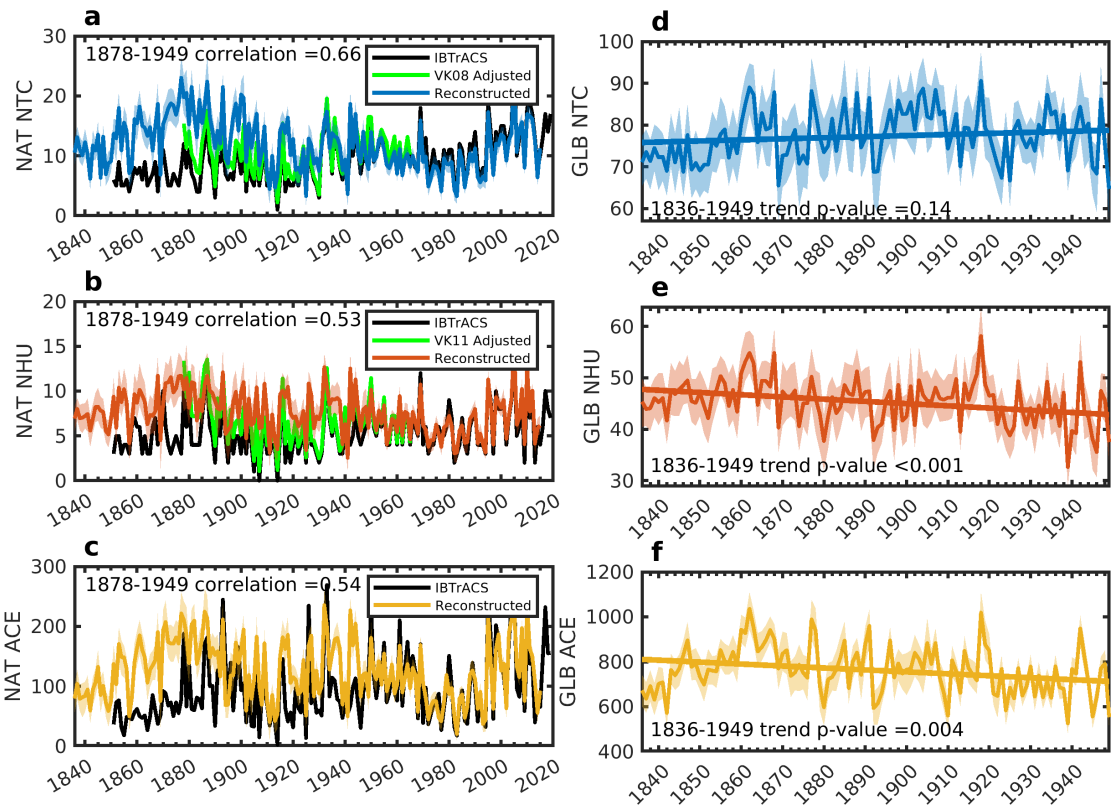

Figure 9: Time-series of NAT seasonal mean (a) NTC, (b) NHU, and (c) ACE from the IBTrACS observation in black, adjusted NTC and NHU observation based on Vecchi and Knutson (2008, 2011; VK08 and VK11 for short) in green, and CNN reconstruction based on the NOAA-CIRES-DOE 20th Century Reanalysis (20CRv3) 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.

