








RESEARCH ARTICLE

10.1029/2023MS004138

Special Collection:

The Energy Exascale Earth System Model (E3SM): Advancements in a Decade of Earth System Modeling

A Machine Learning Bias Correction on Large-Scale Environment of High-Impact Weather Systems in E3SM Atmosphere Model

Shixuan Zhang¹ , Bryce Harrop¹ , L. Ruby Leung¹ , Alexis-Tzianni Charalampopoulos², Benedikt Barthel Sorensen², Wenwei Xu¹ , and Themistoklis Sapsis² 

¹Pacific Northwest National Laboratory, Richland, WA, USA, ²Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Key Points:

- A machine learning approach with a long short-term memory network is used to bias correct climate simulations from the E3SM atmosphere model
- The approach effectively reduces biases in the simulated large-scale model states while preserving their responses to climate change
- Improving large-scale storm environment with bias correction facilitates modeling of high-impact weather systems and their future changes

Correspondence to:

L. R. Leung and S. Zhang,
Ruby.Leung@pnnl.gov;
Shixuan.Zhang@pnnl.gov

Citation:

Zhang, S., Harrop, B., Leung, L. R., Charalampopoulos, A.-T., Barthel Sorensen, B., Xu, W., & Sapsis, T. (2024). A machine learning bias correction on large-scale environment of high-impact weather systems in E3SM atmosphere model. *Journal of Advances in Modeling Earth Systems*, 16, e2023MS004138. <https://doi.org/10.1029/2023MS004138>

Received 20 NOV 2023

Accepted 11 JUL 2024

Author Contributions:**Conceptualization:** Shixuan Zhang,

L. Ruby Leung

Data curation: Shixuan Zhang,

Benedikt Barthel Sorensen

Formal analysis: Shixuan Zhang**Funding acquisition:** L. Ruby Leung,

Themistoklis Sapsis

Abstract Large-scale dynamical and thermodynamical processes are common environmental drivers of high-impact weather systems causing extreme weather events. However, such large-scale environmental conditions often display systematic biases in climate simulations, posing challenges to evaluating high-impact weather systems and extreme weather events. In this paper, a machine learning (ML) approach was employed to bias correct the large-scale wind, temperature, and humidity simulated by the atmospheric component of the Energy Exascale Earth System Model (E3SM) at $\sim 1^\circ$ resolution. The usefulness of the ML approach for extreme weather analysis was demonstrated with a focus on three high-impact weather systems, including tropical cyclones (TCs), extratropical cyclones (ETCs), and atmospheric rivers (ARs). We show that the ML model can effectively reduce climate bias in large-scale wind, temperature, and humidity while preserving their responses to imposed climate change perturbations. The bias correction is found to directly improve water vapor transport associated with ARs, and representations of thermodynamical flows associated with ETCs. When the bias-corrected large-scale winds are used to drive a synthetic TC track forecast model over the Atlantic basin, the resulting TC track density agrees better with that of the TC track model driven by observed winds. In addition, the ML model insignificantly interferes with the mean climate change signals of large-scale storm environments as well as the occurrence and intensity of three weather systems. This study suggests that the proposed ML approach can be used to improve the downscaling of extreme weather events by providing more realistic large-scale storm environments simulated by low-resolution climate models.

Plain Language Summary A machine learning model is employed to bias correct the large-scale dynamical and thermodynamical fields simulated by a low-resolution global climate model. The impact of the machine learning model on the large-scale storm environment associated with tropical cyclones (TCs), extratropical cyclones (ETCs), and atmospheric rivers (ARs) was evaluated. It is found that the ML bias correction can effectively reduce the mean climate biases in large-scale wind, temperature, and humidity fields associated with the three types of high-impact weather systems. For storms such as ETCs and ARs that can be partly resolved by the low-resolution climate models, the machine learning bias correction shows skills in improving the long-term statistics of these weather events. For storms such as TCs that can not be well resolved in the low-resolution climate models, the machine learning approach produces more realistic statistics of the tropical cyclone tracks by providing more realistic large-scale steering winds for downscaling approaches. By reducing model biases without affecting the climate change signals in large-scale storm environments derived from the low-resolution climate model simulations, machine learning bias correction has the potential to provide more reliable projections for assessing future changes in extreme weather events.

1. Introduction

General circulation models (GCMs) are the most common approach used in projecting climate change including future changes in high-impact weather systems such as atmospheric rivers (ARs), tropical cyclones (TCs), and extratropical cyclones (ETCs) which have substantial societal and economic impacts (e.g., Angéilil et al., 2016; Dai & Nie, 2022; Guan & Waliser, 2017; Merz et al., 2020; Moon et al., 2018; Seneviratne et al., 2012, 2023; Wehrli et al., 2018). However, achieving a proper representation of these weather systems for hazard assessment requires high spatial resolutions (on the order of a few kilometers) to realistically simulate the storm processes (e.g., convection), which is computationally demanding for global modeling (e.g., Kanada et al., 2017; Kanada & Wada, 2016; Kendon et al., 2014; Kitoh & Endo, 2016; Lucas-Picher et al., 2021; Mori et al., 2019; Nie

© 2024 Battelle Memorial Institute and The Author(s). 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 License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Investigation: Shixuan Zhang, Bryce Harrop
Methodology: Shixuan Zhang, Benedikt Barthel Sorensen
Project administration: L. Ruby Leung, Themistoklis Sapsis
Resources: Shixuan Zhang
Software: Shixuan Zhang, Benedikt Barthel Sorensen, Wenwei Xu
Supervision: L. Ruby Leung
Validation: Shixuan Zhang, Bryce Harrop
Visualization: Shixuan Zhang
Writing – original draft: Shixuan Zhang
Writing – review & editing: Shixuan Zhang, Bryce Harrop, L. Ruby Leung, Benedikt Barthel Sorensen, Themistoklis Sapsis

et al., 2018; Willison et al., 2013). Consequently, downscaling techniques have been widely used in combination with GCMs at low resolution (on the order of hundreds of kilometers) to yield important scientific insights on past and future changes of high-impact weather events (e.g., Emanuel, 2013; Emanuel et al., 2006; Knutson et al., 2013, 2019, 2020; Lee et al., 2020).

Downscaling approaches rely on the large-scale storm environments simulated by GCMs to project future changes of high-impact weather events through established statistical relationships (e.g., Balaguru et al., 2023; Colle et al., 2015; Dixon et al., 2016; Emanuel, 2013), or to provide boundary conditions for limited area or regional models to simulate the local climate and high-impact weather events (e.g., Fu et al., 2005; Giorgi et al., 1994; Gutowski et al., 2020). Accurate simulation of the large-scale storm environments by GCMs is therefore essential to achieve reliable downscaling to evaluate future changes in frequency, intensity, and characteristics of high-impact weather events. However, the large-scale storm environments governing regional to local-scale weather systems are often not well represented in the GCMs due to varying levels of systematic biases and uncertainties in representing smaller-scale processes that interact with the large-scale environments (e.g., Collins et al., 2013; Flato et al., 2013; Mueller & Seneviratne, 2014; Volosciuk et al., 2015; Zappa et al., 2013). As a result, GCM bias corrections have been an important research topic in downscaling studies and many bias correction methods have been developed to provide more reliable regional climate information (e.g., Christensen et al., 2008; Deque, 2007; François et al., 2020; Gudmundsson et al., 2012; Vaithinada Ayar et al., 2021; Vrac et al., 2012; Z. Xu & Yang, 2012; Z. Xu et al., 2021). Studies have demonstrated that correction of the GCM mean bias may improve dynamical downscaling of local-scale high-impact weather systems such as tropical cyclones over the North Atlantic Ocean (e.g., Bruyère et al., 2014; Done et al., 2015).

In recent years, advances in machine learning (ML) techniques have enabled the application of modern artificial neural network architectures in bias correction and statistical downscaling of GCMs (e.g., Fulton et al., 2023; Han et al., 2021; Moghim & Bras, 2017; Steininger et al., 2020; W. Xu et al., 2021; F. Wang & Tian, 2022). Several types of ML approaches have proven to successfully reduce spatial and temporal biases in GCMs (Fulton et al., 2023; F. Wang & Tian, 2022). In this paper, we introduce a long short-term memory neural network (LSTM) machine learning (ML) approach (Barthel Sorensen et al., 2024; Charalampopoulos et al., 2023) developed to bias correct the climate simulations produced by the U.S. Department of Energy (DOE) Energy Exascale Earth System Model (E3SM, Golaz et al., 2022). Specifically, the developed ML approach is employed to postprocess the large-scale wind (U, V), temperature (T), and humidity (Q) from long-term present-day and future climate simulations conducted with version 2 of the E3SM Atmosphere Model (EAMv2) at a horizontal grid spacing of $\sim 1^\circ$, driven by prescribed sea surface temperature and sea ice as lower boundary conditions. With an ultimate goal of improving the modeling of extreme weather events, we evaluate the impact of ML bias correction on large-scale storm environments and the long-term statistics of the high-impact weather systems simulated by EAMv2, with a focus on the ARs, TCs, and ETCs that frequently cause extreme weather events over the globe. Importantly, we also evaluate the impact of the ML bias correction on the responses of the three high-impact weather systems to future climate change. The goal of this study is to determine how well low-resolution climate models with ML bias correction may provide more skillful simulations of large-scale storm environment conditions than those without bias correction, thus providing a path to improved assessments of extreme weather events.

In Section 2, we introduce the experimental design and ML bias correction method. In Section 3 we evaluate the impacts of the ML model on the GCM model biases and the GCM projected climate change signals in large-scale model states (i.e. U, V, T, Q). Then the long-term statistics of ARs, TCs, and ETCs and their associated large-scale storm environment with and without ML bias correction are compared and evaluated for the present-day climate (1979–2014) (Section 4). Section 5 presents the impact of ML bias correction on the responses of ARs, TCs, and ETCs to climate change projected by pseudo global warming (PGW) simulations with and without postprocessing by the ML bias correction. Lastly, conclusions and discussions are given in Section 5.

2. Methodology

2.1. A Brief Overview of E3SM Atmosphere Model (EAM)

E3SM is a global Earth system model developed by the U.S. Department of Energy (Leung et al., 2020) with the first version released in 2018 (Golaz et al., 2019). This study uses the E3SM Atmosphere Model version 2 (EAMv2, Golaz et al., 2022) at standard resolution (also referred to as the “low-resolution” configuration). In

Table 1
List of Simulations for Present-Day and Future Climate Scenarios

Group	Short name	Time period	Bias correction	Scenario
1	CLIM	1979–2014	No	Present-day
1	SSP245	pseudo 1979–2014	No	Future climate
1	SSP585	pseudo 1979–2014	No	Future climate
2	ML (CLIM)	1979–2014	Yes	Present-day
2	ML (SSP245)	pseudo 1979–2014	Yes	Future climate
2	ML (SSP585)	pseudo 1979–2014	Yes	Future climate

Note. The three simulations in Group 1 were conducted with the coupled atmosphere-land components of EAMv2 with default “low-resolution” configuration at ~110 km grid spacing (see Golaz et al., 2022). All simulations were conducted with prescribed sea surface temperature (SST) and sea ice concentration (SIC) (see context in Section 2.2 for details). Group 2 consists of three simulations that apply ML bias corrections to the corresponding simulations in Group 1. Here “pseudo 1979–2014” refers to simulations driven by SST and SIC with climate perturbations corresponding to the difference between (2014–2060) and (1991–2010) added to the SST and SIC of the CLIM period of 1979–2014.

brief, EAMv2 uses separate computational grids for dynamics and column physics parameterizations. The dynamical core is configured with the “np4” cubed-sphere mesh with a horizontal resolution of ~110 km to solve the equations for large-scale dynamics and tracer transport (e.g., Dennis et al., 2012; Taylor & Fournier, 2010). The column parameterizations are run on a “pg2 grid” which shares the element grid with the dynamics but has a 2 × 2 subgrid of quadrilaterals for a total of four columns per element (e.g., Herrington et al., 2019; Lauritzen et al., 2018). The key subgrid-scale physical parameterizations in EAMv2 include representations of deep convection (G. J. Zhang & McFarlane, 1995), turbulence and shallow convection (Golaz et al., 2002; Larson et al., 2002), cloud microphysics (Gettelman & Morrison, 2015; Morrison & Gettelman, 2008; Y. Wang et al., 2014), aerosol life cycle (Liu et al., 2016; H. Wang et al., 2020), and shortwave and longwave radiation (Iacono et al., 2008; Mlawer et al., 1997). In addition, EAMv2 is interactively coupled with a land model (Oleson et al., 2013) that uses the same “pg2” grid for column parameterizations. EAMv2 is configured with 72 vertical layers, extending from the Earth’s surface to ~0.1 hPa (~64 km). The vertical grid spacing is uneven, with the layer thickness ranging typically from 20 to 100 m near the surface and up to 600 m near the model top.

2.2. Model Simulation

The simulations conducted for this study use prescribed sea surface temperatures and sea-ice concentrations, following the Atmospheric Model Intercomparison Project protocol (AMIP, Gates et al., 1999). Table 1 summarizes the key configurations of these E3SMv2 simulations.

The first group (Group 1) consists of one baseline simulation (i.e., “CLIM”) and two PGW simulations (i.e., “SSP245” and “SSP585”). CLIM is a present-day free-running simulation driven by prescribed observed monthly mean sea surface temperature (SST) and sea ice concentration (SIC) from the input4mip data sets (Reynolds et al., 2002) as the lower boundary conditions. Other external forcing data, including volcanic aerosols, solar variability, concentrations of greenhouse gases, and anthropogenic emissions of aerosols and their precursors, were prescribed following the World Climate Research Program Coupled Model Intercomparison Project-Phase 6 (CMIP6, Eyring et al., 2016; Feng et al., 2020; Hoesly et al., 2018). Emissions of aerosols and their precursor gases were set to the values of the year 2010 to represent the present-day condition. The simulation was run from 1 January 1978 to December 2014. The first year of model output was discarded as model spin-up, and the remaining 36 years of model output were used for analysis.

The SSP245 and SSP585 are two EAMv2 simulations conducted with the PGW approach that have been widely used in climate modeling (e.g., Schär et al., 1996; Xue et al., 2023). The PGW approach has been proven as a useful experiment strategy that enables targeted exploration of regional impacts from future climate change while avoiding the large ensembles typically required to address internal variability (Xue et al., 2023). In a PGW experiment, the large-scale changes in the climate system were imposed on a control climate simulation by modifying the boundary conditions. In this study, the imposed climate change perturbations were added to the boundary conditions of SST and SIC for CLIM that represent present-day climate conditions. Specifically,

- *SSP245*: the patterned SST and SIC perturbations (i.e. Δ) associated with the Shared Socioeconomic Pathways 2–4.5 scenario were added to the SST and SIC boundary conditions used in CLIM. Specifically, the monthly mean SST and SIC model outputs during the 1991–2010 (present-day) and 2041–2060 (future climate) periods were extracted from the coupled simulations conducted with 15 CMIP6 models (See Table A1). The climatological mean of the monthly SST and SIC over the two periods were then computed by averaging each quantity over the 20 years, and the Δ SST and Δ SIC were derived as the difference between the present-day and future climatological mean SSTs and SICs for each month and each grid point. Finally, the multi-model ensemble mean (MME) of Δ SST and Δ SIC were computed and added to the SST and SIC that were prescribed in the CLIM simulation as climate change perturbations for the pseudo SSP245 global warming experiment. We note that the monthly mean Δ SST and Δ SIC were added to each corresponding month to preserve the monthly and seasonal cycles of Δ SST and Δ SIC. Overall, the SSP245 perturbations correspond to a 1–2 K

warming in the annual mean SST, and 5–10% reduction in the annual mean SICs in most regions over the globe compared to the observed SST and sea ice in the present-day climate (see Figures A1a and A1b).

- *SSP585*: same as *SSP245*, except that the CMIP6 coupled simulations conducted under the Shared Socioeconomic Pathways 5–8.5 scenario during 2041–2060 were used to derive the SST and SIC perturbations for the PGW experiment. Compared to the *SSP245*, *SSP585* with an end-of-century forcing of 8.5 W m^{-2} instead of 4.5 W m^{-2} , results in stronger warming of SSTs and larger reductions of SICs (see Figures A1c and A1d).

We note that the two PGW simulations, *SSP245* and *SSP585*, used the same configurations as for *CLIM*, except with imposed climate change perturbations added to the boundary conditions of SST and SIC. Therefore, the differences between *SSP245* (*SSP585*) and *CLIM* were derived in this study to represent the model responses (e.g., large-scale storm environment and statistics of extreme events) to the climate-change-induced SST and SIC perturbations.

2.3. Machine Learning Bias Correction

The second group (Group 2) consists of three simulations that are the same as Group 1 except that the machine learning (ML) model was used to post-process the three simulations in Group 1 to bias correct the EAMv2 simulations of the present-day and future climates. In brief, the ML model employed for bias correction was proposed by Barthel Sorensen et al. (2024) in which a neural network (NN) operator acts on the coarse-resolution climate model simulation in a postprocessing manner. The NN operator was trained to learn a map between a coarse resolution EAMv2 historical simulation and the ERA5 reanalysis data (Hersbach et al., 2020) that represents the real atmosphere, allowing for the correction of large-scale dynamics of the EAMv2 model. A full description of the mathematical framework for the ML model is included in Appendix A2. Specifically, the ML training input was provided by a *nudged* simulation instead of using arbitrary coarse trajectories from the *free-running* simulations (i.e., *CLIM*). Here, the nudged simulation was conducted with the same configuration of EAMv2 as in *CLIM*, except that the model state (i.e. U, V, T, Q) was constrained toward the ERA5 reanalysis that has been remapped to the EAMv2 horizontal grid following S. Zhang et al. (2022) and Sun et al. (2019). With nudging the trajectory of the model state, which predominately obeys the dynamics of the coarse resolution EAMv2 model, is constrained from systematically as well as chaotically diverging from the ERA5 reanalysis. To counteract the artificial dissipation introduced by the nudging tendency, the spectrum of the nudged solution is corrected to match the free-running EAMv2 model (see Equations A2–A5 in Appendix A2). Training on this specific pair of trajectories (the spectrally corrected nudged EAMv2 solution and the ERA5 data) allows the network to learn a map from the chaotic attractor of the coarse resolution EAMv2 model to that of the reference data (i.e., ERA5 reanalysis) without being corrupted by chaotic divergence. At test time, the correction operator is then applied to the output of *free-running* EAMv2 simulation which is mapped into a trajectory residing in the attractor of the reference data. More detailed information for the development of the ML model can be found in Barthel Sorensen et al. (2024).

The performance of the ML model in correcting large-scale model states, including U, V, T, and Q, has been verified in Barthel Sorensen et al. (2024) using EAMv2 historical simulations from 1979 to 2014. It is found that the ML model can correct the coarse E3SM output to closely reflect the 36-year ERA5 statistics for all prognostic variables and significantly reduce their spatial biases (see Section 4 of their paper). In this study, the same ML model was applied to bias correct the 3-hourly instantaneous U, V, T, and Q model output at each grid point and 72 model levels from the three simulations in Group 1. In this study, the pair of EAMv2 simulations with and without ML bias correction during the historical period of 1979–2014, that is, *CLIM* and ML (*CLIM*) are the same as those verified in Barthel Sorensen et al. (2024). In addition, we further applied the ML bias correction to the PGW simulations for future climate, which is referred to as ML (*CLIM*), ML (*SSP245*), and ML (*SSP585*) in Table 1 and throughout the rest of this manuscript. Here, no new training was carried out for the bias correction of future scenarios. Therefore, the implied hypothesis here is that the bias correction model trained using the historical simulation can be applied to correct similar biases in future climate scenarios. Further analyses and evaluations in this regard will be presented in Section 3.2.

Overall, different from Barthel Sorensen et al. (2024) focusing on the development and assessment of the ML methodology, this paper focuses on presenting a detailed comparison between the Group 1 and Group 2

simulations to quantify the effects of the ML bias correction on the EAMv2 simulations of mean climate, extreme events, and climate change signals, which will be presented in Sections 3–4.

2.4. Analysis Strategy

As discussed above, the current ML bias correction model was only designed to correct the model state variables of U, V, T, and Q in the EAMv2 model output to improve the large-scale atmospheric flows. Therefore, the discussion in Sections 3–4 will rely on the metrics that can be derived from these model state variables through offline diagnostic equations or post-processing diagnostic package (see Appendix B). Other climate variables such as precipitation can be valuable for climate communities and impact studies, but they have not been included in our ML bias correction yet and therefore will not be discussed in this study. The limitations of our ML bias correction approach will be further discussed in Section 5.

In addition, the raw data sets from the simulations in Table 1 were the 3-hourly output of U, V, T, Q during 1979–2014 on the EAMv2 model grid as mentioned in Section 2.1. The data were post-processed into a $1.5^\circ \times 1.5^\circ$ normal lat-lon grid with the bilinear interpolation to facilitate the convenience of analysis in this study. A similar remapping process was also applied to the ERA5 reanalysis data for comparison and the derivation of model biases metrics in Sections 3–4. All quantities derived from U, V, T, and Q, as discussed in Appendix B, were diagnosed or processed using the 3-hourly data. The average of all 3-hourly data samples at each month was computed to form the monthly mean data sets during 1979–2014, and these monthly mean data sets will be further used to derive the seasonal and annual means for the diagnostic figures in Section 3. Moreover, the diagnostic metrics in Section 4 for TCs, ETCs, and ARs were generated with the algorithms and functions provided by the TempestExtremes package (P. A. Ullrich et al., 2021). The input to TempestExtreme is 3-hourly data derived from the U, V, T, and Q model output. The detailed configuration used by our study for TempestExtremes is described in Appendix B. A similar approach as mentioned above was used to process the high-frequency diagnostics from TempestExtreme into monthly, seasonal, or annual mean metrics for discussions in Section 4. Extra information for plotting details was also added to the captions under each figure in Sections 3–4.