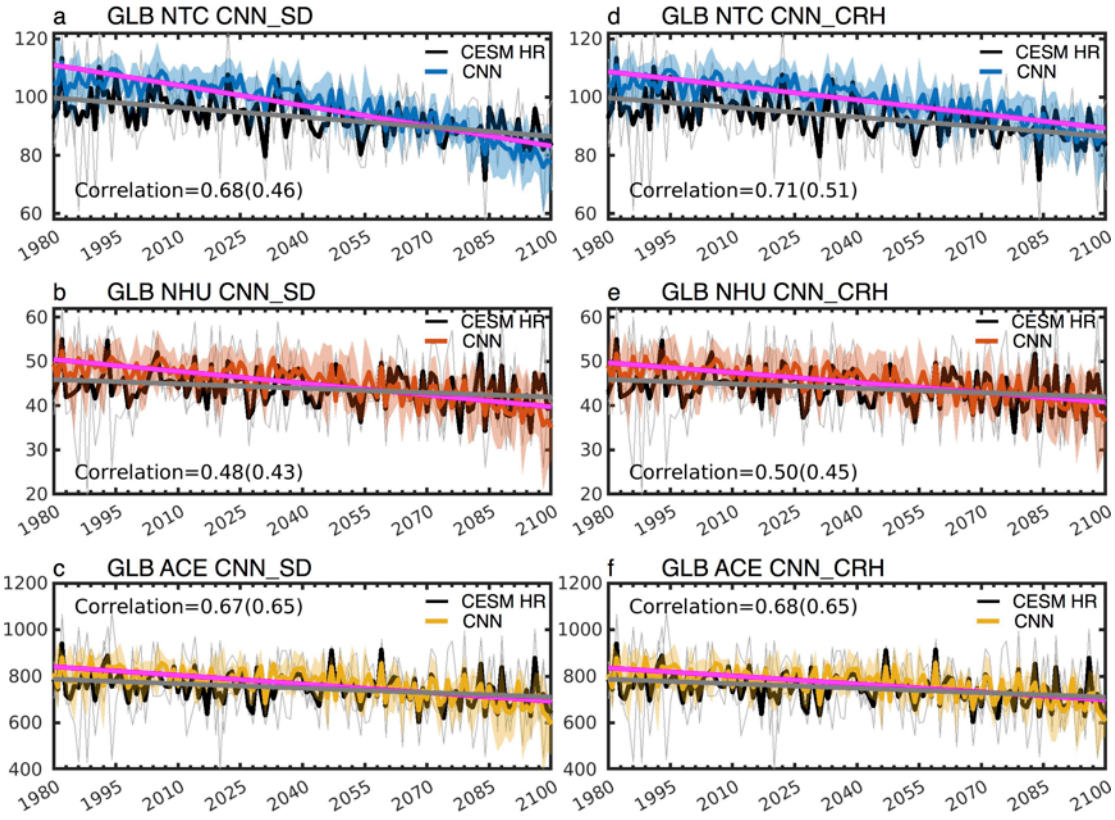

Figure 10: Time-series of 3-member ensemble mean CESM HR dynamically resolved TC activity (black) and CNN_SD (without transfer-learning, trained on climate model only) emulated global (a) NTC (blue), (b) NHU (red) and (c) ACE (yellow). Thin grey lines denote CESM HR 3 individual ensemble member, and color shadings indicate ranges across CNN ensembles. Pearson correlation coefficients between 1980-2100 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 deficient (CNN_SD), as the CNN moisture predictor variables.