Finally, the two-tailed Student's *t*-test was employed for generating the significance test on the 2-D plots (stipplings overlaid on the figures) in Sections 3 and 4. The significance test in our study was performed in a consistent manner, namely, we first processed the annual mean values at each grid point of the 2-D map over 36 years from 1979 to 2014, and then we calculated the sample average and variances from the 36 annual mean values and generated the probabilities for the Student-*t* test. Grid points with probability above the critical significance level of 0.1 or 0.05 (provided in the captions under the specific figure) are dotted to indicate that the means are statistically different at these grid points. While more rigorous statistical tests (e.g., Wilks, 2016) may be used to determine the significance of the differences, the Student's *t*-test results in combination with our understanding of the physical processes represented in E3SM are helpful for interpreting the differences between the results from EAMv2 simulations with and without ML bias correction.

3. Validation of Machine Learning (ML) Bias Correction

3.1. Bias Correction on Historical Simulation

A thorough evaluation of the performance of ML bias correction in improving the mean climate statistics was detailed in Barthel Sorensen et al. (2024). In this section, we present a brief evaluation of the ability of the ML approach to reduce the mean climate biases in EAMv2, which will be used to facilitate our discussion in the next subsection and Section 4. Figure 1 shows the zonal mean and annual mean biases in the zonal wind (U), temperature (T), and specific humidity (Q) fields of the EAMv2 simulations without and with ML bias correction. The metrics were derived by comparing CLIM and ML (CLIM) in Table 1 with the ERA5 reanalysis during the 1979–2014 period. The biases in the U, T, and Q fields from ML (CLIM) (Figures 1b, 1e, and 1h) are systematically smaller than those in CLIM (Figures 1a, 1d, and 1g) over most regions and vertical levels, meaning that the ML model leads to promising bias reduction in the mean climate fields simulated by EAMv2. We also note that the performance of ML bias correction varies with the quantities and spatial locations. Compared ML (CLIM) to CLIM, more significant bias corrections are seen in the temperature (Figure 1d versus Figure 1e) and humidity fields (Figure 1g versus Figure 1h), while relatively weaker bias reductions are seen in wind fields (Figure 1a versus Figure 1b). In addition, more promising bias corrections are seen in the near-surface levels relative to the upper model levels in most regions over the globe (also see Figure C1 in Appendix), especially for the wind fields

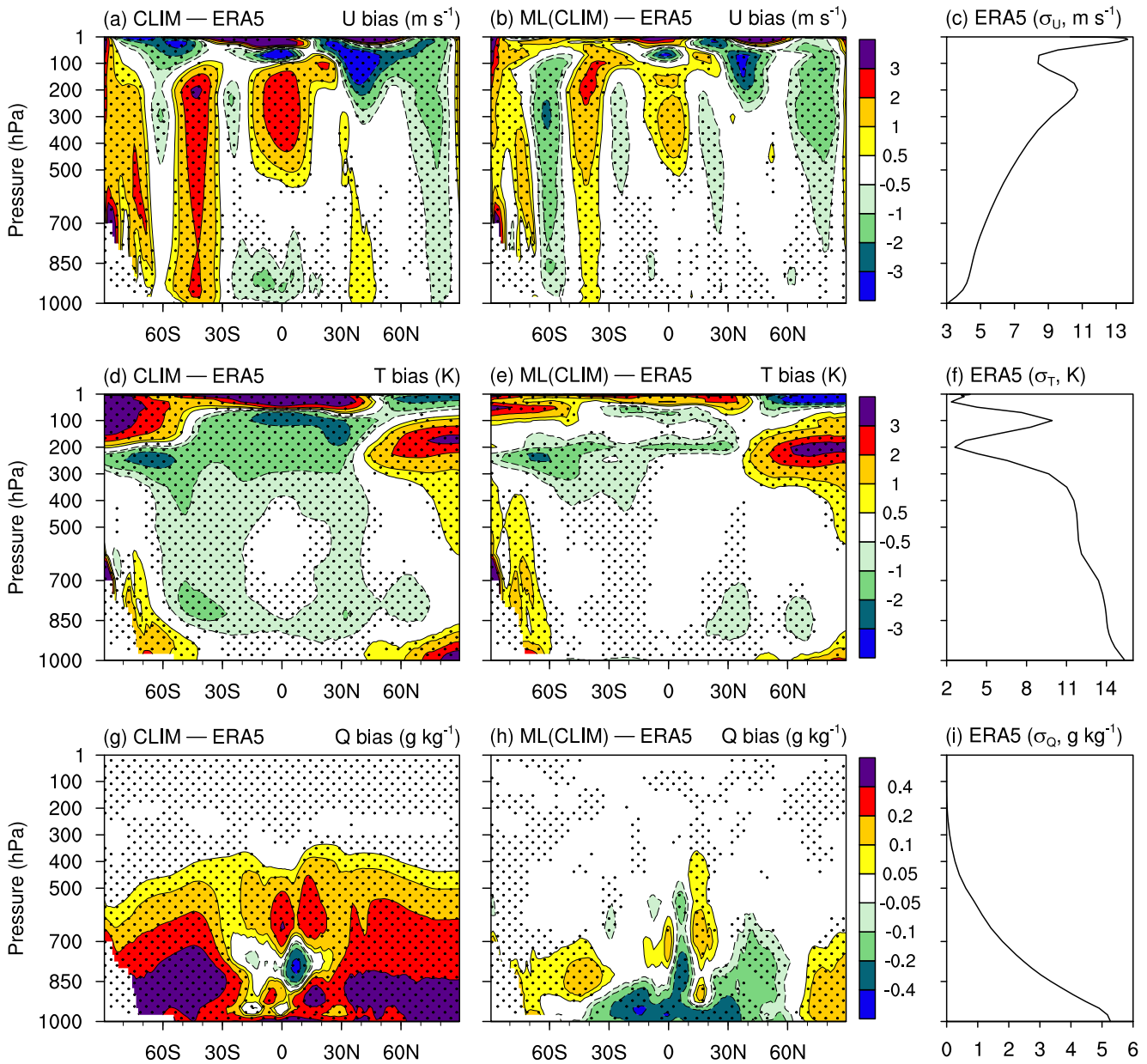


Figure 1. Zonal and annual mean model biases in the zonal wind (U, unit: m s^{-1} , panels a–b), temperature (T, unit: K, panels d–e), and water vapor mixing ratio (Q, unit: g kg^{-1} , panels g–h) averaged from 1979 to 2014. Shown are the results from free-running (i.e., CLIM, first column) and ML bias-corrected (i.e., ML (CLIM), second column) EAMv2 simulations. The biases are derived by comparing annual mean quantities between EAMv2 simulations with ERA5 reanalysis. The dotted region indicates that the differences are significant at a 95% confidence level from a Student's *t*-test. The third column shows the standard deviation of U, T, and Q at each pressure level from the ERA5 reanalysis. The log-linear algorithm is used to interpolate the EAMv2 data on the hybrid sigma-pressure level to the pressure level for comparison with ERA5 reanalysis. Details of the simulation setups can be found in Table 1.

(Figure 1a versus Figure 1b). Further investigation on the profiles of the global and annual mean standard deviation of U, T, and Q fields (Figures 1c, 1f, and 1i) from ERA5 reanalysis shows the clear vertical structure in the variability of these fields. In particular, the standard deviation of U (Figure 1c) in the upper troposphere (10–15 m s^{-1}) is about 3–5 times larger than those at near-surface levels (3–5 m s^{-1}). The differences in variability with altitude in these fields could limit the skill of the ML bias correction model at those levels.

For a more quantitative evaluation, Figure 2 further shows the mean biases and root-mean-square error (RMSE) of selected physical quantities from CLIM and ML (CLIM). The mean biases are normalized by the observed values, while the RMSE is normalized by the RMS of the observed values to demonstrate the relative rank of the biases in

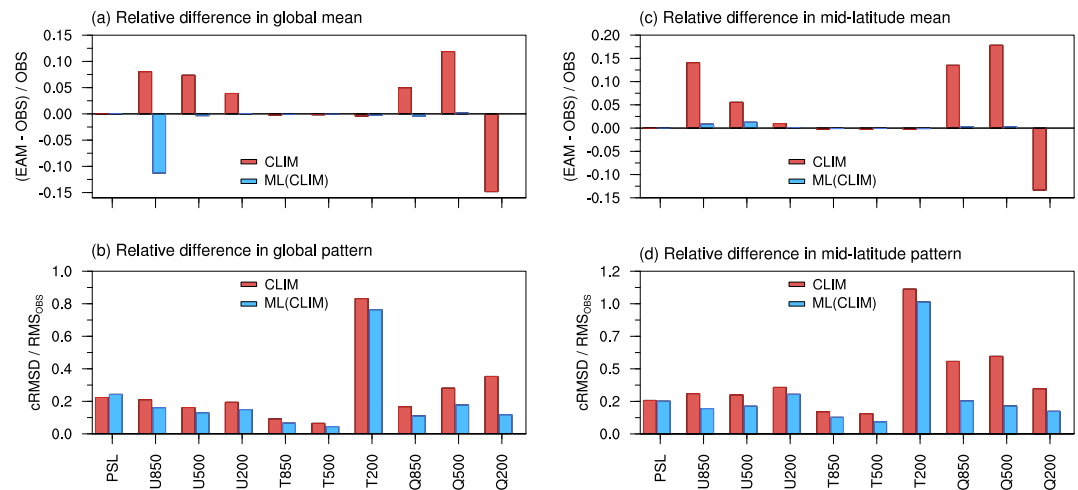


Figure 2. First row: mean biases in selected physical quantities averaged over the globe (panel a) and mid-latitude region ($[30\text{--}60\text{N}]$ and $[30\text{--}60\text{S}]$ latitude bands, panel c) from EAMv2 simulations without (i.e., “CLIM” in red) and with (i.e., “ML (CLIM)” in blue) ML bias correction, normalized by the observed value (i.e., ERA5 reanalysis); Second row: same as the first row, but for root-mean-square errors (RMSE) of anomaly patterns between EAMv2 simulations and observations, normalized by the root-mean-square (RMS) of the observed values. All metrics are calculated using the monthly mean model output and ERA5 reanalysis (i.e., observation) from 1979 to 2014. The y-axis shows the selected physical quantities, including surface pressure (PS, unit: hPa), sea level pressure (PSL, unit: hPa), zonal wind (U, unit: m/s) and temperature (T, unit: K) at bottom model level, 850-, 500- and 200-hPa pressure levels, as well as the specific humidity (Q, unit: g/kg) at 925-, 850-, 500- and 200-hPa pressure levels. The log-linear interpolation is used to regrid the EAM model output on the hybrid sigma-pressure level to pressure-level, and compared with ERA5 reanalysis on pressure levels. Details of the simulation setups can be found in Table 1.

different variables. We can see that the ML bias correction effectively reduces the biases by 10%–20% in the wind, temperature, and humidity fields over the globe (Figure 2a), especially in the mid-latitude regions. Meanwhile, the global and regional patterns of large-scale wind, temperature, and humidity fields are also systematically improved as evidenced by the RMSE metrics (Figures 2c and 2d). With these results, we conclude that the ML approach is capable of reducing biases in the mean climate simulated by EAMv2, which may produce a more realistic representation of the large-scale dynamics and thermodynamics fields associated with extreme events.

3.2. Impact of Bias Correction on Mean Climate Change Signals

As discussed in Section 2.2, the ML model trained with EAMv2 nudged simulations and ERA5 reanalysis during the historical period is directly used to correct the EAMv2 simulations for historical and future climate scenarios. Unlike many previous studies (e.g., Chen et al., 2020, 2021; Gutiérrez et al., 2019; Teutschbein & Seibert, 2012) which assumed that the bias correction was identical in present-day and future climates, our ML bias correction does not assume that the biases in climate model simulations are independent of the mean climate states. In other words, we do not assume that the error correction terms computed in the present-day climate can be simply added to the future runs. It is likely, however, that the large-scale structure and magnitude of model biases are very similar between the climates of the two time periods. If this is true, the bias correction should not significantly interfere with the large-scale climate change signals resulting from the imposed perturbations for the PGW simulations. We demonstrate that this is the case for our employed ML model by checking the features of the large-scale climate change signals before and after the ML bias correction in this section.

Figure 3 shows the spatial distribution of the temperature changes at 850-hPa due to the imposed climate change perturbations in SST and SIC for the SSP2-4.5 (top row) and SSP5-8.5 (bottom row) future scenarios. The patterns and magnitude of changes in near-surface temperature comparing the PGW and CLIM simulations (Figures 3a and 3c) are largely consistent with the PGW perturbations of SST (Figures A1a and A1c). These responses are expected due to the direct impact of the prescribed SST perturbations on the temperatures in the lower atmosphere. Compared with the climate change signals without bias correction (Figures 3a and 3c), ML bias correction applied to CLIM and the PGW simulations overall does not significantly modify the patterns and

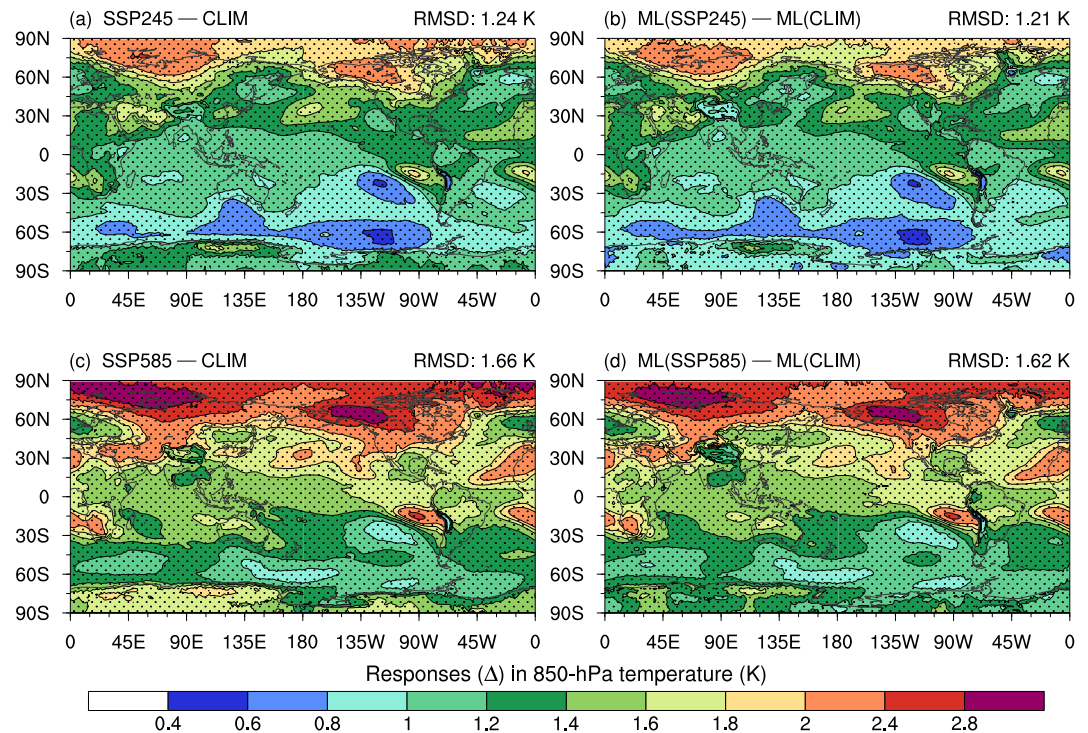


Figure 3. Differences of air temperature at 850-hPa (unit: K) between the present-day and pseudo global warming EAMv2 simulations averaged over the whole simulation period of 1979–2014. Shown are panels (a) SSP245—CLIM and (c) SSP585—CLIM for EAMv2 simulations without ML bias correction, as well as (b) ML (SSP245)—ML (CLIM) and (d) ML (SSP585)—ML (CLIM) for EAMv2 simulations with ML bias correction. The dotted regions indicate the differences are significant with a 95% confidence level. The SSP245 and SSP585 denote two PGW EAMv2 simulations with imposed climate change perturbations in sea surface temperature (SST) and sea-ice concentrations (SIC) derived from the CMIP6 historical simulations following SSP2-4.5 and SSP5-8.5 future scenarios, respectively. Detailed description on simulations can be found in Table 1.

magnitude of 850-hPa temperature responses in most regions over the globe (Figures 3b and 3d). An exception is the regions around 30° S and 30° N where a moderate modification on the magnitude of 850-hPa temperature changes is observed (e.g., Figure 3c versus Figure 3d). Further analysis indicates that the ML model applies corrections to the large cold temperature biases over these regions (see. Figures 1d–1f) during the historical period. Such corrections from the ML model are expected to take effect in the SSP245 and SSP585 simulations as well. Through non-linear processes in the atmosphere, the moderate adjustment of the climate change signals in these regions is not unexpected.

To further study the correction patterns between the present-day climate and future scenarios, the probability density functions (PDFs) of monthly temperature and humidity at 850 hPa during 1979–2014 are also plotted and shown in Figure 4. Consistent with the results in the previous section, we can see that the ML bias correction adjusts the PDF of CLIM in the present-day climate toward the PDF of ERA5 reanalysis data (dashed and solid blue lines vs. gray bars) for both near-surface temperature and humidity fields. For future climate scenarios, consistent with the imposed positive radiative forcing associated with the SSP scenarios, both EAMv2 simulations with and without ML bias correction predict warmer 850-hPa temperatures and higher specific humidity relative to present-day conditions (red lines vs. blue lines). Therefore, the physical climate change effects, that is, shifting toward warmer temperatures, are not erroneously removed as systematic biases by the ML model. More importantly, the differences between the ML-corrected future scenarios and present-day climate (solid blue and red lines) are very similar to those between the uncorrected data sets (dashed blue and red lines, see also Figure C2), suggesting that the ML bias correction does not significantly interfere with the PGW-induced climate change signals. In fact, we note that the PDFs of T850 and Q850 after ML bias correction (i.e., ML (SSP245) and ML (SSP585), solid red lines) are also quantitatively closer to the ERA5 global statistics (gray bars), compared with the uncorrected free-running simulations (i.e., SSP245 and SSP585, dashed red lines). This implies that the

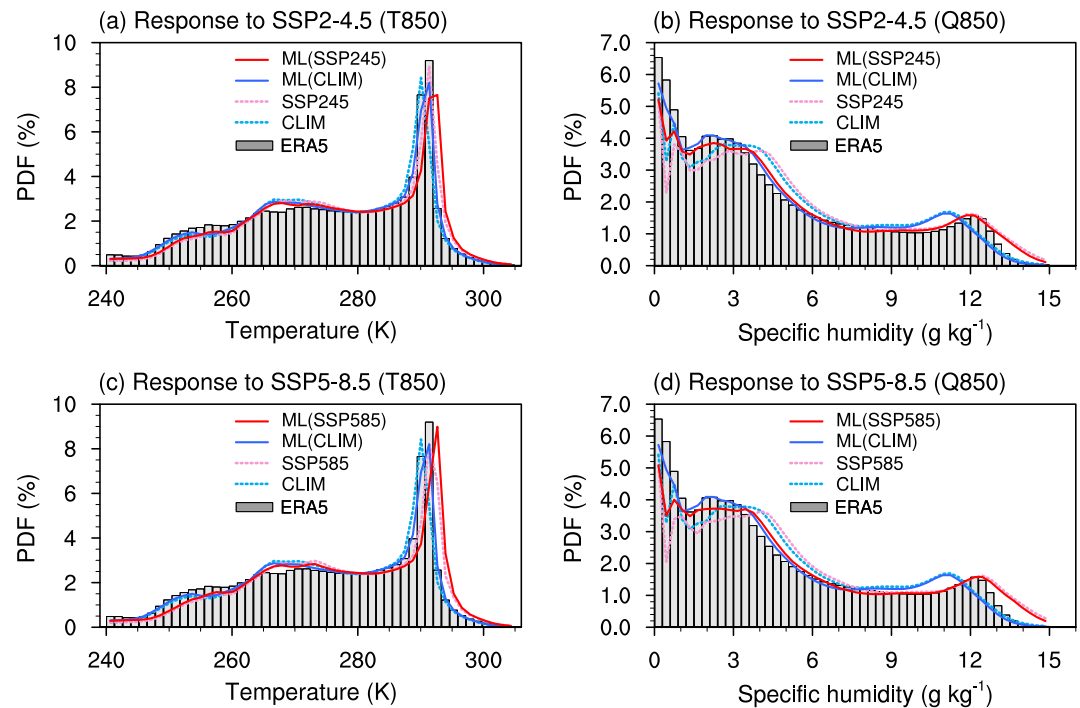


Figure 4. Long-term statistics for monthly mean air temperature (first column, unit: K) and specific humidity (second column, unit: g kg^{-1}) at 850-hPa pressure levels over the whole global domain during the simulation period of 1979–2014. Shown is the comparison among ERA5 reanalysis (gray bars), uncorrected (dashed lines), and ML-corrected (solid lines) EAMv2 simulation for present-day (blue lines) and future climate scenarios (red lines). The future climate simulations with SSP2-4.5 (top row) and SSP5-8.5 (bottom row) perturbations are shown in the top and bottom rows, respectively. The detailed descriptions on simulations can be found in Table 1.

ML bias correction constrains the EAMv2 simulations with corrections of the same sign and similar magnitude in both present-day and future climate simulations. For instance, the near-surface humidity correction shifts the right-side tail of the distribution by a similar amount to that seen in the present-day results (Figures 4b and 4d). Overall, it is encouraging that the ML model effectively reduces the model biases in large-scale dynamical and thermodynamical atmospheric conditions, while introducing insignificant interference on the climate change signals (or preserving the climate change signals imposed from external forcing).

4. Impact of Bias Correction on Statistics of Extreme Weather Events

In this section, we further discuss the value of ML bias correction for the study of extreme weather events and their underlying processes in EAMv2. We selected three types of high-impact weather systems to analyze: atmospheric rivers (ARs), ETCs, and tropical cyclones (TCs). The evaluation metrics rely on feature tracking using the TempestExtremes package (P. A. Ullrich et al., 2021), and are detailed in Appendix C. These three types of systems are of interest partly because they have the potential to generate extreme weather events, and they operate at spatial and temporal scales that are largely resolved (e.g., ARs), or under-resolved (e.g., ETCs and TCs) by low-resolution climate models (e.g., EAMv2) at $\sim 1^\circ$ horizontal resolution. With analyses of these events, we aim to demonstrate the value of the ML bias correction for improving simulations and projections of extreme weather events by typical GCMs.

4.1. Atmospheric Rivers (ARs)

Atmospheric rivers (ARs) are characterized by intense moisture transport, which, upon landfall, can produce precipitation that can be both beneficial and destructive (Payne et al., 2020). This is because the precipitation rate is proportional to the convergence of the zonal and meridional moisture transport (model output fields TUQ and TVQ, respectively) associated with ARs (e.g., Mo et al., 2021). The major features of ARs are reflected by TUQ

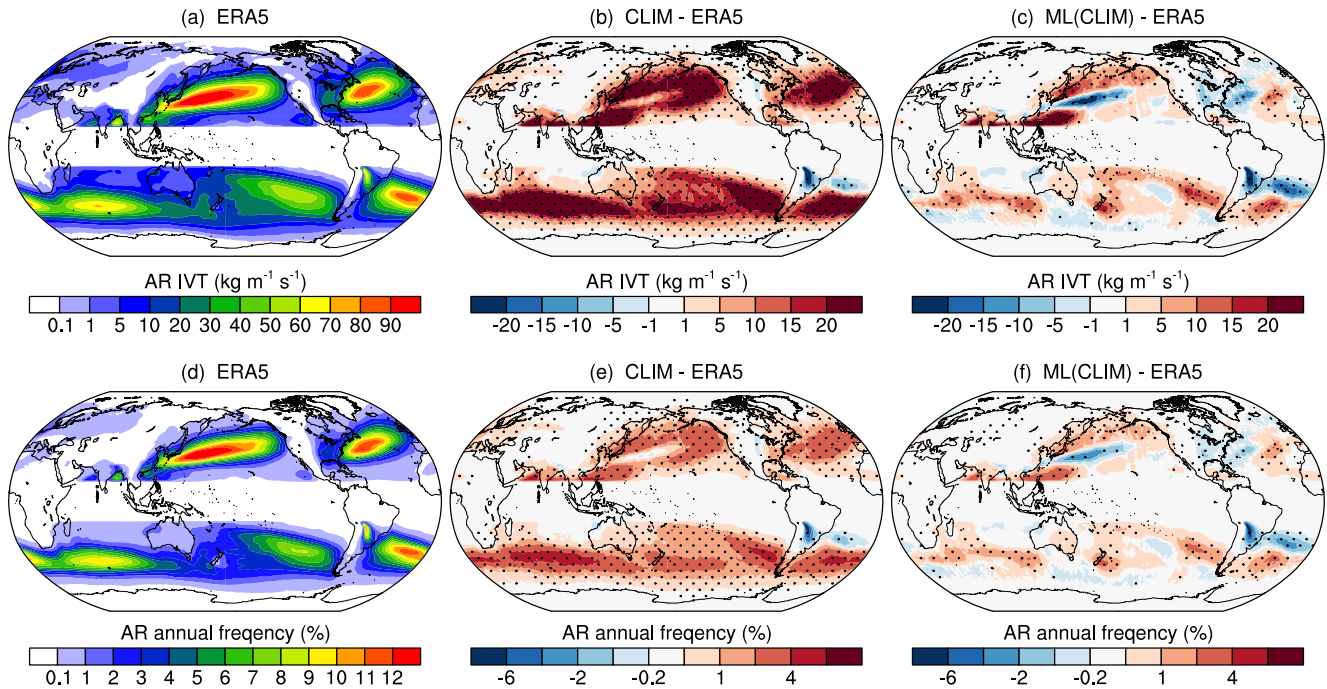


Figure 5. Top row: distribution of the vertically integrated horizontal water vapor transport (IVT, units $\text{kg m}^{-1} \text{s}^{-1}$) from ERA5 reanalysis averaged over all identified AR events at each grid point during 1979–2014 (panel a), and the model biases in the EAMv2 simulations without (i.e., CLIM, panel b) and with (i.e., ML (CLIM), panel c) ML bias correction. The AR events and composite IVT are tracked with TempestExtremes using the 6-hourly TUQ and TVQ data from ERA5 reanalysis and EAMv2 simulations; Bottom row: same as the top row, but for the annual AR occurrence frequency (unit: %) in ERA5 reanalysis (panel d) and the mean biases in CLIM (panel e) and ML (CLIM) (panel f). The annual frequency of AR is defined as the percentage of the number of time steps (6 hr) a grid point was part of an AR, divided by the total number of 6-hr time steps in each year during 1979–2014. The dotted region in panels (b–c and e–f) indicates that the differences between EAMv2 simulation and ERA5 are significant at a 95% significance level.

and TVQ, which, as in Equation B5, are directly linked to large-scale wind and specific humidity. Therefore, the AR systems are expected to be well simulated by EAMv2 at at $\sim 1^\circ$ resolution.

Despite the model being capable of resolving ARs, biases still exist in the simulated ARs. As reported in Kim et al. (2022), version 1 of the E3SM model overestimates the occurrence frequency and the water vapor transport of ARs. Therefore, it is worth checking whether EAMv2 with the ML bias correction can reduce these AR biases. Here, TempestExtremes is employed to track ARs using the 6-hourly TUQ and TVQ fields derived from the EAMv2 simulations with and without ML bias correction (see detailed tracking algorithm in Appendix B). The occurrence frequency of ARs and vertically integrated horizontal water vapor transport (IVT) associated with ARs are then calculated and shown in Figure 5. We can see that the significant overestimation of IVT in E3SM v1 still exists in the EAMv2 model (Figure 5b), meaning that spurious large moisture transport associated with ARs persists in both versions of the EAM model. These model biases can introduce biases in the AR-driven precipitation in model simulation as found in previous studies (Kim et al., 2022).

With the ML bias correction, the spurious large moisture transports associated with ARs are significantly reduced in the EAMv2 simulations (Figure 5c). The remaining model biases in the composite IVT field are statistically insignificant in most regions over the globe. Following Equation B5, such improvements in ARs are obtained because of the effective bias reductions by the ML model in both large-scale wind and humidity fields (Figures 1h and 1i). Consistently, the AR annual occurrence frequency also agrees better with ERA5 reanalysis after the ML bias correction (Figure 5e versus Figure 5f). Note that TempestExtremes uses the Laplacian of IVT instead of an IVT threshold for AR tracking. Therefore, biases in large-scale humidity on their own are not responsible for the AR frequency biases in Figure 5e. The improvements in the occurrence frequency of ARs suggest that the ML bias correction not only modifies the IVT value at each grid point but also inherently improves the gradient of IVT simulated by EAMv2.

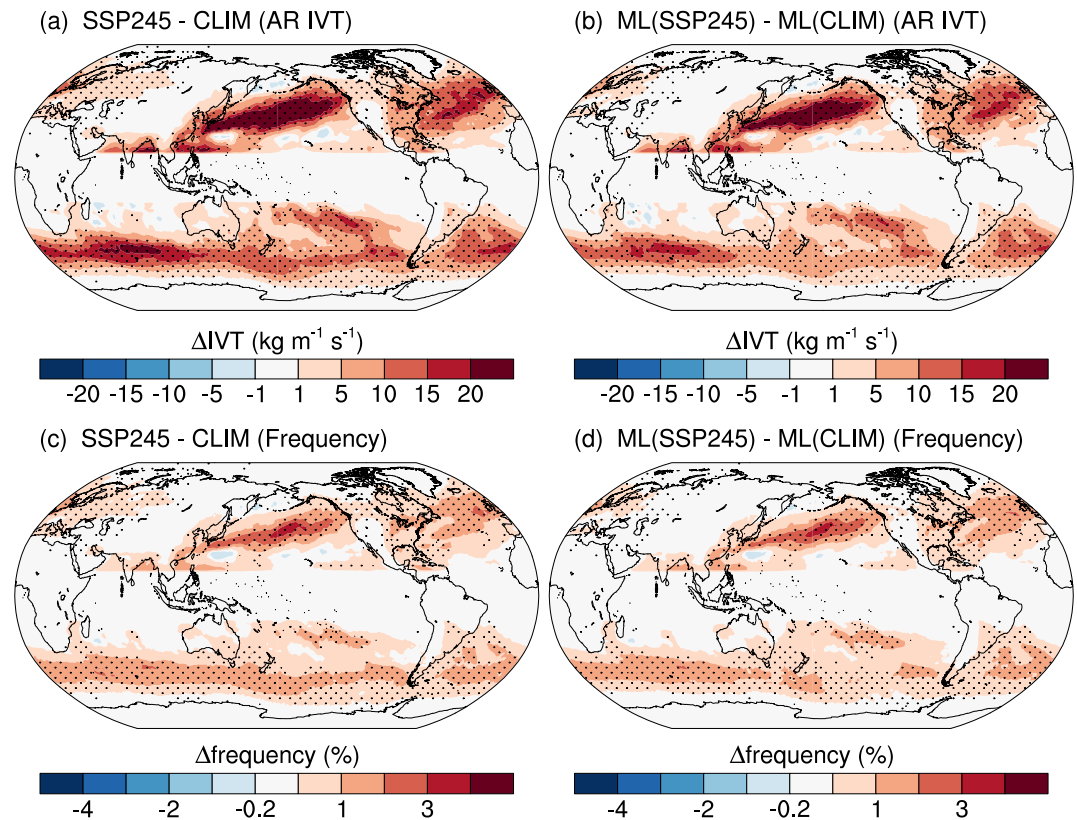


Figure 6. Changes of the vertically integrated horizontal water vapor transport (IVT, units $\text{kg m}^{-1} \text{s}^{-1}$, top row) and annual occurrence frequency of ARs (unit: %, bottom row) in the EAMv2 simulations with imposed climate change perturbations in sea-surface temperature (SST) and sea-ice concentration (SIC) for the SSP2-4.5 scenarios. Shown are the differences of SSP245—CLIM (panels a and c) and ML (SSP245)—ML (CLIM) (panels b and d) for the EAMv2 simulations without and with ML bias correction, respectively. The composite of IVT is derived using all AR events tracked by TempestExtremes during the present-day or pseudo-global warming period of 1979–2014. The definitions of the annual frequency of AR are the same as in Figure 5a. The dotted region in panels (b–c and e–f) indicates that the differences are significant at a 95% significance level.

Figure 6 further shows the responses of the IVT and occurrence frequency of ARs to the climate change perturbations in SSTs and SICs used in the PGW EAMv2 simulations. The differences between SSP245 and CLIM suggest an increase of the IVT (Figure 6a) and the occurrence of AR events (Figure 6c), which can be explained by the higher atmospheric humidity associated with warmer temperature as shown in Figure 4. However, circulation changes such as changes in the jet stream and subtropical high-pressure systems (Kim et al., 2022) likely also play a role since as noted earlier, ARs are tracked based on the Laplacian of IVT instead of an IVT threshold so an increase in atmospheric humidity alone does not translate to more frequent AR occurrence. These climate change signals are also seen in the same pair of simulations with the ML bias correction (ML (CLIM) and ML (SSP245)), indicating that the bias correction preserves the climate change signals. Compared with the free-running E3SM simulations, the ML bias correction results in a weaker increase of IVT and occurrence frequency of AR over the Northeast Pacific and Southern Ocean regions, which is likely due to the correction of ML on the overestimation of IVT in the E3SM model. Similar responses in the intensity (in terms of IVT) and occurrence frequency of ARs are also seen in the PGW simulations with stronger imposed climate changes in SST and SIC (i.e., SSP585), but the magnitudes of change in IVT and occurrence of ARs are more pronounced due to the stronger external forcing in SST and SIC (Figure C3). Again, the ML bias correction preserves the climate change signals, while adjusting the strength of the responses in IVT associated with ARs. Overall, the results in this section suggest that the ML bias correction reduces the systematic model biases in large-scale wind and humidity and improves the representation of ARs in EAMv2. Meanwhile, the ML bias correction does not have a significant impact on the climate change signals associated with ARs derived from the PGW simulations. By

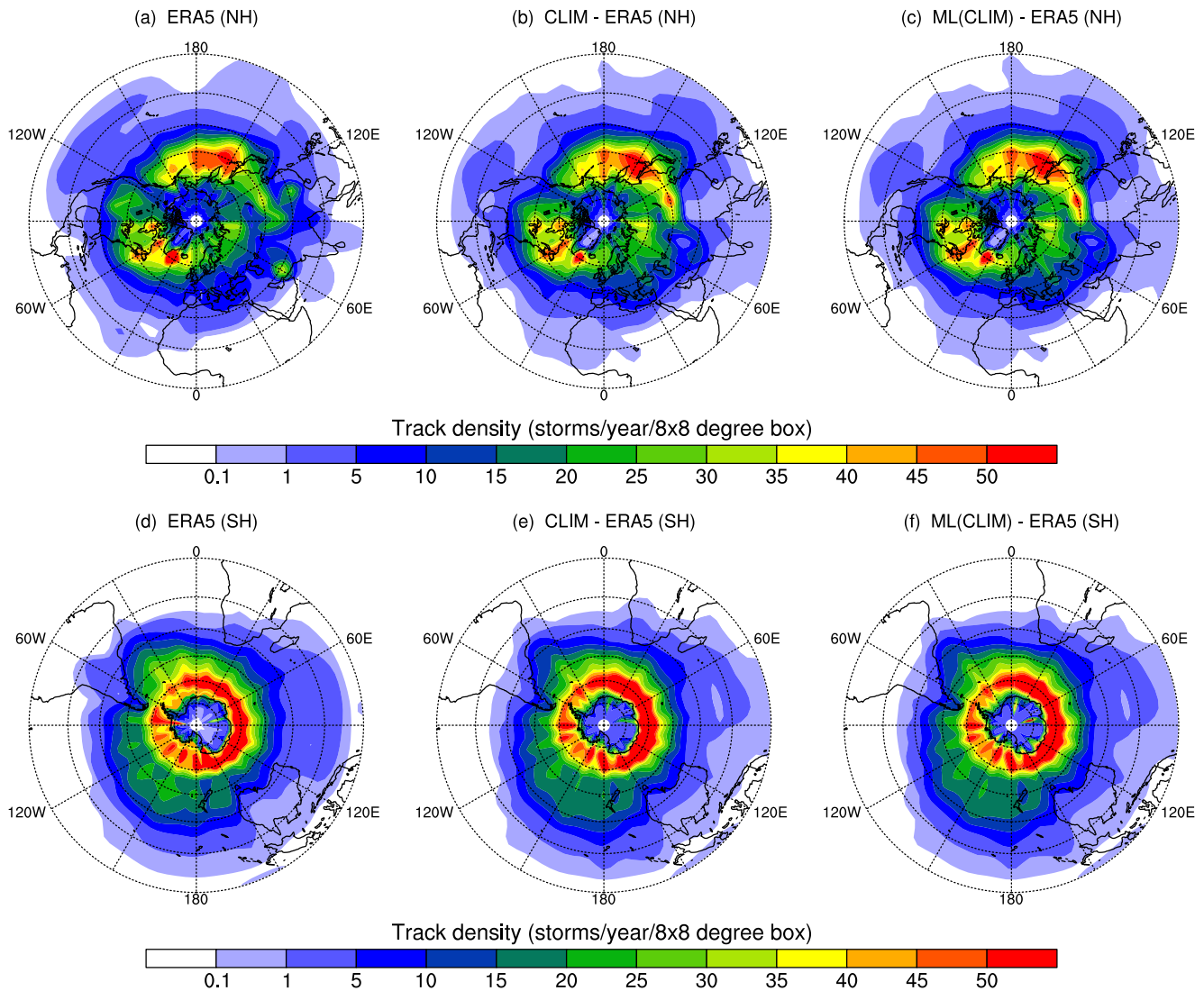


Figure 7. Track density maps for total annual ETCs over the Northern Hemisphere (NH, top row) and Southern Hemisphere (SH, bottom row) tracked in the ERA5 reanalysis (panels a, d) and EAMv2 climate simulations without (i.e., CLIM, panels b, e) and with (i.e., ML (CLIM), panels c, f) ML bias correction. The ETC events and composite IVT are tracked with the TempestExtremes using the 6-hourly mean sea level pressure (PSL) data from ERA5 reanalysis and EAMv2 simulations. The warm-core tropical-cyclone-like vortices were excluded during the feature tracking. The track densities shown are defined as the total number of time steps (6 hr) the ETCs passed the $8^\circ \times 8^\circ$ grid over the globe at each year from 1979 to 2014. Units are the number of 6-hourly ETC occurrences per $8^\circ \times 8^\circ$ grid box per year.

eliminating the systematic model biases, the bias-corrected AR environments provide more reliable information for downscaling ARs for assessing future changes in precipitation and flood hazards associated with ARs.

4.2. Extratropical Cyclones (ETCs)

Extra-tropical cyclones (ETCs) are a fundamental part of the atmospheric circulation that modulates the transportation of heat, moisture, and momentum in the mid-latitudes (Hawcroft et al., 2012; Sinclair et al., 2020). ARs discussed in the previous section are typically associated with a low-level jet stream ahead of the cold front of an ETC. The heavy precipitation and strong winds accompanying ETCs are known to cause extreme weather-induced damages in midlatitude regions such as Europe and North America (Fink et al., 2009; Hoskins & Hodges, 2002).

We begin our discussion by showing the track densities of ETCs that are tracked with TempestExtremes using the 6-hourly sea level pressure (PSL) model output in the two hemispheres (see detailed tracking algorithms in Appendix B). The annual ETC storm tracks over the Northern Hemisphere (NH) in ERA5 reanalysis (Figure 7a)

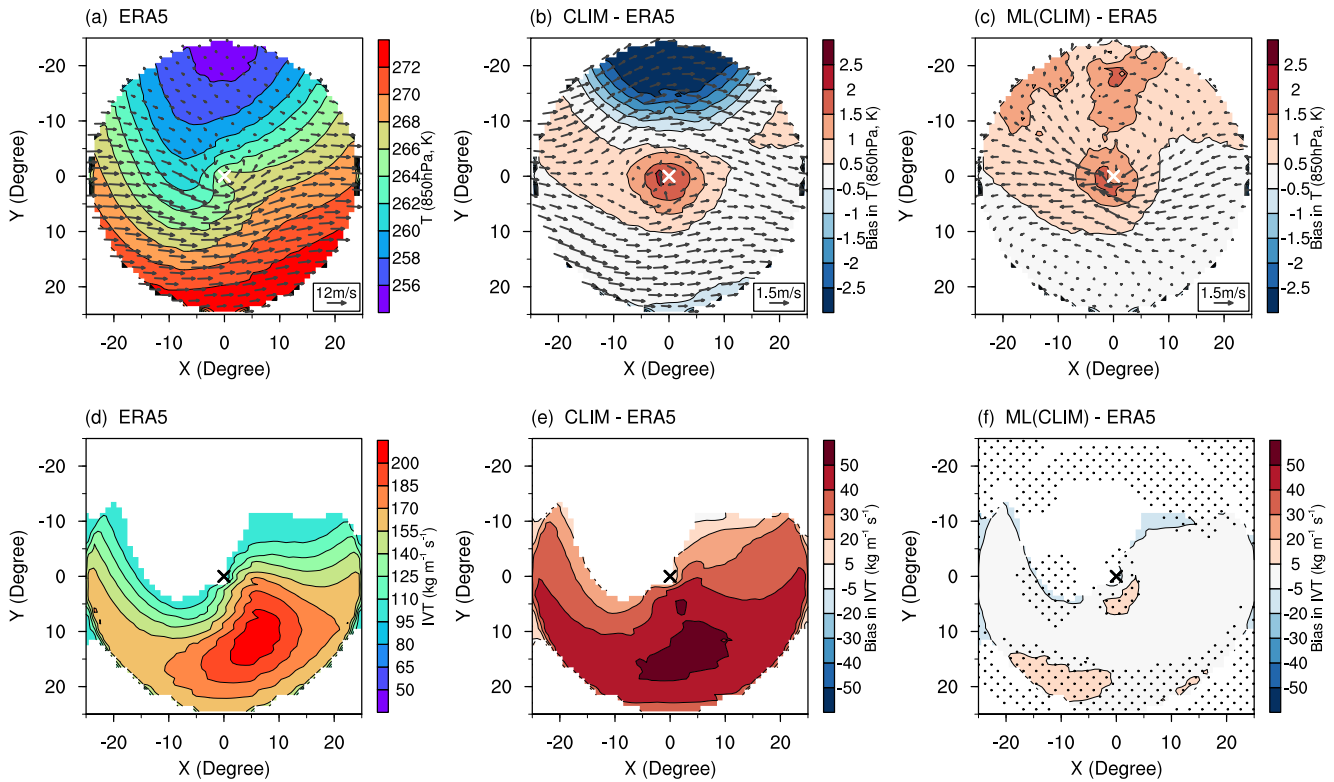


Figure 8. Composites of meteorological quantities centered on ETC storm center of all filtered storms with mean sea level pressure (PSL) less than or equal to 990-hPa in the ERA5 reanalysis (first column) and the differences between EAMv2 simulations and ERA5 reanalysis before (i.e., CLIM, second column) and after (i.e., ML (CLIM), third column) applying ML bias correction. The top row shows the composite of air temperature (contour, unit: K) and wind (vector, unit: m s^{-1}) at 850-hPa pressure level for (a) ERA5 reanalysis, (b) CLIM—ERA5 and (c) ML (CLIM)—ERA5; the bottom row shows the integrated vapor transport (IVT, unit: g kg^{-1}) for (d) ERA5 reanalysis, (e) CLIM—ERA5 and (f) ML (CLIM)—ERA5. All ETCs tracked with 6-hourly PSL fields during 1979–2014 are included in the composite by filtering out the storms with centered PSL >990-hPa. The white (panels a–c) and black (panels d–f) cross markers indicate the center of ETC storms.

show very clear high track densities over two regions separated by orographic features: the first region extends from high topography in East Asia (i.e., the Tibetan Plateau and the Altai–Sayan–Stonovoy range) into the western North Pacific, while the second region extends from the lee of the Rocky Mountains in North America, across the North Atlantic into Scandinavia and northern Russia. Different from the NH, the annual ETC tracks in the Southern Hemisphere (SH) show more continuous features with the highest track densities between 50°S and 70°S (Figure 7d). The low-resolution EAMv2 model produces a good representation of the observed spatial patterns of ETC track densities in both hemispheres (Figures 7b and 7e). Compared with the free-running simulations (i.e., CLIM, Figures 7b and 7e), no significant differences in the ETC track densities are seen in simulations with the ML bias correction (i.e., ML (CLIM), Figures 7c and 7f). The small differences between ML (CLIM) and CLIM are likely because the corrections by the ML model on wind and temperature do not lead to significant adjustments on the derived PSL (see Equation B1). As shown in Figure C4, the systematic model biases of PSL in CLIM are less than 2-hPa in most regions over the globe (Figure C4b). Therefore, we will not expect a strong correction from the ML model in these regions with small PSL biases (Figure C4c). However, the CLIM simulation indeed reveals large low-pressure biases in the southern ocean region (50°–70°S) (Figure C4b), which is co-located with the highest ETC track density region over the Southern Hemisphere (Figure 7d). In the same region, we indeed see a reduction of maximum PSL biases in ML (CLIM) due to the correction by the ML model (Figure C4c).