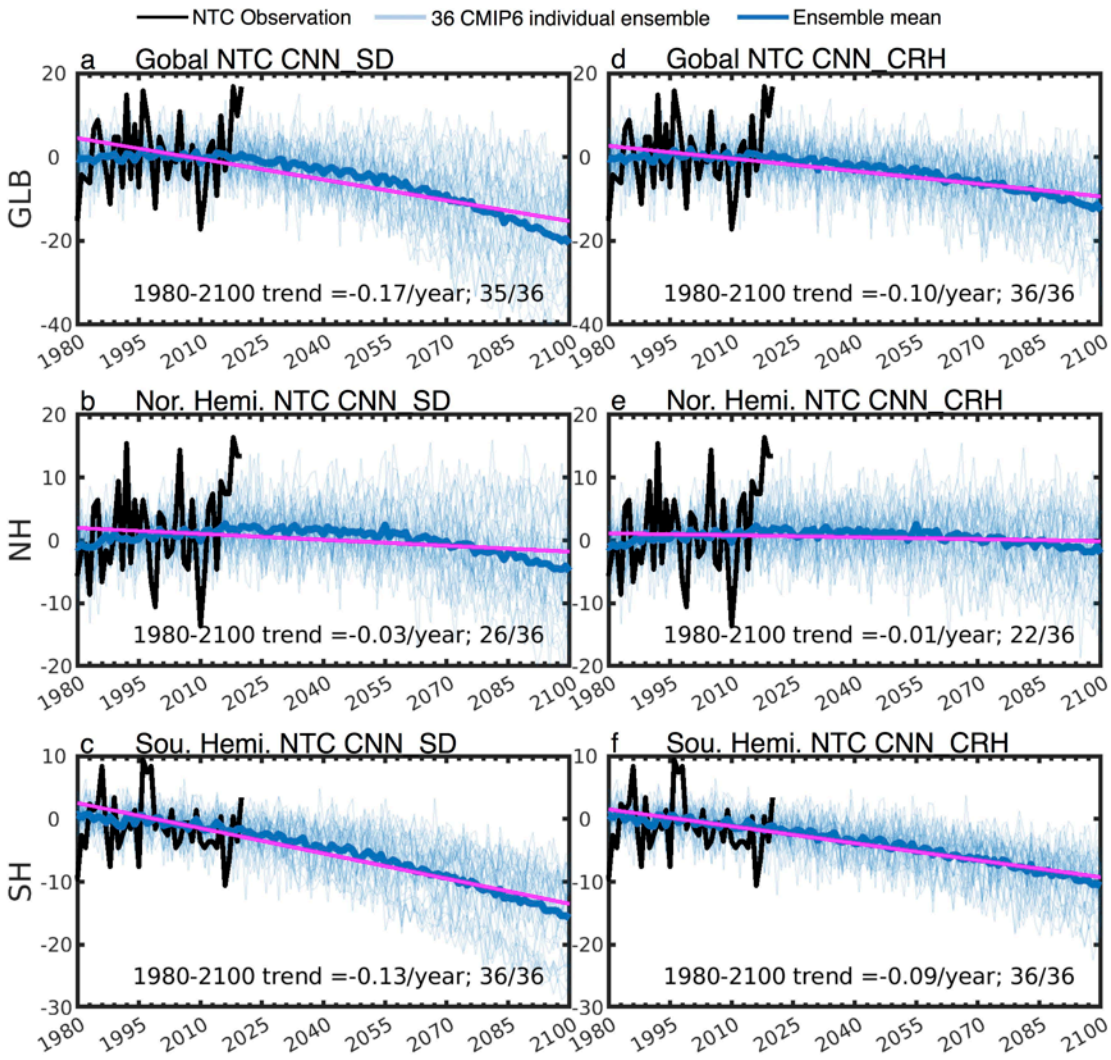

Figure 11: Time-series of CNN_SD emulated anomalous NTC for (a) global, (b) Northern Hemisphere and (c) Southern Hemisphere integration with the large-scale environmental conditions projected by 36 different CMIP6 model under historical forcing and shared socio-economic pathway 5-8.5 (SSP585). Anomalies 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.