The reasonable representation of ETC occurrence in both CLIM and ML (CLIM) enables a fair comparison of the large-scale storm environment associated with ETCs through feature-oriented composite analyses. Figure 8 shows the composited 850-hPa temperature field, along with analogously calculated composites of 850-hPa wind vectors (Figures 8a–8c) and IVT (Figures 8d–8f). ERA5 reanalysis clearly showed the “warm conveyor belt” feature (e.g., Dettinger et al., 2015) with the advection of warm and moist air wrapping

cyclonically around the eastern side of the storm center (Figures 8a and 8d). The CLIM simulation without ML bias correction shows systematic warm biases and spurious large water vapor transport around the storm center, suggesting an overestimation of the advection of warm and moist air associated with the ETCs in the model (Figures 8b and 8e). A sizable cold temperature bias is also pronounced on the north side of the composite storm (Figure 8b). For ML (CLIM) with the ML bias correction, the biases in the temperature, wind, and vapor transport are reduced in the composite storms (Figures 8c and 8f), with more significant improvements in the IVT fields (Figure 8f). This results in a more realistic advection of temperature and humidity associated with the ETCs in the EAMv2 simulations. We also notice that the ML bias correction produces a weakening of the westward wind around the storm center (Figure 8d) compared with those in CLIM. This seems to be a physical response as the corrections on the warm temperature bias around the storm center reduce the west-to-east temperature gradient featured in Figure 8b, leading to an adjustment of wind according to the thermal wind balance relationship. The results likely suggest that the ML model indeed makes physically meaningful corrections on the EAMv2 simulations.

The responses of the ETC track densities to the future climate change in the Northern Hemisphere (NH) are shown in Figure 9. The results suggest that climate change with warmer sea surface temperature and lower sea-ice concentrations leads to a reduction of the storm track density around the Arctic (Figure 9a). Stronger signals in the ETC track density responses are observed in the simulations with the higher-emission climate change scenario of SSP5-85 (Figure 9c). In addition, the signals of the ETC track density responses to the climate change from ML bias corrected simulation (ML (CLIM) (Figures 9b and 9d), highly agree with those from the free-running EAMv2 simulations (CLIM, Figures 9a and 9b). Similar conclusions can be drawn for ETC track density responses over the Southern Hemisphere (SH, Figure C5).

For the responses of the ETC intensity, Figure 10 shows the changes of composited mean sea level pressure (PSL) fields in response to the imposed climate change perturbations from the SSP2-4.5 (first row) and SSP5-8.5 (bottom row) emission scenarios. The results suggest that global warming may favor more intense ETCs as there is a reduction of the storm center sea level pressure in the EAMv2 simulations of SSP2-4.5 (Figure 10a) and SSP5-8.5 (Figure 10c). Like tropical cyclones, the intensity of ETCs is not expected to be well simulated by the low-resolution EAMv2 model as it lacks the resolution to fully resolve the storm dynamics. However, analyses of the composite large-scale storm environment suggest that global warming leads to warmer temperatures on the west side of the storm and increased water vapor transport on the east side of the storm (contour and shading in Figures 11a and 11c). Meanwhile, there is an enhanced cyclonic circulation in the boundary layer regions (i.e., 850-hPa) due to climate change (vectors in Figures 11a and 11c). These changes in the storm environment suggest enhanced warm and moist air advection wrapping cyclonically around the storm center, favoring the development and formation of more intense ETCs, consistent with the sea level pressure changes.

The climate change signals in the ETC intensity and storm environment from the EAMv2 simulations with ML bias correction largely agree with those in the free-running PGW simulations for both SSP2-4.5 and SSP5-8.5. This was concluded by viewing the responses of ETC-composited PSL (Figure 10) and large-scale circulation and moisture transport at 850-hPa pressure level (Figure 11). Again, this suggests that the ML bias correction preserves the climate change signals associated with ETCs. However, different from the ETC track density, the ML bias correction shows noticeable impacts on the magnitude of the responses of the ETC intensity and storm environment to climate change. Specifically, the responses of PSL to the perturbations of SSP2-4.5 and SSP5-8.5 become weaker after applying the ML bias correction (Figures 10b and 10d), compared with those in free-running simulations (Figures 10a and 10c). Accordingly, a weaker change of the temperature and water vapor transport is also evidenced in the simulations with ML bias correction (Figures 11b and 11d versus Figures 11a and 11c). The modifications of the ML bias correction on the climate change signals are likely reasonable. As shown in Figure C1, the EAMv2 model significantly overestimates the humidity over the ETC active regions (e.g., 50–70°S), which likely explains the significant overestimation of the composite water vapor transport in Figure 8e. The effective corrections by the ML model on these biases tend to reduce the model-simulated humidity and water vapor transport associated with ETCs. Such corrections from the ML model are also expected to take effect in the ML (SSP245) and ML (SSP585) simulations. Also noteworthy is that the ML bias correction preserves the physical relationships between the storm environment and storm intensity, both showing smaller changes in the future compared to the changes simulated without bias correction.

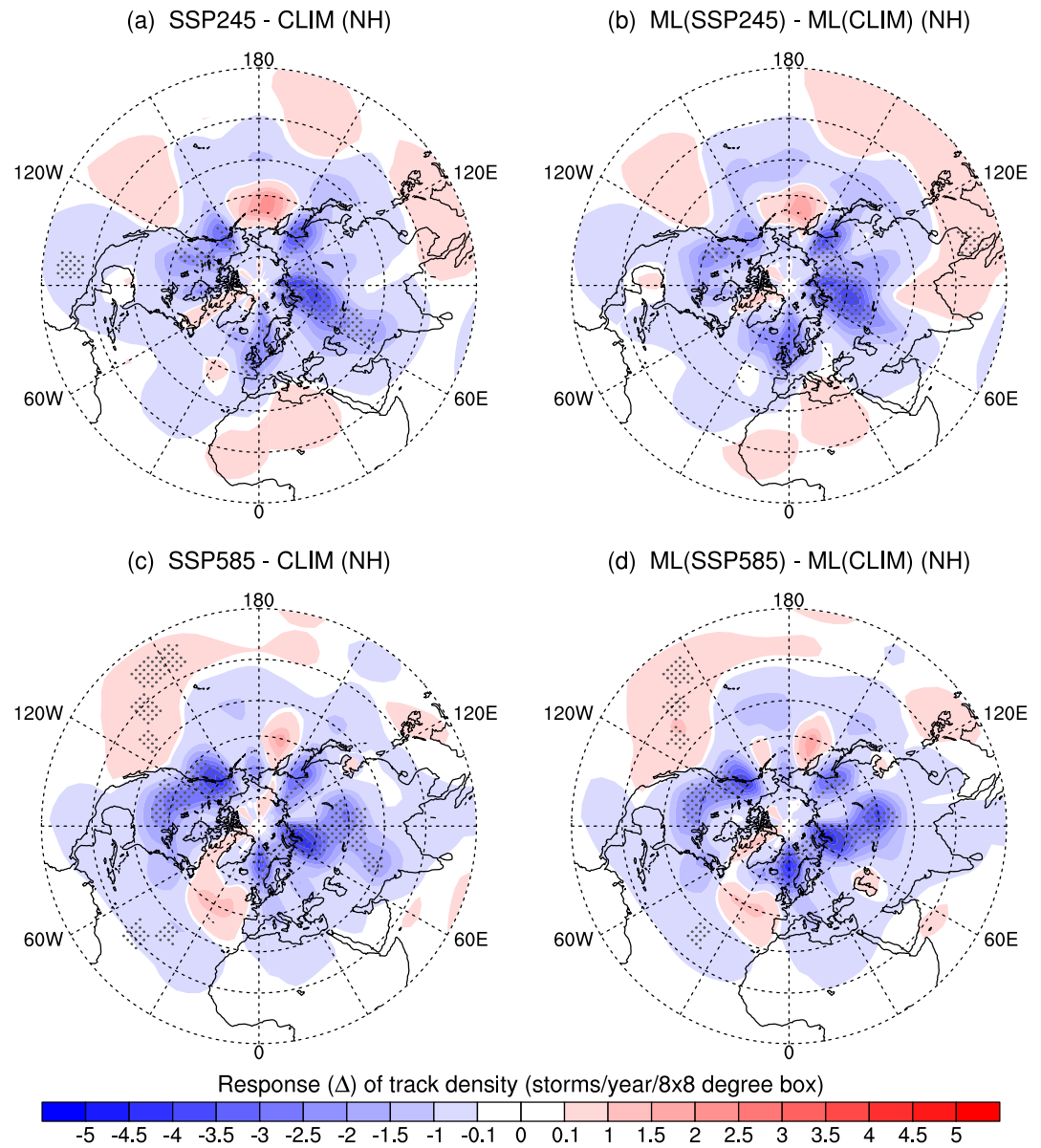


Figure 9. Responses (Δ) of annual ETC track densities over the North Hemisphere (NH) to the imposed climate change perturbations in sea-surface temperature (SST) and sea-ice concentration (SIC) for the SSP2-4.5 (top row) and the SSP5-8.5 scenarios (bottom row). Shown are the differences of SSP245—CLIM (panel a), SSP585—CLIM (panel c) for the EAMv2 simulations without ML bias correction, as well as ML (SSP245)—ML (CLIM) (panel b) and ML (SSP585)—ML (CLIM) (panel d) for the EAMv2 simulations and with ML bias correction. All ETCs with mean sea level pressure (PSL) less than or equal to 990-hPa are included for the metrics. The ETCs are tracked with 6-hourly PSL fields from EAMv2 simulations from 1979 to 2014. The annual track density is defined as the total number of time steps (6 hr) the ETCs passed over an $8^\circ \times 8^\circ$ grid box per year. The dotted region indicates that the differences are significant at a 90% significance level.

4.3. Tropical Cyclones (TCs)

Tropical cyclones (TCs) are low-pressure systems that typically form in lower-latitude regions, which can cause some devastating and widespread geophysical hazards in the global tropics and subtropics. Previous studies have evaluated the frequency and distribution of TCs in an earlier version of the EAM model at ~ 100 km resolution and found that the model significantly underestimates the occurrence frequency and intensity of TCs (Balaguru et al., 2020). The same conclusions can be drawn for the low-resolution EAMv2 simulations in our study (see Figure C6 in Appendix). Therefore, direct evaluation of TempestExtreme-derived metrics for TCs provides

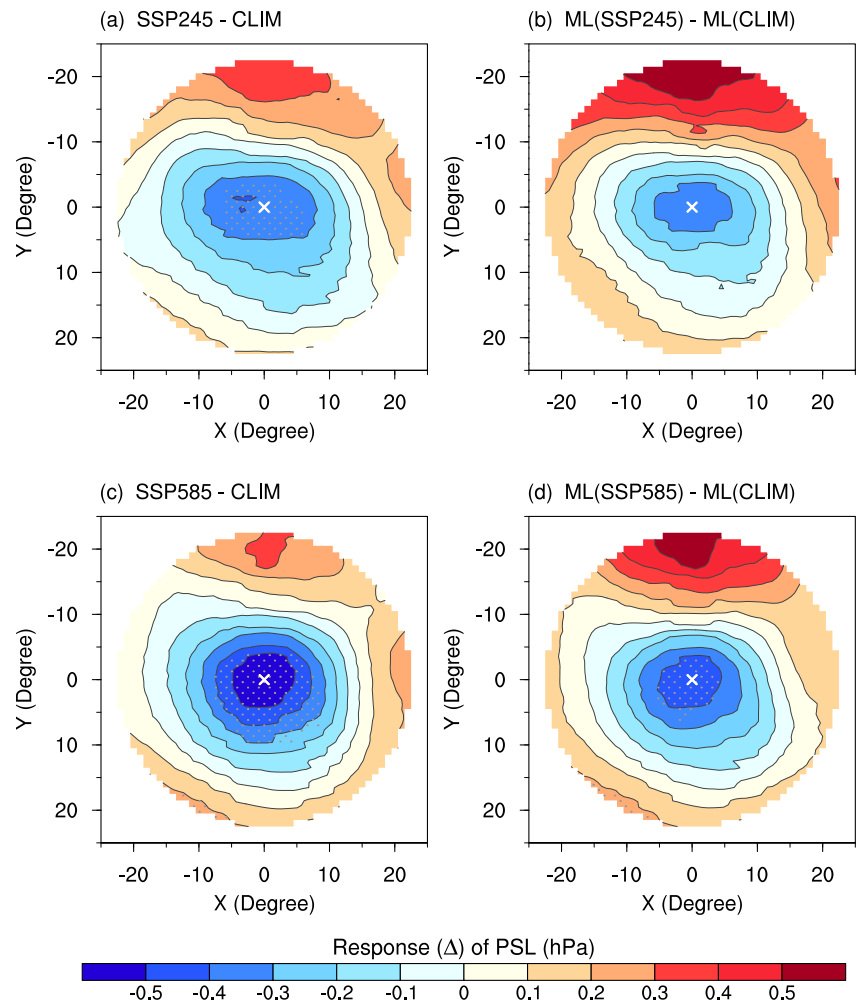


Figure 10. Responses (Δ) of the composite mean sea level pressure (PSL, unit: hPa) centered on the ETC storm center of all filtered storms in EAMv2 future climate simulations following SSP2-4.5 (first row) and SSP5-8.5 (second row) emission scenarios. The Δ s are derived by subtracting the composite PSL in SSP245 and SSP585 simulations from the CLIM simulations. All ETCs tracked with 6-hourly PSL fields during 1979–2014 are included in the composite by filtering out the storms with centered PSL >990-hPa. The white cross markers (panels a–d) indicate the center of ETC storms.

limited value to conclude the impacts of the ML bias correction model. In this section, we instead focus on the evaluation of large-scale environmental conditions that are key drivers governing TC formation and development. These large-scale storm environments usually operate on the order of tens of thousands of kilometers that can be resolved by the EAMv2 model.

Figure 12 shows the climatological seasonal mean TC cyclone genesis potential index (GPI), potential intensity (PI), and vertical wind shear between 200 and 850 hPa. Here, the GPI and PI are defined using the large-scale vorticity, vertical wind shear, potential intensity, and humidity fields following Camargo et al. (2007) (also see discussion in Appendix B). The differences between CLIM and ERA5 in terms of the GPI and low-level wind shear (Figures 12b and 12h) are systematically reduced over most of the TC basins when the ML bias correction is applied to the U, V, T, and Q fields (Figures 12c and 12i). Although ML (CLIM) relative to CLIM produces larger biases in PI over the tropical Indian and western Pacific Ocean regions within 5°S – 5°N, ML (CLIM) with bias correction (Figure 12f) effectively reduced the significant positive biases of PI in CLIM (Figure 12e) over the regions with maximum track densities (see Figure C6b). Overall, we conclude that ML bias correction shows the potential to improve the representation of large-scale storm environments associated with TCs in the low-resolution EAMv2. Further analysis from Equation B7 suggested that these improvements are possibly due to the bias reduction in large-scale U, V, T, and Q fields with the ML bias correction. In addition, the CLIM

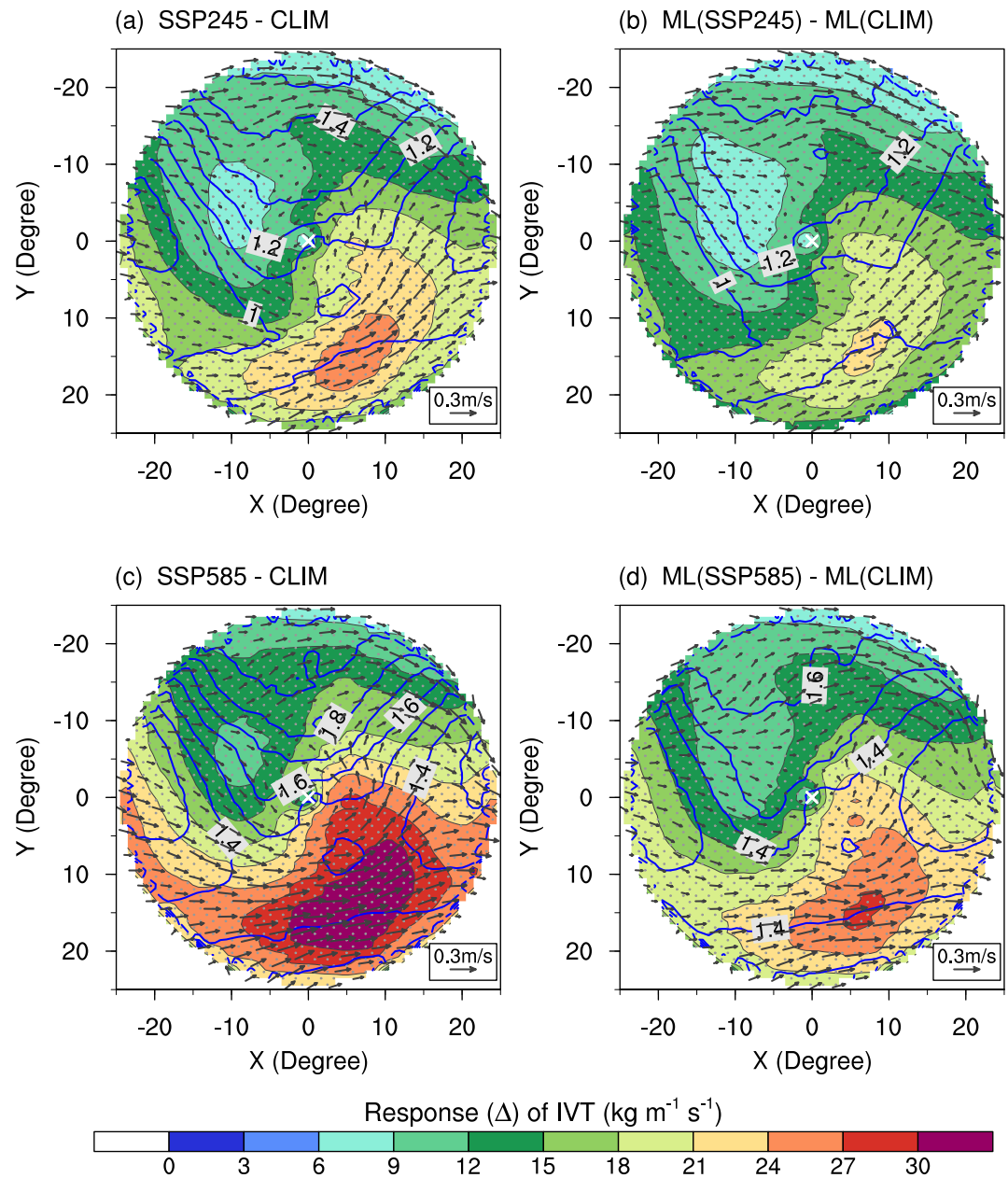


Figure 11. Same as Figure 10 but for responses (Δ) of temperature (blue contours, unit: K) and wind (black vectors) at 850-hPa pressure level as well as the vertically integrated vapor transport (IVT, unit: $\text{kg m}^{-1} \text{s}^{-1}$) to climate change following SSP2-4.5 (first row) and SSP5-8.5 (second row) emission scenarios. See Table 1 for a detailed description of EAMv2 simulations.

simulation, as shown in Figure 12h, features a noticeable overestimation of wind shear over tropical eastern Pacific and Atlantic ocean regions. This could also partly account for the significant underestimation of TC track densities over the Northeast Pacific and North Atlantic basin (Figure C6b versus Figure C6a) because the activity of the tropical easterly waves over tropical eastern Pacific and Atlantic oceans are known as key drivers for TC genesis. Interestingly, the biases in wind shear over these two regions in CLIM are significantly reduced in ML (CLIM) after applying the ML corrections (Figure 12i).