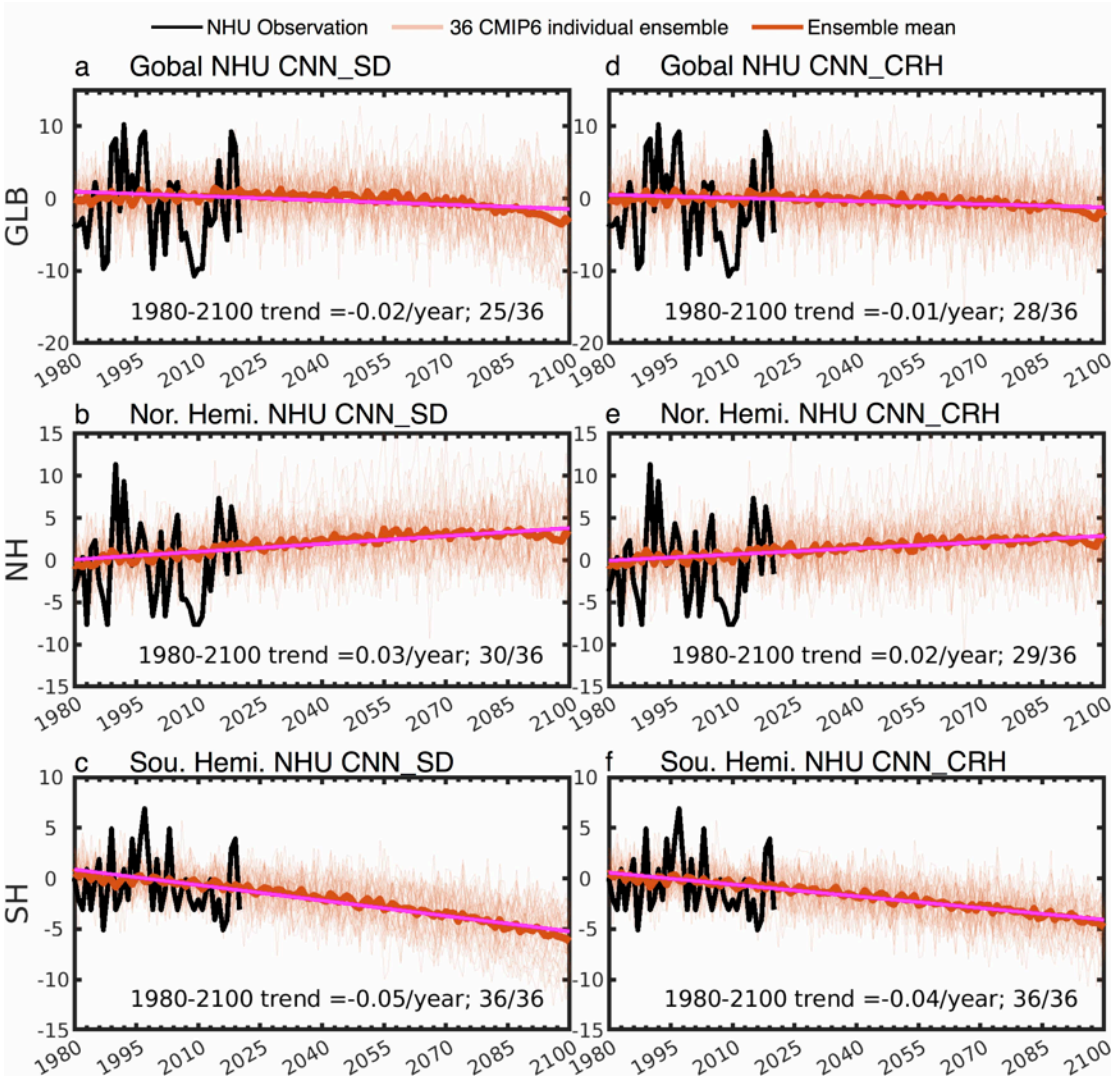

Figure 12: Similar to Figure 11, but for the NHU projections.

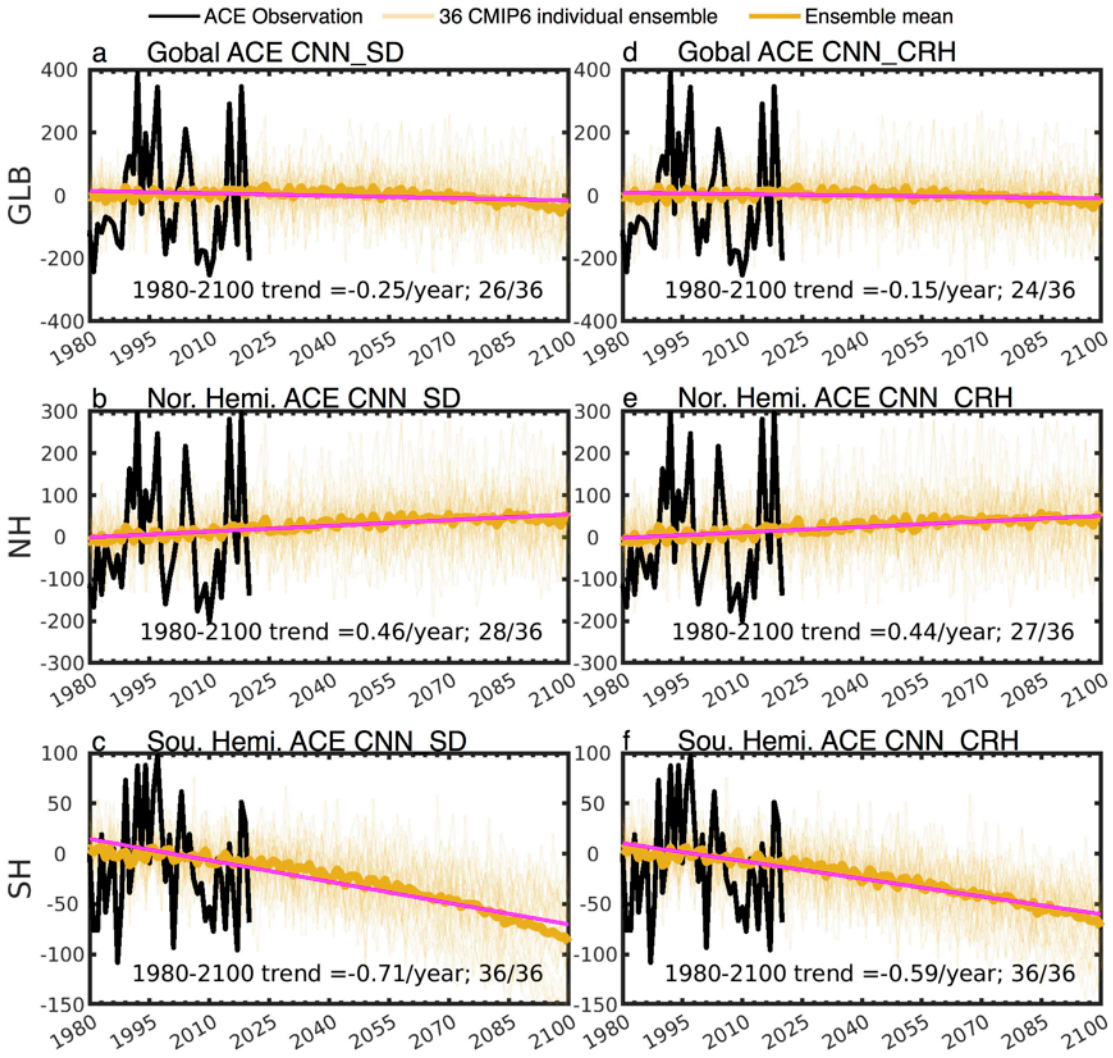

Figure 13: Similar to Figure 11, but for the ACE projections.