To demonstrate if the ML corrections on large-scale wind indeed lead to changes in TC activities, we employ the Risk Analysis Framework for Tropical Cyclones (RAFT, W. Xu et al., 2021) for a complementary assessment. The TC track model of RAFT is used to simulate TC track density given the climatological steering winds.

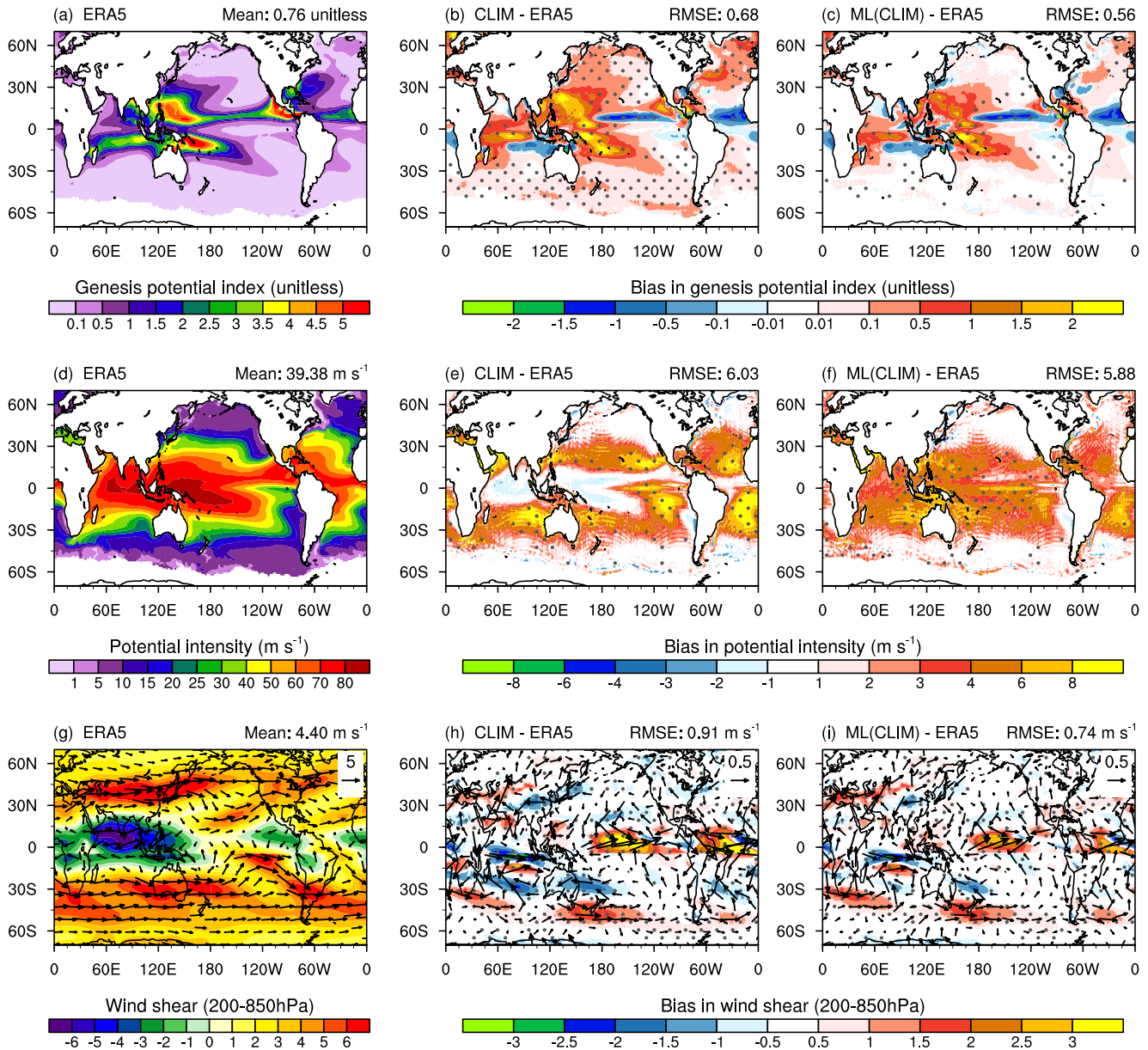


Figure 12. First column: Seasonal mean tropical cyclone Genesis Potential Index (GPI, unitless, panel a), potential intensity (PI, unit: m s^{-1} , panel d), and vertical wind shear between 200 and 850 hPa (unit: m s^{-1} , panel g) from ERA5 reanalysis averaged from 1979 to 2014; Second column: bias in GPI (panel b), PI (panel e) and 200–850 hPa vertical wind shear (panel h) in the EAMv2 simulation without ML bias correction (i.e., CLIM—ERA5); Third column: the same as the second column but for the EAMv2 simulations with ML bias correction (i.e., ML (CLIM)—ERA5, panels c, f, i). The monthly mean model output from ERA5 reanalysis and EAMv2 simulations are used to calculate the GPI, PI, and vertical wind shear following Equation B7. For all panels, the seasonal mean values are computed for August to October in the Northern Hemisphere and for January to March in the Southern Hemisphere. Detailed description of EAMv2 simulations can be found in Table 1.

Comparison of the TC track density produced by RAFT as driven by EAMv2 simulated steering winds with and without ML bias correction provides an assessment of the large-scale TC environment in the simulations. Following W. Xu et al. (2021), we used the 6-hourly large-scale wind fields (i.e. U and V) at 200-hPa and 850-hPa from ERA5 reanalysis, CLIM, and ML (CLIM) to generate three sets of synthetic TC tracks with RAFT, respectively. We hypothesize that the synthetic tracks should agree better with those obtained with ERA5 reanalysis if the ML bias correction improves the large-scale wind fields.

Figure 13 shows the annual mean TC track density over the Atlantic basin from the RAFT forecast. Compared to CLIM (Figure 13b), ML (CLIM) shows a better agreement with ERA5 reanalysis, with a clear reduction of track

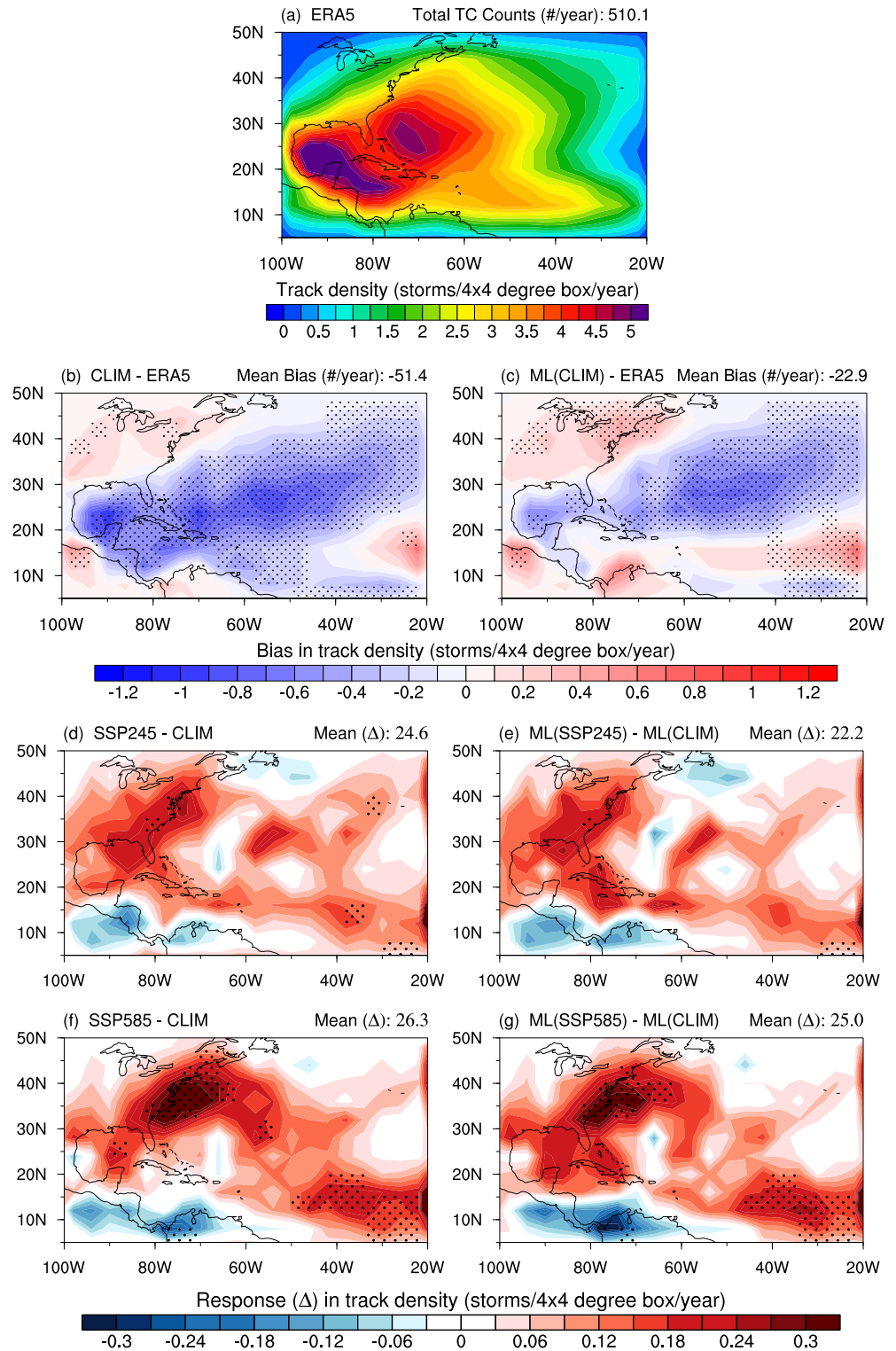


Figure 13. Annual mean TC track density over the Atlantic basin from RAFT forecast driven by the large-scale environmental wind fields from ERA5 reanalysis (panel a), and the differences in track densities between RAFT forecasts driven by ERA5 and EAMv2 simulations (panels b–c), and by EAMv2 simulations in present-day and pseudo-global warming scenarios (panels d–f), respectively. Panels (b–c) show the CLIM—ERA5 (panel b) and ML (CLIM)—ERA5 (panel c) for present-day EAMv2 simulations without and with ML bias correction, respectively. Panels (d–g) show the SSP245—CLIM (panel d) and SSP585—CLIM (panel f) for EAMv2 simulations without ML bias correction, as well as ML

density biases (Figure 13c). The basin mean track density biases are reduced by more than 50% percent as shown by the numbers on the top right corner in Figures 13b and 13c). These results validate our hypothesis as discussed above, and demonstrate that the ML bias correction improves the RAFT TC track forecasts and thus the downscale analysis of the statistics of TC track densities. Similar track density forecasts from RAFT are also generated for EAMv2 PGW simulations with and without ML bias correction. As shown in Figures 13d and 13f, the climate-change-induced changes in the large-scale wind fields led to a significant increase in the number of TCs over the Atlantic Basin region, especially in the coastal regions over the eastern U.S. These climate change signals, which have been linked to the warmer SSTs over the eastern tropical Pacific Ocean under warming (Balaguru et al., 2023) (also seen in Figures A1a and A1b), are still seen in the simulations after applying the ML bias correction (Figures 13e and 13g), suggesting that the ML bias correction on the large-scale wind fields preserves the climate change signals and the associated TC track responses as seen in the free-running simulations. Moreover, we observe differences in the magnitude of the responses of TC track density over Eastern US coastal and Gulf of Mexico coastal regions before and after ML bias correction (Figures 13e and 13g versus Figures 13d and 13f). This reflects the impact of ML bias correction on the RAFT forecasts and the associated TC track responses through modifications on the large-scale wind fields in EAMv2 simulations. As the ML bias correction produces a more reliable representation of large-scale wind fields in EAMv2, higher confidence could be given to the results drawn from Figures 13e and 13g. Overall, the improved representation of large-scale storm environments (e.g., large-scale wind) by ML bias correction is beneficial for obtaining a more reliable downscaling of high-impact weather systems such as TCs.

5. Conclusions

Bias correction has been a commonly used approach when applying climate model outputs to impact studies. This study employed a machine-learning-based (ML) bias correction approach to improve the representation of the large-scale wind (U, V), temperature (T), and humidity (Q) in the climate simulations conducted with DOE's E3SM Atmosphere Model (EAM). The performance of the ML bias correction method in producing large-scale storm environments associated with high-impact weather systems is evaluated for both present-day (i.e., historical) and climate change scenarios.

Globally, the results show that the ML bias correction method performs well in reducing the overall biases in U, V, T, and Q fields from the climate model simulations. Compared with the wind fields, more promising corrections are found in the thermodynamical fields (i.e. T and Q), especially in the tropics and midlatitude regions and over the lower troposphere (see Figures 1 and C1). As reported in previous studies (S. Zhang et al., 2022), biases are more pronounced in these fields compared to the winds. Therefore, there is more room for these larger biases to be corrected during training for these fields compared to winds. When looking at the mean values (global and regional means), bias correction is very efficient at removing the biases in all fields at most model levels, with a systematic bias reduction of 10%–20% quantitatively (Figure 2). The same ML bias correction approach is then applied to process the PGW simulations from EAMv2 forced with the imposed climate change perturbations in sea surface temperature (SST) and sea-ice concentration (SIC) derived for the future climate scenarios of SSP2-4.5 and SSP5-8.5. The ML bias correction is found to constrain the probability distribution function (PDF) of the large-scale model state variable in historical simulations toward a better agreement with the observations. Similar shifting of the PDFs by ML bias correction is also seen in the PGW simulations, while the large-scale climate change signals of the model state (e.g., temperature and humidity) are well preserved before and after the ML bias correction (see Figure 4).

This study further demonstrated the value of the employed ML bias correction in the assessment of high-impact weather systems that have the potential to generate extreme weather events. We used the model state of U, V, T, and Q with and without ML bias correction in the low-resolution EAM model to derive the long-term statistics,

(SSP245)—ML (CLIM) (panel e) and ML (SSP585)—ML (CLIM) for EAMv2 simulations with ML bias correction. The SSP245 (or ML (SSP245)) and SSP585 (or ML (SSP585)) are two PWG simulations with imposed climate change perturbations in sea surface temperature and sea-ice concentrations derived for SSP2-4.5 and SSP5-8.5 future climate scenarios. More detailed descriptions of simulations can be found in Table 1. The 6-hourly zonal wind (U) and meridional wind (V) at 200 and 850 hPa during 1979–2014 from ERA5 reanalysis and EAMv2 simulations are used to drive RAFT TC track forecasts following W. Xu et al. (2021). The annual track density is defined as the total number of 6-hourly tracks that pass through a $4^\circ \times 4^\circ$ grid box per year. The gray dots in panels b and c indicate the differences are significant at a 95% confidence level.

and evaluated the skills of bias correction in improving the model representation of the high-impact weather systems (e.g., occurrence frequency, intensity, and storm environment, etc.) in both present-day and PGW scenarios, with a focus on atmospheric rivers (ARs), ETCs and tropical cyclones (TCs). The results show that the large-scale vapor transport associated with ARs is more realistically represented in the bias-corrected data sets than those without bias correction, leading to a better representation of the occurrence frequency and the strength of ARs in the EAM model (see Section 4.1). Similarly, more realistic representations of ETC structure and ETC-induced changes in water vapor transport and thermodynamical flows are also obtained in the simulations with ML bias correction (see Section 4.2). When the ML bias-corrected large-scale winds are used to drive a TC track forecast model for downscaling analysis of TC activities over the Atlantic basin, the resulting TC track forecasts agree better with the results driven by observations (see Section 4.3). In addition, the ML bias correction does not significantly change the patterns of the responses of occurrence frequency and intensity of the three types of extreme events to pseudo-global warming effects, but there are obvious differences in the magnitude of the responses before and after the ML bias correction. Analysis of the ETC response to climate change shows that the ML bias correction preserves the physical relationship between the storm environment and storm intensity. Overall, the findings in this study suggest that the proposed machine learning bias correction is a useful approach to facilitate the downscaling of high-impact weather systems for low-resolution climate models by providing more realistic large-scale environment information.

While the proposed ML bias correction was demonstrated to be effective in the assessment of high-impact weather systems in the low-resolution EAMv2 climate models, some limitations of the current setup should be stated. First, as mentioned in Section 2.3, the ML model employed in this study for bias correction acts on the coarse-resolution climate model simulation in a postprocessing manner. The goal of such debiasing process was to bring the climate statistics of quantities predicted by the coarser-resolution numerical model into better agreement with the reference as quantified by the ERA5 data set. Therefore, our ML bias approach is not expected to correct biases in the dynamics of the model system. Nevertheless, the offline ML bias correction proposed in our study can be potentially combined with an online bias correction approach targeting direct corrections on the model dynamics to derive more optimal bias reduction for coarser-resolution climate model predictions. Second, the ML bias correction in this study was only applied to improve the statistics of large-scale model state variables, namely U, V, T, and Q fields. The ability of the ML approach to bias correct other physical quantities such as precipitation and radiative fluxes and to preserve the physical consistency and conservation laws of mass and energy has not been explored and discussed. However, the results of this study have demonstrated that our ML model is in a good position for further exploration of these issues. Lastly, ML bias correction was currently trained and verified only for the EAMv2 climate model. Applying the ML approach to correct other weather and climate models could be possible with retraining, but this has not been tested and verified. Future work will focus on addressing the limitations of our ML bias correction model as mentioned here and exploring the potential application of our ML architecture to other weather and climate models. We will report our findings in separate publications.

Appendix A: Supplementary Material for Section 2

A1. Additional Tables and Figures for Section 2.2

A2. Machine Learning Framework

The machine learning model for bias correction in Section 2.2 utilizes the same convolutional-LSTM hybrid neural network (NN) architecture described in Barthel Sorensen et al. (2024). The network takes as its input the snapshots of the entire horizontal discretization of all prognostic variables (i.e. U, V, T, Q) at a single sigma level of the EAMv2 model. Afterward, a custom "split" layer separates the input into non-overlapping subregions. Then, each subregion is independently passed through a series of convolutional layers tasked with extracting local flow features. Afterward, the local information extracted from each subregion is concatenated in a single vector via a custom "merge" layer and projected onto a reduced order latent space via a linear fully connected layer. This latent space representation is then fed through an LSTM layer before being projected back to physical space via another linear fully connected layer. In addition, global information is split into the same subregions of the input, and distributed to a series of independent deconvolution layers that upscale the data to the original resolution. Finally, a custom "merge" layer gathers the information from each subregion and produces the final corrected snapshot. The mathematical framework and algorithms for the machine learning operator are introduced below.

Table A1

List of CMIP6 Models Used to Derive the Imposed Climate Change Perturbations in Sea Surface Temperature (SST) and Sea Ice Concentrations (SIC) for Pseudo Global Warming Simulations in Table 1

CMIP6 institution	CMIP6 model
AWI	AWI-CM-1-1-MR
BCC	BCC-CSM2-MR
CAMS	CAMS-CSM1-0
CAS	FGOALS-f3-L
CCCma	CanESM5
CNRM-CERFACS	CNRM-CM6-1
EC-Earth-Consortium	EC-Earth3
IPSL	IPSL-CM6A-LR
MIROC	MIROC-ESM
MOHC	UKESM1-0-LL
MRI	MRI-ESM2-0
NCAR	CESM2
NCC	NorESM2-LM
NOAA-GFDL	GFDL-CM4
NUIST	NESM3

Note. The monthly mean model output of “tos” (SST) and “siconc” (SIC) from the “r1i1p1f1” experiment conducted for “SSP24-5” and “SSP58-5” scenarios were extracted and used.

Consider a coarse discretization of a dynamical system, in this case the EAMv2 model, describing the evolution of the vector quantity v

$$\dot{v} = f(v). \quad (\text{A1})$$

The high fidelity reference solution, in this case ERA5 data, is represented by u . The objective of the ML framework we employ is to capture the long term statistics of u by solving the imperfect model (Equation A1) and then applying a correction operator, \mathcal{G} , to that computed solution.

An ML model naively trained a pair of arbitrary trajectories (v, u) is unlikely to generalize as it will be corrupted by the effects of chaotic divergence. Chaotic divergence is the inherent property of all turbulent systems that any two trajectories, which may initially be arbitrarily close, will eventually diverge—making a mapping between them meaningless. To minimize this effect the correction operator, \mathcal{G} is trained not on an arbitrary pair of trajectories but specifically on the pair (v_τ, u) where v_τ is the solution to the coarse model nudged toward the reference data,

$$\dot{v}_\tau = f(v_\tau) - \frac{1}{\tau}(v_\tau - \mathcal{P}u). \quad (\text{A2})$$

Here \mathcal{P} is an operator which projects the reference solution onto the coarse grid. The constant τ is a user-defined parameter that represents the timescale over which the nudging tendency acts. While this value is chosen such that the nudging term is smaller than all others, it still creates small discrepancies between the spectra of the nudged solution, v_τ , and the free coarse solution, v . If left unaddressed this discrepancy will hinder the ability of the machine-

learned map \mathcal{G} to generalize to free-running data. To remedy this issue the spectrum of the nudged trajectory, v_τ is rescaled to match the spectrum of the free-running coarse model.

Specifically, let $\hat{q}_k = \mathcal{F}[q]$ be the spatial Fourier transform of an arbitrary field q . The spectral energy is then defined as

$$\mathcal{E}_{k,q} = \frac{1}{T} \int_0^T |\hat{q}_k|^2 dt, \quad (\text{A3})$$

and the energy-ratio between v and v_τ is defined as

$$a_k \equiv \sqrt{\frac{\mathcal{E}_{k,v}}{\mathcal{E}_{k,v_\tau}}} \quad (\text{A4})$$

The spectrum-matched nudged solution is then defined as the inverse Fourier transform of the spectrally rescaled nudged solution:

$$v'_\tau = \mathcal{F}^{-1}[a_k \hat{v}_{k,\tau}]. \quad (\text{A5})$$

Training the correction operator then reduces to a supervised learning problem with an objective function

$$\min_{\mathcal{G}} \int_0^T \|\mathcal{G}[v'_\tau(t)] - u(t)\|^2 dt. \quad (\text{A6})$$

After the correction operator, \mathcal{G} , is trained on the spectrally corrected nudged data, during testing it is applied to the free run coarse model trajectory $v(t)$. The resulting corrected trajectory constitutes our ML prediction and is

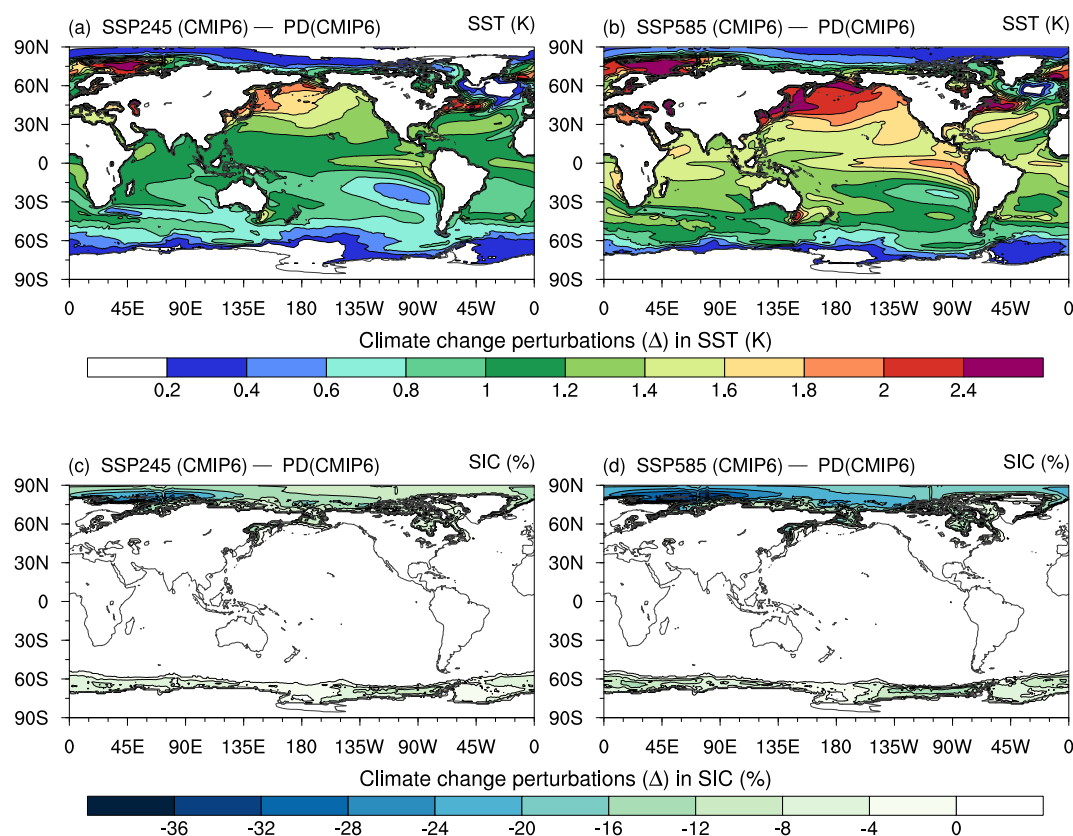


Figure A1. Changes of annual mean sea surface temperature (SST, unit: K, panels a, c) and sea-ice concentration (SIC, unit: %, panels b, d) in response to the forcing pathways of SSP2-4.5 (top row) and SSP5-8.5 (bottom row) from CMIP6 coupled model simulations. Shown are the multi-model ensemble mean climatological differences averaged over the 15 models listed in Table A1. The climatological differences are computed with the output from coupled historical simulations during 1991–2010 and future climate simulations during 2041–2060. More detailed descriptions on the simulations and models in Table A1 can be found in Eyring et al. (2016) and O'Neill et al. (2016).

then used to compute statistics and other properties of interest. We refer the interested reader to Barthel Sorensen et al. (2024) for a more detailed discussion of the mathematical framework and network architecture.

Appendix B: Additional Notes for Section 2.4 and Section 4

Section 4 discussed three high-impact weather systems, including atmospheric rivers (ARs), ETCs, and tropical cyclones (TCs). Table B1 documents the feature tracking information used to derive the metrics for discussions. The TempestExtremes package (P. A. Ullrich et al., 2021) was employed for feature detection and tracking of these weather phenomena using the 6-hourly model output from EAMv2. Specifically, the TCs and ETCs were tracked with mean sea level pressure (PSL), while the ARs were tracked with two components of integrated water vapor flux (TUQ and TVQ). In this paper, the same algorithm and parameter setups described in P. A. Ullrich

Table B1

List of Extreme Weather Events and Model Variables Used for Feature Tracking by the TempestExtremes Package

Feature events	Short name	Feature quantity	EAM output
Tropical cyclones	TCs	Mean sea level pressure (PSL)	U, V, T, Q, PS
Extratropical cyclones	ETCs	Mean sea level pressure (PSL)	U, V, T, Q, PS
Atmospheric rivers	ARs	Integrated Vapor Transport (IVT)	U, V, Q

Note. The EAM model output used to derive the feature quantities is listed in the fourth column. See context in Appendix C for details.

et al. (2021) were used by TempestExtremes for feature tracking of each extreme weather event in EAMv2 simulations:

- For TCs, our configuration for TempestExtreme identifies the features on the grid points that have both a PSL minimum and an upper-level warm core. Specifically, the candidate model grid points are first identified and tagged by minima in PSL and then eliminated if a more intense minimum exists within a great-circle distance (CGD) of 6° . The closed contour criteria are then applied, requiring an increase in SLP of at least 2 hPa within 5.5° GCD of the candidate node, and the average temperature over 200- and 500-hPa pressure levels must decrease by 0.6 K within 5.5° GCD of the node within 1° CGD of the candidate with maximum air temperature. Meanwhile, the maximum magnitude of the vector wind at 10 m altitude (estimated with the wind fields at the bottom model level of EAMv2) within 2° GCD of the candidate, and the surface height at the candidate point was also identified and output by TempestExtreme. All threshold values selected here for TC feature tracking are based on Table 1 in Zarzycki and Ullrich (2017). After identifying TC candidates on each time slice (i.e., 6 hr), the “stitching” step in the TempestExtremes (P. A. Ullrich et al., 2021) was further employed to ensure that the identified features can be sufficiently classified as tropical storms: the wind magnitude (derived from the wind fields at bottom model level) must be greater than 10 m s^{-1} for at least 10 time-slices, the latitude of the feature must be between 50°S and 50°N for at least 10 time-slices (i.e., 60 hours), the feature must exist at an elevation below 150 m for at least 10 time-slices, and the maximum distance between feature candidates are of 8.0° GCD. In addition, we only kept the TC candidates that persist for at least 24 hr with a maximum gap size of 6 hr for evaluation and analysis in Section 4.
- For ETCs, our configuration for TempestExtreme identified the features on the grid points with PSL minima and the minimum PSL must be enclosed by a closed contour of 2 hPa within 5.5° GCD of the feature center. Unlike the TCs, ETCs do not have a unique warm core structure. Therefore, the above-mentioned TC warm core structure criteria defined with average temperature over 200- and 500-hPa pressure levels were applied to exclude the TCs from the tracked ETC candidates. During the “stitching” step, we further filter the tracked feature candidates with the following criteria: the feature must exist at an elevation below 1500 m for at least 8 time-slices (i.e., 48 hours), and the maximum distance between feature candidates is of 6.0° GCD. In addition, we only kept the ETC candidates that persist for at least 60 hr with a maximum gap size of 18 hr for evaluation and analysis in Section 4.
- For ARs, our configuration for TempestExtreme first calculated the Laplacian of the IVT field using eight radial points at a 10° GCD from each candidate model grid point. Following McClenny et al. (2020), the feature of AR was then identified and tagged on the grid points where the Laplacian of the IVT field is below a fixed threshold of $-4 \times 10^4 \text{ kg m}^{-2} \text{ s}^{-1} \text{ rad}^{-2}$. Finally, we removed the features too near the Equator and those that are deemed too small using filtering criteria typically used for AR trackers (Shields et al., 2018): each blob must have a minimum area of $4 \times 10^5 \text{ km}^2$, and the latitude of each tagged grid point must be at least 15° . The grid points belonging to a tropical cyclone identified in the first item were also excluded from the ARs.

With the feature tracking algorithms described above, 6-hourly TC, ETC track forecasts with a similar format as National Hurricane Center's HURDAT2 database (Landsea & Franklin, 2013), and 6-hourly AR mask data on the $1.5^\circ \times 1.5^\circ$ global lat-lon grid will be obtained for EAMv2 simulations in Table 1. These tracking results were used to derive occurrence frequency metrics for ARs (Figure C3, Figures 5 and 6), ETCs (Figure C5 and Figures 7 and 9) and TCs (Figure C6) and ARs as shown in Section 4 and Appendix A1. This was achieved by binning the 6-hourly TCs and ETC track data into a selected bin (described in the captions of each figure), or directly calculating the counts with the AR mask data. In addition, the “NodeFileCompose” function built in TempestExtreme was employed for composite analysis and metrics for ETCs shown in Figures 8, 10, and 11. Specifically, the NodeFileCompose first projects the model fields (e.g., wind, temperature, etc.) onto the stereographic plane centered on the ETC track to form a storm-centered snapshot for each field at each time slice (6 hr). The composite of each field is then derived by averaging all 6-hourly snapshots generated at the first step. More detailed descriptions and verifications on the composite analysis method can be seen in P. A. Ullrich et al. (2021), and the specific rules for the composite analysis in our study are also described in the captions under Figures 8, 10, and 11.

In addition, as the ML bias correction was applied to U, V, T, and Q at each grid point and model level, the feature tracking quantities, including PSL and IVT in this study were diagnosed offline for simulations with and without ML bias correction listed in Table 1. Specifically, the PSL was diagnosed with the algorithm proposed by the European Center for Medium-Range Weather Forecasts (ECMWF, Trenberth et al., 1993):

$$PSL = \begin{cases} P_s, & \text{if } Z_s \leq 1e^{-4} \text{ m} \\ P_s e^{\frac{Z_s}{R_d T_s} \left(1 - \frac{Z_s}{2} + \frac{Z_s^2}{3}\right)}, & \text{otherwise} \end{cases} \quad (\text{B1})$$

where,

$$x = \begin{cases} 0, & \text{if } T_s > 290.5 \text{ K and } T_0 > 290.5 \text{ K} \\ \frac{R_d (290.5 - T_s)}{g T_s}, & \text{if } T_s \leq 290.5 \text{ K and } T_0 > 290.5 \text{ K} \\ \frac{R_d \Gamma_0 Z_s}{g T_s}, & \text{otherwise} \end{cases} \quad (\text{B2})$$

$$T_0 = T_b \left[1 + \frac{R_d \Gamma_0}{g} \left(\frac{P_s}{P_b} - 1 \right) \right] + \Gamma_0 Z_s \quad (\text{B3})$$

$$T_s = T_{bot} \left[1 + \frac{R_d \Gamma_0}{g} \left(\frac{P_s}{P_b} - 1 \right) \right], \text{ where} \quad (\text{B4})$$

$$T_s = \begin{cases} \frac{T_s + 290.5}{2}, & \text{if } T_s > 290.5 \text{ K and } T_0 > 290.5 \text{ K} \\ \frac{T_s + 255.0}{2}, & \text{if } T_s < 255.5 \text{ K} \end{cases}$$

Here, P_s and Z_s are surface pressure and Geopotential height, respectively. T_{bot} and P_{bot} are air temperature and pressure at the bottom model level, respectively. These quantities are directly from the model output of simulations in Table 1. Γ_0 ($=6.5e^{-3} \text{ K m}^{-1}$) is the temperature lapse rate. R_d and g denote the dry air gas constant and gravitational acceleration, respectively.

Moreover, the water vapor transport in the atmosphere consists of the eastward (TUQ) and northward (TVQ) components, which were derived directly with the U, V, and Q from the model output using:

$$TUQ = \frac{1}{g} \int_{P_s}^{P_t} UQ dP \quad (\text{B5})$$

$$TVQ = \frac{1}{g} \int_{P_s}^{P_t} VQ dP \quad (\text{B6})$$

where dP is layer thickness, and the integral was computed from surface (i.e., P_s) to top model level (i.e., P_t) at $\sim 0.2 \text{ hPa}$.

Finally, the cyclone genesis potential index (GPI), and potential intensity (PI) shown in Figure 12 are defined using the large-scale vorticity, vertical wind shear, potential intensity, and humidity fields following (Camargo et al., 2007):

$$GPI = 10^5 \eta^3 \left(\frac{H_{700}}{50} \right)^3 \left(\frac{V_{pot}}{70} \right)^3 (1 + 0.1 V_{shear})^{-2} \quad (\text{B7})$$

where η is the absolute vorticity at 850 hPa in s^{-1} , H_{700} is the relative humidity at 700 hPa in percent, V_{pot} is the potential intensity (PI) computed with the method proposed by (Emanuel, 2000; Knutson et al., 2013). The unit of PI is in m s^{-1} . V_{shear} is the magnitude of the vertical wind shear between 850 and 200 hPa in m s^{-1} .

Appendix C: Additional Tables and Figures

This section contains supplemental figures used for discussions in Section 3 and Section 4.

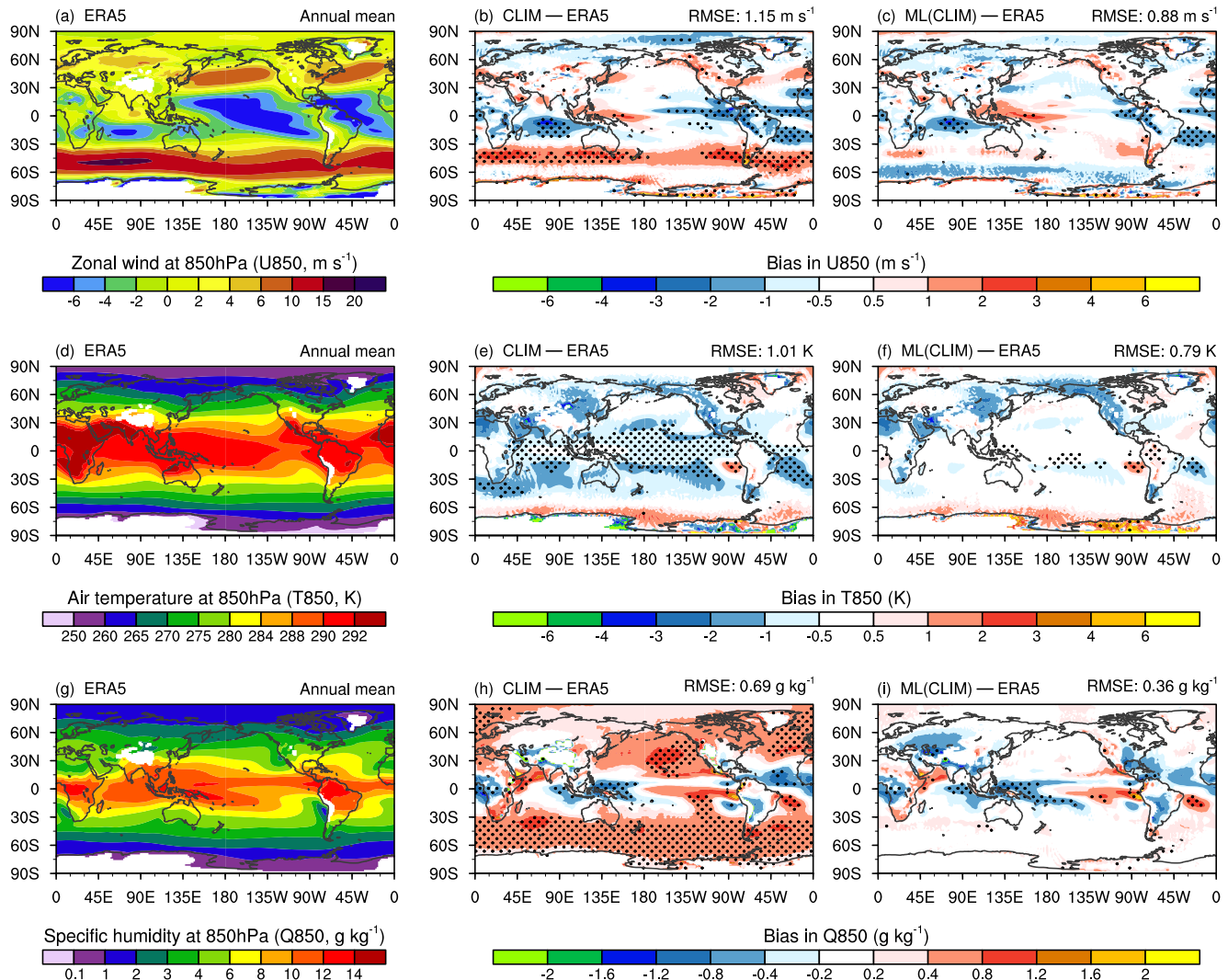


Figure C1. First column: Distribution of annual mean zonal velocity (U850, unit: m s^{-1} , panel a), temperature (T850, unit: K, panel d) and specific humidity (Q850, unit: g kg^{-1} , panel g) at 850-hPa pressure level from ERA5 reanalysis averaged over 1979–2014; Second and Third columns: the same as in first column, but for biases (with respect to ERA5 reanalysis) in U850 (panels b–c), T850 (panels e–f), and Q850 (panels h–i) from EAMv2 simulations without (i.e., CLIM, second column) and with (i.e., ML (CLIM), third column) bias correction. The numbers on the top right of panels in the second and third columns denote the global mean root-mean-square errors (RMSE). The log-linear interpolation is applied to map the EAMv2 data on the hybrid sigma-pressure level into pressure level, and to be compared with ERA5 reanalysis. Details of the simulation setups can be found in Table 1.

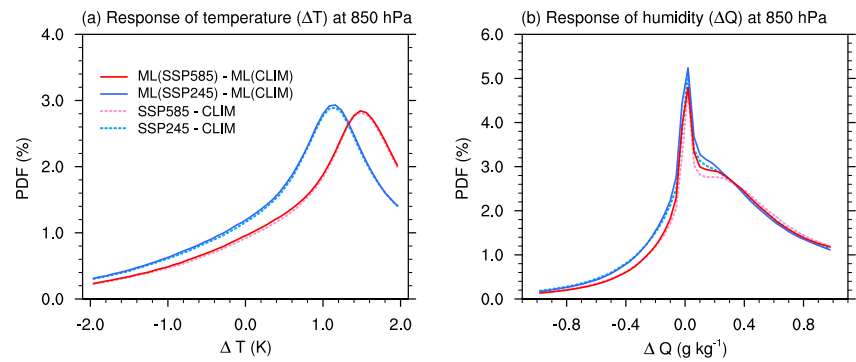


Figure C2. Probability Density Function (PDF) of monthly mean temperature (ΔT , unit: K, panel a) and humidity (ΔQ , unit: g kg^{-1} , panel b) differences between present-day and pseudo global warming simulations conducted with EAMv2. Shown are the SSP245–CLIM (dashed blue line) and SSP585–CLIM (dashed red line) for EAMv2 simulations without ML bias correction, as well as ML (SSP245)–ML (CLIM) (solid blue line) and ML (SSP585)–ML (CLIM) (solid red lines) for EAMv2 simulations with ML bias correction. The monthly mean data from each simulation during the 1979–2014 period were used to derive the metrics. The detailed descriptions of simulations can be found in Table 1.

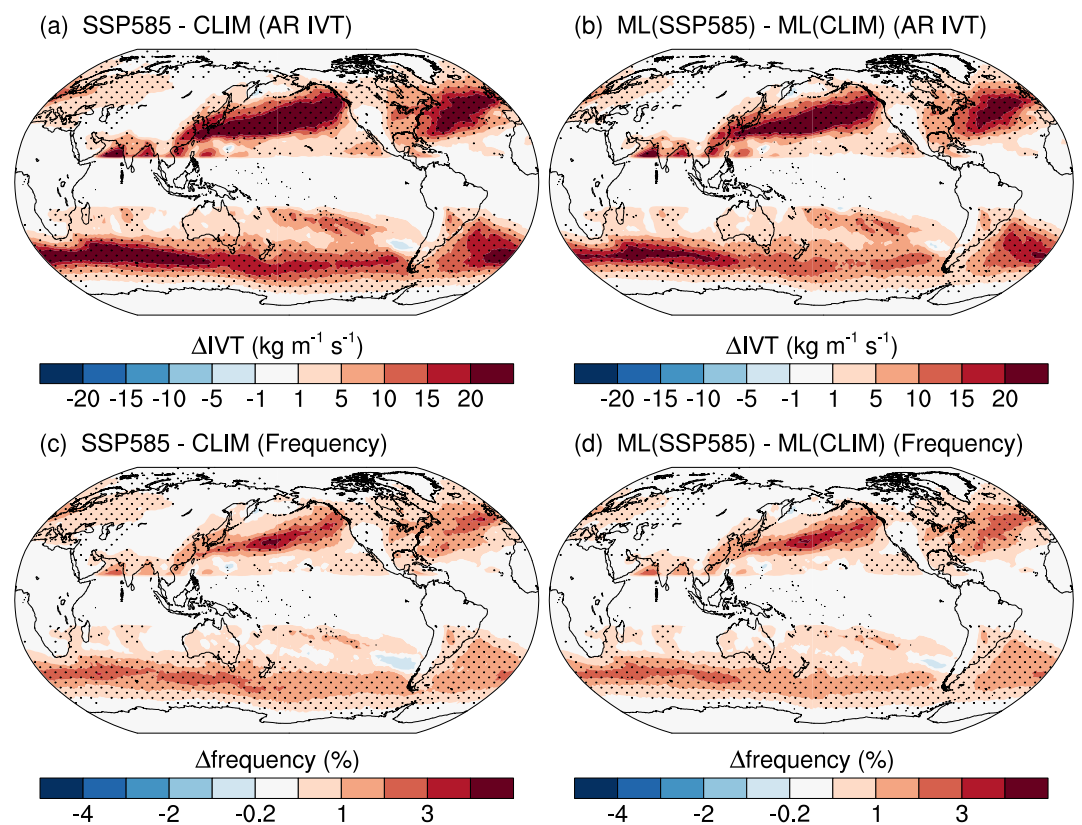


Figure C3. Same as Figure 6, but for the pseudo global warming simulations conducted with EAMv2 using imposed climate change perturbations in SST and SIC derived from SSP5-8.5 future scenarios.

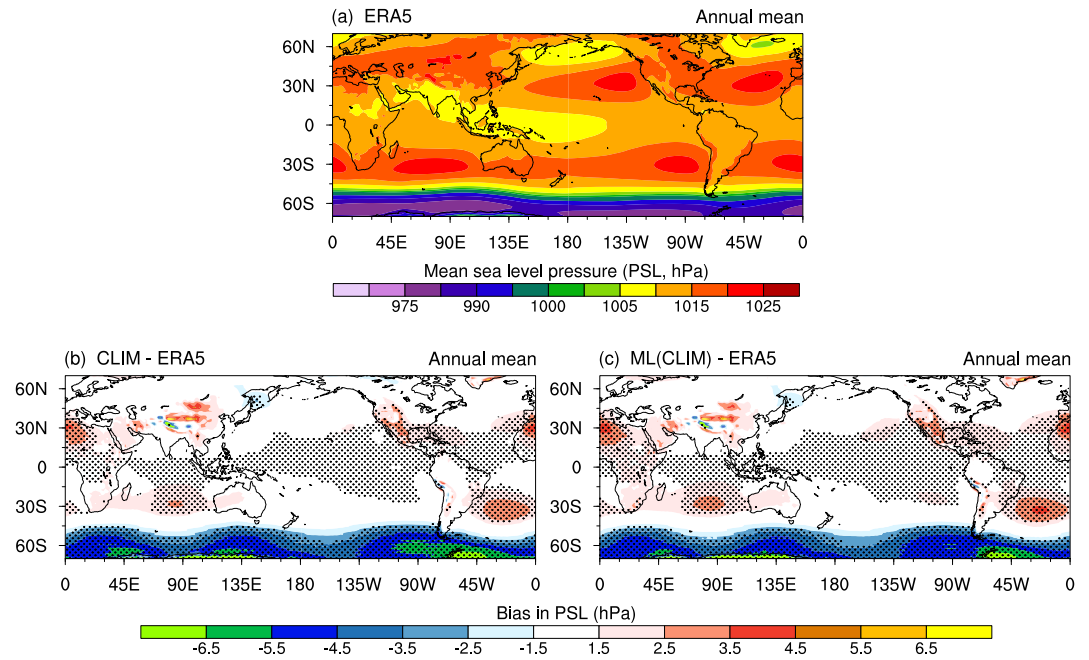


Figure C4. Horizontal distribution of mean sea level pressure averaged from 1979 to 2014 from ERA5 reanalysis (panel a), and mean model biases in EAMv2 free-running simulations without (i.e CLIM, panel b) and with (i.e., ML (CLIM), panel c) ML bias correction. The dotted regions in panels (b, c) indicate the differences between the model and ERA5 reanalysis are significant at a 95% significance level.

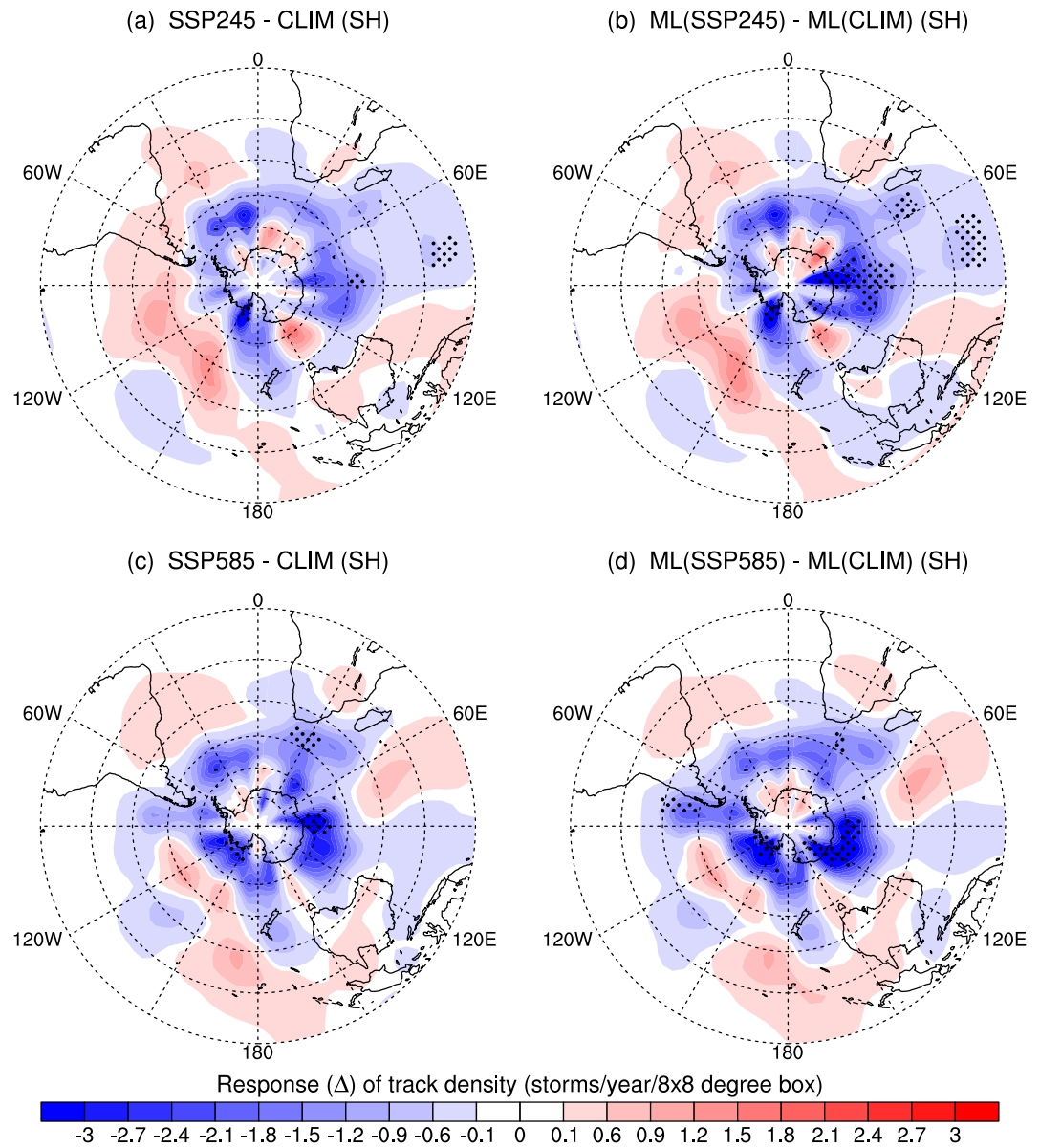


Figure C5. Same as Figure 9 but for the responses of ETC track density to imposed climate change perturbations in sea surface temperature (SST) and sea-ice concentration (SIC) from SSP2-4.5 (top row) and SSP5-8.5 (bottom row) over Southern Hemisphere.

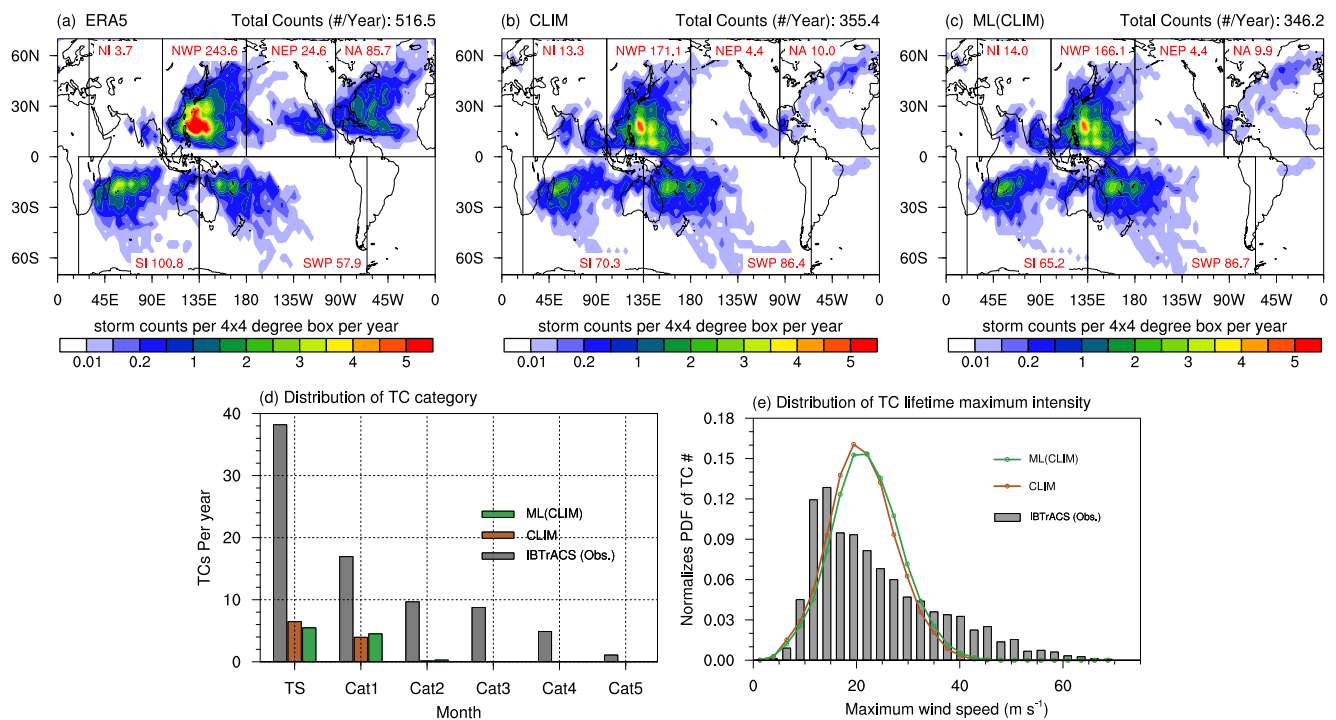


Figure C6. Top row: track density maps for tropical cyclones (TCs) tracked in ERA5 reanalysis (panel a) and EAMv2 free-running simulations without (i.e., CLIM, panel b) and with (i.e., ML (CLIM), panel c) ML bias correction. The 6-hourly sea level pressure (PSL) data from 1979 to 2014 are used to track the TC-like vortices at each model grid using the TempestExtremes. The TC track density is defined as the average number of 6-hourly TC track locations within a $4^\circ \times 4^\circ$ grid box per year. Bottom row: Climatological mean distribution of the TC numbers fall into the Saffir-Simpson wind scale (d) and the normalized probability distribution function (PDF) of the 10-m maximum wind speed (e) in IBTrACS observations (gray colored bars) and EAMv2 simulations without (brown colored bars and lines) and with (green colored bars and lines) ML bias correction. The x-axis in panel (d) corresponds to the Saffir-Simpson wind scale: TS, tropical storm ($17.5\text{--}32\text{ m s}^{-1}$); Cat1, Category 1 ($33\text{--}42\text{ m s}^{-1}$); Cat2, Category 2 ($43\text{--}49\text{ m s}^{-1}$); Cat3, Category 3 ($50\text{--}58\text{ m s}^{-1}$); Cat4, Category 4 ($59\text{--}69\text{ m s}^{-1}$); Cat5, Category 5 ($>69\text{ m s}^{-1}$). The IBTrACS refers to the International Best Track Archive for Climate Stewardship, which contains the TC track and intensity data in historical observations. The statistics in panels (d–e) are obtained with 6-hourly data from 1979 to 2014.

Data Availability Statement

The source code for EAMv2 (E3SM Project, 2021) used for simulations in this study was obtained from the Energy Exascale Earth System Model project, sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. The TempestExtremes package used for feature tracking of extreme weather events was extracted from the Github at <https://github.com/ClimateGlobalChange/tempest-extremes> (P. Ullrich, 2022), and the user guide for this package can be found at <https://climate.ucdavis.edu/tempestextremes.php> (P. A. Ullrich et al., 2021). The CMIP6 data used to derive the climate change perturbations of sea surface temperature (SST) and sea-ice concentration (SIC) are available at <https://esgf-node.llnl.gov/projects/cmip6/> (Eyring et al., 2016; O'Neill et al., 2016). The ERA5 reanalysis data used for machine learning training and evaluation in this study are available at the Copernicus Climate Change Service (C3S) Climate Data Store via <https://doi.org/10.24381/cds.bd0915c6> (Hersbach et al., 2020). The scripts and post-processed data for the analyses in this study can be found on Zenodo at <https://zenodo.org/doi/10.5281/zenodo.11053624> (S. Zhang & Charalampopoulos, 2024).

References

- Angéil, O., Perkins-Kirkpatrick, S., Alexander, L. V., Stone, D., Donat, M. G., Wehner, M., et al. (2016). Comparing regional precipitation and temperature extremes in climate model and reanalysis products. *Weather and Climate Extremes*, *13*, 35–43. <https://doi.org/10.1016/j.wace.2016.07.001>
- Balaguru, K., Leung, L. R., Van Roekel, L. P., Golaz, J.-C., Ullrich, P. A., Caldwell, P. M., et al. (2020). Characterizing tropical cyclones in the energy exascale earth system model version 1. *Journal of Advances in Modeling Earth Systems*, *12*(8), e2019MS002024. <https://doi.org/10.1029/2019MS002024>
- Balaguru, K., Xu, W., Chang, C.-C., Leung, L. R., Judi, D. R., Hagos, S. M., et al. (2023). Increased U.S. coastal hurricane risk under climate change. *Science Advances*, *9*(14), eadf0259. <https://doi.org/10.1126/sciadv.adf0259>

Acknowledgments

The authors thank George Karniadakis (Brown University) for valuable discussions during this research. This research has been supported by the DARPA grant HR00112290029 with the program “AI-assisted Climate Tipping-point Modeling” managed by Dr. Joshua Elliott. Computing resources for simulations shown in this paper were carried out using the DOE Biological and Environmental Research (BER) Earth System Modeling program’s Compy computing cluster located at Pacific Northwest National Laboratory (PNNL) and using the resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231. PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract DE-AC05-76RL01830.

- Barthel Sorensen, B., Charalampopoulos, A., Zhang, S., Harrop, B. E., Leung, L. R., & Sapsis, T. P. (2024). A non-intrusive machine learning framework for debiasing long-time coarse resolution climate simulations and quantifying rare events statistics. *Journal of Advances in Modeling Earth Systems*, *16*(3), e2023MS004122. <https://doi.org/10.1029/2023MS004122>
- Bruyère, 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*(7–8), 1847–1856. <https://doi.org/10.1007/s00382-013-2011-6>
- Camargo, S. J., Emanuel, K. A., & Sobel, A. H. (2007). Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *Journal of Climate*, *20*(19), 4819–4834. <https://doi.org/10.1175/JCLI4282.1>
- Charalampopoulos, A.-T., Zhang, S., Harrop, B., Leung, L. Y. R., & Sapsis, T. (2023). Statistics of extreme events in coarse-scale climate simulations via machine learning correction operators trained on nudged datasets. <https://doi.org/10.48550/arXiv.2304.02117>
- Chen, J., Arsenault, R., Brissette, F. P., & Zhang, S. (2021). Climate change impact studies: Should we bias correct climate model outputs or post-process impact model outputs? *Water Resources Research*, *57*(5), e2020WR028638. <https://doi.org/10.1029/2020WR028638>
- Chen, J., Brissette, F. P., & Caya, D. (2020). Remaining error sources in bias-corrected climate model outputs. *Climatic Change*, *162*(2), 563–582. <https://doi.org/10.1007/s10584-020-02744-z>
- Christensen, J. H., Boberg, F., Christensen, O. B., & Lucas-Picher, P. (2008). On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters*, *35*(20), L20709. <https://doi.org/10.1029/2008GL035694>
- Colle, B. A., Booth, J. F., & Chang, E. K. (2015). A review of historical and future changes of extratropical cyclones and associated impacts along the US East Coast. *Current Climate Change Reports*, *1*(3), 125–143. <https://doi.org/10.1007/s40641-015-0013-7>
- Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. In T. Stocker et al. (Eds.), *Climate change 2013 - The physical science basis* (pp. 1029–1136). Cambridge University Press. <https://doi.org/10.1017/CBO9781107415324.024>
- Dai, P., & Nie, J. (2022). Robust expansion of extreme midlatitude storms under global warming. *Geophysical Research Letters*, *49*(10), e2022GL099007. <https://doi.org/10.1029/2022GL099007>
- Dennis, J. M., Edwards, J., Evans, K. J., Guba, O., Lauritzen, P. H., Mirin, A. A., et al. (2012). CAM-SE: A scalable spectral element dynamical core for the community atmosphere model. *The International Journal of High Performance Computing Applications*, *26*(1), 74–89. <https://doi.org/10.1177/1094342011428142>
- Deque, M. (2007). Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. *Global and Planetary Change*, *57*(1), 16–26. (Extreme Climatic Events). <https://doi.org/10.1016/j.gloplacha.2006.11.030>
- Dettinger, M. D., Ralph, F. M., & Lavers, D. A. (2015). Setting the stage for a global science of atmospheric rivers. *Eos, Earth and Space Science News*, *96*. <https://doi.org/10.1029/2015EO038675>
- Dixon, K. W., Lanzante, J. R., Nath, M. J., Hayhoe, K., Stoner, A., Radhakrishnan, A., et al. (2016). Evaluating the stationarity assumption in statistically downscaled climate projections: Is past performance an indicator of future results? *Climatic Change*, *135*(3–4), 395–408. <https://doi.org/10.1007/s10584-016-1598-0>
- Done, J. M., Holland, G. J., Bruyère, C. L., Leung, L. R., & Suzuki-Parker, A. (2015). Modeling high-impact weather and climate: Lessons from a tropical cyclone perspective. *Climatic Change*, *129*(3–4), 381–395. <https://doi.org/10.1007/s10584-013-0954-6>
- E3SM Project, D. (2021). Energy exascale earth system model v2.0. [Computer Software]. <https://doi.org/10.11578/E3SM/dc.20210927.1>
- Emanuel, K. A. (2000). A statistical analysis of tropical cyclone intensity. *Monthly Weather Review*, *128*(4), 1139–1152. [https://doi.org/10.1175/1520-0493\(2000\)128\(1139:ASAOTC\)2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128(1139:ASAOTC)2.0.CO;2)
- Emanuel, K. A. (2013). Downscaling CMIP5 climate models shows increased tropical cyclone activity over the 21st century. *Proceedings of the National Academy of Sciences*, *110*(30), 12219–12224. <https://doi.org/10.1073/pnas.1301293110>
- Emanuel, K. A., Ravela, S., Vivant, E., & Risi, C. (2006). A statistical deterministic approach to hurricane risk assessment. *Bulletin of the American Meteorological Society*, *87*(3), 299–314. <https://doi.org/10.1175/BAMS-87-3-299>
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model inter-comparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Feng, L., Smith, S. J., Braun, C., Crippa, M., Gidden, M. J., Hoesly, R., et al. (2020). The generation of gridded emissions data for CMIP6. *Geoscientific Model Development*, *13*(2), 461–482. <https://doi.org/10.5194/gmd-13-461-2020>
- Fink, A. H., Brücher, T., Ermert, V., Krüger, A., & Pinto, J. G. (2009). The European storm Kyrill in January 2007: Synoptic evolution, meteorological impacts and some considerations with respect to climate change. *Natural Hazards and Earth System Sciences*, *9*(2), 405–423. <https://doi.org/10.5194/nhess-9-405-2009>
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., et al. (2013). Evaluation of climate models. In T. F. Stocker et al. (Eds.), *Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change* (pp. 741–882). Cambridge University Press. <https://doi.org/10.1017/CBO9781107415324.020>
- François, B., Vrac, M., Cannon, A. J., Robin, Y., & Allard, D. (2020). Multivariate bias corrections of climate simulations: Which benefits for which losses? *Earth System Dynamics*, *11*(2), 537–562. <https://doi.org/10.5194/esd-11-537-2020>
- Fu, C., Wang, S., Xiong, Z., Gutowski, W. J., Lee, D.-K., McGregor, J. L., et al. (2005). Regional climate model intercomparison project for Asia. *Bulletin of the American Meteorological Society*, *86*(2), 257–266. <https://doi.org/10.1175/BAMS-86-2-257>
- Fulton, D. J., Clarke, B. J., & Hegerl, G. C. (2023). Bias correcting climate model simulations using unpaired image-to-image translation networks. *Artificial Intelligence for the Earth Systems*, *2*(2), e220031. <https://doi.org/10.1175/AIES-D-22-0031.1>
- Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., et al. (1999). An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bulletin of the American Meteorological Society*, *80*(1), 29–56. [https://doi.org/10.1175/1520-0477\(1999\)080\(0029:AOOTRO\)2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080(0029:AOOTRO)2.0.CO;2)
- Gottelman, A., & Morrison, H. (2015). Advanced two-moment bulk microphysics for global models, Part I: Off-line tests and comparison with other schemes. *Journal of Climate*, *28*(3), 1268–1287. <https://doi.org/10.1175/JCLI-D-14-00102.1>
- Giorgi, F., Brodeur, C. S., & Bates, G. T. (1994). Regional climate change scenarios over the United States produced with a nested regional climate model. *Journal of Climate*, *7*(3), 375–399. [https://doi.org/10.1175/1520-0442\(1994\)007\(0375:RCCSOT\)2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007(0375:RCCSOT)2.0.CO;2)
- Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., et al. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2089–2129. <https://doi.org/10.1029/2018MS001603>
- Golaz, J.-C., Larson, V., & Cotton, W. (2002). A pdf-based model for boundary layer clouds. Part I: Method and model description. *Journal of the Atmospheric Sciences*, *59*(24), 3540–3551. [https://doi.org/10.1175/1520-0469\(2002\)059\(3540:APBMBF\)2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059(3540:APBMBF)2.0.CO;2)

- Golaz, J.-C., Van Roekel, L. P., Zheng, X., Roberts, A. F., Wolfe, J. D., Lin, W., et al. (2022). The DOE E3SM model version 2: Overview of the physical model and initial model evaluation. *Journal of Advances in Modeling Earth Systems*, *14*(12), e2022MS003156. <https://doi.org/10.1029/2022MS003156>
- Guan, B., & Waliser, D. E. (2017). Atmospheric rivers in 20 year weather and climate simulations: A multimodel, global evaluation. *Journal of Geophysical Research: Atmospheres*, *122*(11), 5556–5581. <https://doi.org/10.1002/2016JD026174>
- Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen-Skaugen, T. (2012). Technical note: Downscaling RCM precipitation to the station scale using statistical transformations – A comparison of methods. *Hydrology and Earth System Sciences*, *16*(9), 3383–3390. <https://doi.org/10.5194/hess-16-3383-2012>
- Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., et al. (2019). An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the value perfect predictor cross-validation experiment. *International Journal of Climatology*, *39*(9), 3750–3785. <https://doi.org/10.1002/joc.5462>
- Gutowski, W. J., Ullrich, P. A., Hall, A., Leung, L. R., O'Brien, T. A., Patricola, C. M., et al. (2020). The ongoing need for high-resolution regional climate models: Process understanding and stakeholder information. *Bulletin of the American Meteorological Society*, *101*(5), E664–E683. <https://doi.org/10.1175/BAMS-D-19-0113.1>
- Han, L., Chen, M., Chen, K., Chen, H., Zhang, Y., Lu, B., et al. (2021). A deep learning method for bias correction of ECMWF 24–240 h forecasts. *Advances in Atmospheric Sciences*, *38*(9), 1444–1459. <https://doi.org/10.1007/s00376-021-0215-y>
- Hawcroft, M. K., Shaffrey, L. C., Hodges, K. I., & Dacre, H. F. (2012). How much northern hemisphere precipitation is associated with extratropical cyclones? *Geophysical Research Letters*, *39*(24). <https://doi.org/10.1029/2012GL053866>
- Herrington, A. R., Lauritzen, P. H., Reed, K. A., Goldhaber, S., & Eaton, B. E. (2019). Exploring a lower-resolution physics grid in CAM-SE-CLAM. *Journal of Advances in Modeling Earth Systems*, *11*(7), 1894–1916. <https://doi.org/10.1029/2019MS001684>
- 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>
- Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the community emissions data system (CEDS). *Geoscientific Model Development*, *11*(1), 369–408. <https://doi.org/10.5194/gmd-11-369-2018>
- Hoskins, B. J., & Hodges, K. I. (2002). New perspectives on the northern hemisphere winter storm tracks. *Journal of the Atmospheric Sciences*, *59*(6), 1041–1061. [https://doi.org/10.1175/1520-0469\(2002\)059<1041:NPTNH>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<1041:NPTNH>2.0.CO;2)
- Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*(D13), D13103. <https://doi.org/10.1029/2008JD009944>
- Kanada, S., Takemi, T., Kato, M., Yamasaki, S., Fudeyasu, H., Tsuboki, K., et al. (2017). A multimodel intercomparison of an intense typhoon in future, warmer climates by four 5-km-mesh models. *Journal of Climate*, *30*(15), 6017–6036. <https://doi.org/10.1175/JCLI-D-16-0715.1>
- Kanada, S., & Wada, A. (2016). Sensitivity to horizontal resolution of the simulated intensifying rate and inner-core structure of typhoon Ida, an extremely intense typhoon. *Journal of the Meteorological Society of Japan*, *94*(0), 181–190. <https://doi.org/10.2151/jmsj.2015-037>
- Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., & Senior, C. A. (2014). Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change*, *4*(7), 570–576. <https://doi.org/10.1038/nclimate2258>
- Kim, S., Leung, L. R., Guan, B., & Chiang, J. C. H. (2022). Atmospheric river representation in the energy exascale earth system model (E3SM) version 1.0. *Geoscientific Model Development*, *15*(14), 5461–5480. <https://doi.org/10.5194/gmd-15-5461-2022>
- Kitoh, A., & Endo, H. (2016). Changes in precipitation extremes projected by a 20-km mesh global atmospheric model. *Weather and Climate Extremes*, *11*, 41–52. <https://doi.org/10.1016/j.wace.2015.09.001>
- Knutson, T. R., Camargo, S. J., Chan, J. C. L., Emanuel, K. A., Ho, C.-H., Kossin, J., et al. (2019). Tropical cyclones and climate change assessment: Part I: Detection and attribution. *Bulletin of the American Meteorological Society*, *100*(10), 1987–2007. <https://doi.org/10.1175/BAMS-D-18-0189.1>
- Knutson, T. R., Camargo, S. J., Chan, J. C. L., Emanuel, K. A., Ho, C.-H., Kossin, J., et al. (2020). Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bulletin of the American Meteorological Society*, *101*(3), E303–E322. <https://doi.org/10.1175/BAMS-D-18-0194.1>
- Knutson, T. R., Sirutis, J. J., Vecchi, G. A., Garner, S., Zhao, M., Kim, H.-S., et al. (2013). Dynamical downscaling projections of twenty-first-century Atlantic hurricane activity: CMIP3 and CMIP5 model-based scenarios. *Journal of Climate*, *26*(17), 6591–6617. <https://doi.org/10.1175/JCLI-D-12-00539.1>
- 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>
- Larson, V. E., Golaz, J.-C., & Cotton, W. R. (2002). A pdf-based model for boundary layer clouds. Part I: Method and model description. *Journal of the Atmospheric Sciences*, *59*(24), 3519–3539. [https://doi.org/10.1175/1520-0469\(2002\)059<3519:SSAMVI>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<3519:SSAMVI>2.0.CO;2)
- Lauritzen, P. H., Nair, R. D., Herrington, A. R., Callaghan, P., Goldhaber, S., Dennis, J. M., et al. (2018). NCAR release of CAM-SE in CESM2.0: A reformulation of the spectral element dynamical core in dry-mass vertical coordinates with comprehensive treatment of condensates and energy. *Journal of Advances in Modeling Earth Systems*, *10*(7), 1537–1570. <https://doi.org/10.1029/2017MS001257>
- Lee, C.-Y., Camargo, S. J., Sobel, A. H., & Tippett, 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>
- Leung, L. R., Bader, D. C., Taylor, M. A., & McCoy, R. B. (2020). An introduction to the E3SM special collection: Goals, science drivers, development, and analysis. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2019MS001821. <https://doi.org/10.1029/2019ms001821>
- Liu, X., Ma, P.-L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., et al. (2016). Description and evaluation of a new four–mode version of the modal aerosol module (MAM4) within version 5.3 of the Community Atmosphere Model. *Geoscientific Model Development*, *9*(2), 505–522. <https://doi.org/10.5194/gmd-9-505-2016>
- Lucas-Picher, P., Argüeso, D., Brisson, E., Trambly, Y., Berg, P., Lemonsu, A., et al. (2021). Convection-permitting modeling with regional climate models: Latest developments and next steps. *WIREs Climate Change*, *12*(6), e731. <https://doi.org/10.1002/wcc.731>
- McClenny, E. E., Ullrich, P. A., & Grotjahn, R. (2020). Sensitivity of atmospheric river vapor transport and precipitation to uniform sea surface temperature increases. *Journal of Geophysical Research: Atmospheres*, *125*(21), e2020JD033421. <https://doi.org/10.1029/2020JD033421>
- Merz, B., Kuhllicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D. N., et al. (2020). Impact forecasting to support emergency management of natural hazards. *Reviews of Geophysics*, *58*(4), e2020RG000704. <https://doi.org/10.1029/2020RG000704>
- Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*(D14), 16663–16682. <https://doi.org/10.1029/97JD00237>

- Mo, R., So, R., Brugman, M. M., Mooney, C., Liu, A. Q., Jakob, M., et al. (2021). Column relative humidity and primary condensation rate as two useful supplements to atmospheric river analysis. *Water Resources Research*, 57(11), e2021WR029678. <https://doi.org/10.1029/2021WR029678>
- Moghim, S., & Bras, R. L. (2017). Bias correction of climate modeled temperature and precipitation using artificial neural networks. *Journal of Hydrometeorology*, 18(7), 1867–1884. <https://doi.org/10.1175/jhm-d-16-0247.1>
- Moon, H., Gudmundsson, L., & Seneviratne, S. I. (2018). Drought persistence errors in global climate models. *Journal of Geophysical Research: Atmospheres*, 123(7), 3483–3496. <https://doi.org/10.1002/2017JD027577>
- Mori, N., Shimura, T., Yoshida, K., Mizuta, R., Okada, Y., Fujita, M., et al. (2019). Future changes in extreme storm surges based on mega-ensemble projection using 60-km resolution atmospheric global circulation model. *Coastal Engineering Journal*, 61(3), 295–307. <https://doi.org/10.1080/21664250.2019.1586290>
- Morrison, H., & Gettelman, A. (2008). A new two-moment bulk stratiform cloud microphysics scheme in the community atmosphere model, version 3 (CAM3). Part I: Description and numerical tests. *Journal of Climate*, 21(15), 3642–3659. <https://doi.org/10.1175/2008JCLI2105.1>
- Mueller, B., & Seneviratne, S. I. (2014). Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophysical Research Letters*, 41(1), 128–134. <https://doi.org/10.1002/2013GL058055>
- Nie, J., Sobel, A. H., Shaevitz, D. A., & Wang, S. (2018). Dynamic amplification of extreme precipitation sensitivity. *Proceedings of the National Academy of Sciences*, 115(38), 9467–9472. <https://doi.org/10.1073/pnas.1800357115>
- Oleson, K. W., Lawrence, M., Bonan, B., Drewniak, B. A., Huang, M., Koven, D., et al. (2013). Technical description of version 4.5 of the community land model (CLM). NCAR Technical Note NCAR/TN-503+STR, No. 420pp.) NCAR, USA. <https://doi.org/10.5065/D6RR1W7M>
- O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The scenario model intercomparison project (scenariosmip) for CMIP6. *Geoscientific Model Development*, 9(9), 3461–3482. <https://doi.org/10.5194/gmd-9-3461-2016>
- Payne, A. E., Demory, M.-E., Leung, L. R., Ramos, A. M., Shields, C. A., Rutz, J. J., et al. (2020). Responses and impacts of atmospheric rivers to climate change. *Nature Reviews Earth & Environment*, 1(3), 143–157. <https://doi.org/10.1038/s43017-020-0030-5>
- Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *Journal of Climate*, 15(13), 1609–1625. [https://doi.org/10.1175/1520-0442\(2002\)015<1609:AIISAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2)
- Schär, C., Frei, C., Lüthi, D., & Davies, H. C. (1996). Surrogate climate-change scenarios for regional climate models. *Geophysical Research Letters*, 23(6), 669–672. <https://doi.org/10.1029/96GL00265>
- Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., et al. (2012). Changes in climate extremes and their impacts on the natural physical environment. In C. B. Field, V. Barros, T. F. Stocker, & Q. Dahe (Eds.), *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the intergovernmental panel on climate change* (pp. 109–230). Cambridge University Press. <https://doi.org/10.1017/CBO9781139177245.006>
- Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., et al. (2023). Weather and climate extreme events in a changing climate. In V. Masson-Delmotte et al. (Eds.), *Climate change 2021 – The physical science basis: Working group I contribution to the sixth assessment report of the intergovernmental panel on climate change* (pp. 1513–1766). Cambridge University Press. <https://doi.org/10.1017/9781009157896.013>
- Shields, C. A., Rutz, J. J., Leung, L.-Y., Ralph, F. M., Wehner, M., Kawzenuk, B., et al. (2018). Atmospheric river tracking method inter-comparison project (ARTMIP): Project goals and experimental design. *Geoscientific Model Development*, 11(6), 2455–2474. <https://doi.org/10.5194/gmd-11-2455-2018>
- Sinclair, V. A., Rantanen, M., Haapanala, P., Räisänen, J., & Järvinen, H. (2020). The characteristics and structure of extra-tropical cyclones in a warmer climate. *Weather and Climate Dynamics*, 1(1), 1–25. <https://doi.org/10.5194/wcd-1-1-2020>
- Steininger, M., Abel, D., Ziegler, K., Krause, A., Paeth, H., & Hotho, A. (2020). Deep learning for climate model output statistics. <https://doi.org/10.48550/arXiv.2012.10394>
- Sun, J., Zhang, K., Wan, H., Ma, P.-L., Tang, Q., & Zhang, S. (2019). Impact of nudging strategy on the climate representativeness and hindcast skill of constrained EAMv1 simulations. *Journal of Advances in Modeling Earth Systems*, 11(12), 3911–3933. <https://doi.org/10.1029/2019MS001831>
- Taylor, M. A., & Fournier, A. (2010). A compatible and conservative spectral element method on unstructured grids. *Journal of Computational Physics*, 229(17), 5879–5895. <https://doi.org/10.1016/j.jcp.2010.04.008>
- Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, 456–457, 12–29. <https://doi.org/10.1016/j.jhydrol.2012.05.052>
- Trenberth, E., Berry, C. W., & Buja, E. (1993). Vertical interpolation and truncation of model-coordinate data (NCAR Technical Report NCAR/TN-396+STR). <https://doi.org/10.5065/D6HX19NH>
- Ullrich, P. (2022). Tempestextremes v2.2.1 [Computer Software]. <https://github.com/ClimateGlobalChange/tempestextremes/releases/tag/v2.2.1>
- Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., & Reed, K. A. (2021). Tempestextremes v2.1: A community framework for feature detection, tracking, and analysis in large datasets. *Geoscientific Model Development*, 14(8), 5023–5048. <https://doi.org/10.5194/gmd-14-5023-2021>
- Vaittinada Ayar, P., Vrac, M., & Mailhot, A. (2021). Ensemble bias correction of climate simulations: Preserving internal variability. *Scientific Reports*, 11(1), 3098. <https://doi.org/10.1038/s41598-021-82715-1>
- Volosciuk, C., Maraun, D., Semenov, V. A., & Park, W. (2015). Extreme precipitation in an atmosphere general circulation model: Impact of horizontal and vertical model resolutions. *Journal of Climate*, 28(3), 1184–1205. <https://doi.org/10.1175/JCLI-D-14-00337.1>
- Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L., & Somot, S. (2012). Dynamical and statistical downscaling of the French mediterranean climate: Uncertainty assessment. *Natural Hazards and Earth System Sciences*, 12(9), 2769–2784. <https://doi.org/10.5194/nhess-12-2769-2012>
- Wang, F., & Tian, D. (2022). On deep learning-based bias correction and downscaling of multiple climate models simulations. *Climate Dynamics*, 59(11–12), 3451–3468. <https://doi.org/10.1007/s00382-022-06277-2>
- Wang, H., Easter, R. C., Zhang, R., Ma, P.-L., Singh, B., Zhang, K., et al. (2020). Aerosols in the E3SM version 1: New developments and their impacts on radiative forcing. *Journal of Advances in Modeling Earth Systems*, 12(1), e2019MS001851. <https://doi.org/10.1029/2019MS001851>
- Wang, Y., Liu, X., Hoose, C., & Wang, B. (2014). Different contact angle distributions for heterogeneous ice nucleation in the community atmospheric model version 5. *Atmospheric Chemistry and Physics*, 14(19), 10411–10430. <https://doi.org/10.5194/acp-14-10411-2014>
- Wehrli, K., Guillod, B. P., Hauser, M., Leclair, M., & Seneviratne, S. I. (2018). Assessing the dynamic versus thermodynamic origin of climate model biases. *Geophysical Research Letters*, 45(16), 8471–8479. <https://doi.org/10.1029/2018GL079220>

- Wilks, D. S. (2016). "The stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it. *Bulletin of the American Meteorological Society*, 97(12), 2263–2273. <https://doi.org/10.1175/BAMS-D-15-00267.1>
- Willison, J., Robinson, W. A., & Lackmann, G. M. (2013). The importance of resolving mesoscale latent heating in the North Atlantic storm track. *Journal of the Atmospheric Sciences*, 70(7), 2234–2250. <https://doi.org/10.1175/JAS-D-12-0226.1>
- Xu, W., Balaguru, K., August, A., Lalo, N., Hodas, N., DeMaria, M., & Judi, D. (2021). Deep learning experiments for tropical cyclone intensity forecasts. *Weather and Forecasting*, 36(4), 1453–1470. <https://doi.org/10.1175/WAF-D-20-0104.1>
- Xu, Z., Han, Y., Tam, C.-Y., Yang, Z.-L., & Fu, C. (2021). Bias-corrected CMIP6 global dataset for dynamical downscaling of the historical and future climate (1979–2100). *Scientific Data*, 8(1), 293. <https://doi.org/10.1038/s41597-021-01079-3>
- Xu, Z., & Yang, Z.-L. (2012). An improved dynamical downscaling method with GCM bias corrections and its validation with 30 years of climate simulations. *Journal of Climate*, 25(18), 6271–6286. <https://doi.org/10.1175/JCLI-D-12-00005.1>
- Xue, Z., Ullrich, P., & Leung, L.-Y. R. (2023). Sensitivity of the pseudo-global warming method under flood conditions: A case study from the northeastern US. *Hydrology and Earth System Sciences*, 27(9), 1909–1927. <https://doi.org/10.5194/hess-27-1909-2023>
- Zappa, G., Shaffrey, L. C., Hodges, K. I., Sansom, P. G., & Stephenson, D. B. (2013). A multimodel assessment of future projections of North Atlantic and European extratropical cyclones in the CMIP5 climate models. *Journal of Climate*, 26(16), 5846–5862. <https://doi.org/10.1175/JCLI-D-12-00573.1>
- Zarzycki, C. M., & Ullrich, P. A. (2017). Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. *Geophysical Research Letters*, 44(2), 1141–1149. <https://doi.org/10.1002/2016GL071606>
- Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model. *Atmosphere-Ocean*, 33(3), 407–446. <https://doi.org/10.1080/07055900.1995.9649539>
- Zhang, S., & Charalampopoulos, A.-T. (2024). Analysis scripts and figure dataset for Zhang et al. (2024) [Dataset]. *Zenodo*. <https://doi.org/10.5281/zenodo.11053624>
- Zhang, S., Zhang, K., Wan, H., & Sun, J. (2022). Further improvement and evaluation of nudging in the E3SM atmosphere model version 1 (EAMv1): Simulations of the mean climate, weather events, and anthropogenic aerosol effects. *Geoscientific Model Development*, 15(17), 6787–6816. <https://doi.org/10.5194/gmd-15-6787-2022>