A Machine Learning Bias Correction of Large-scale Environment of Extreme Weather Events in E3SM Atmosphere Model

Shixuan Zhang[1], Bryce Harrop [1], L. Ruby Leung[1], Alexis-Tzianni Charalampopoulos[2], Benedikt Barthel[2], Wenwei Xu[1], Themistoklis Sapsis[2]

[1]Pacific Northwest National Laboratory, Richland, WA 99354, USA
[2]Department of Mechanical Engineering, Massachusetts Institute of Technology,Cambridge, MA 02139, 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 extreme weather events and their future changes.

Corresponding author: L. Ruby Leung, Shixuan Zhang, `Ruby.Leung@pnnl.gov`, `Shixuan.Zhang@pnnl.gov`

Abstract

Large-scale dynamical and thermodynamical processes are common environmental drivers of extreme weather events. However, such large-scale environmental conditions often display systematic biases in climate simulations, posing challenges to evaluating extreme weather events and associated risks in current and future climate. In this paper, a machine learning (ML) approach was employed to bias correct the large-scale wind, temperature, and humidity simulated by the E3SM atmosphere model at $\sim 1^\circ$ resolution. The usefulness of the proposed ML approach for extreme weather analysis was demonstrated with a focus on three extreme weather events, 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 the water vapor transport associated with ARs, and the 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 extreme events. 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 extreme weather events. 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 extreme 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 extreme storms such as atmospheric rivers (ARs), tropical cyclones (TCs) and extratropical cyclones (ETCs) which have substantial societal and economic impacts (S. Seneviratne et al., 2012; Angélil et al., 2016; Guan & Waliser, 2017; Moon et al., 2018; Wehrli et al., 2018; Merz et al., 2020; Dai & Nie, 2022; S. I. Seneviratne et al., 2023). However, achieving a proper representation of these extreme events 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., Willison et al., 2013; Kendon et al., 2014; Kitoh & Endo, 2016; Kanada & Wada, 2016; Kanada et al., 2017; Lucas-Picher et al., 2021; Nie et al., 2018; Mori et al., 2019). Consequently, downscaling tech-

niques 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 extreme storms (Emanuel et al., 2006; Emanuel, 2013; 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 extreme weather events through established statistical relationships (e.g, Emanuel, 2013; Colle et al., 2015; Dixon et al., 2016; Balaguru et al., 2023a), or to provide boundary conditions for limited area or regional models to simulate the local climate and extreme weather events (e.g, Giorgi et al., 1994; Fu et al., 2005; 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 storms. However, the large-scale storm environments governing regional- to local-scale extreme events 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; Zappa et al., 2013; Volosciuk et al., 2015). 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, Deque, 2007; Christensen et al., 2008; Z. Xu & Yang, 2012; Vrac et al., 2012; Gudmundsson et al., 2012; François et al., 2020; W. Xu et al., 2021; Vaittinada Ayar et al., 2021). Studies have demonstrated that correction of the GCM mean bias may improve dynamical downscaling of local-scale extreme weather events 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 application of modern artificial neural network architectures in bias correction and statistical downscaling of GCMs (e.g., Moghim & Bras, 2017; Steininger et al., 2020; Han et al., 2021; Z. Xu et al., 2021; F. Wang & Tian, 2022; Fulton et al., 2023). Several types of ML approaches have proven to successfully reduce spatial- and temporal- biases in GCMs (F. Wang & Tian, 2022; Fulton et al., 2023). In this paper, we introduce a long short-term memory neural network (LSTM) machine learning (ML) approach (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 modeling of extreme weather events, we evaluate the impact of ML bias correction on the large-scale storm environments and the long-term statistics of extreme weather events simulated by EAMv2, with a focus on three types of extreme weather events, including ARs, TCs and ETCs. Importantly, we also evaluate the impact of the ML bias correction on the responses of the three types of extreme weather events to future climate change. The goal of this study is to determine how well low-resolution climate models with the 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 and downscaling 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 6

2 Model and Simulations

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 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 ∼110km 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, Lauritzen et al., 2018; Herrington et al., 2019). 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 (Morrison & Gettelman, 2008; Gettelman & Morrison, 2015; 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 m to 100 m near the surface and up to 600 m near the model top.

2.2 Simulation and Bias Correction

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 pseudo global warming (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 datasets (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 Programme (WCRP) Coupled Model Intercomparison Project-Phase 6 (CMIP6, Eyring et al., 2016; Hoesly et al., 2018; Feng et al., 2020). 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 two PGW simulations, SSP245 and SSP585, use the same configurations as for CLIM, except with imposed climate change perturbations added to the boundary conditions of SST and SIC:

- **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-year period, and the ΔSST and ΔSIC 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 ΔSST and ΔSIC 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 ΔSST and ΔSIC were added to each corresponding month to preserve the monthly and seasonal cycles of ΔSST and ΔSIC. 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 Figs. A1a–b).

- **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 pseudo global warming 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 Figs. A1c–d).

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 et al. (2023) in which a neural network (NN) operator acts on the coarse-resolution climate model simulation in a postprocessing manner. Specifically, the NN operator was trained to learn a map between a *nudged* coarse resolution EAMv2 historical simulation and the ERA5 reanalysis data (Hersbach et al., 2020) that represents the real atmosphere. The nudged simulation was conducted by constraining the large-scale model state (i.e. U, V, T, Q) towards the ERA5 reanalysis following S. Zhang et al. (2022) and Sun et al. (2019) so that the trajectory of model state predominately obeys the dynamics of the coarse resolution EAMv2 model, yet is constrained from 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. 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. A full description of the mathematical framework for the ML model is included in appendix A2. More detailed information for the development and verification of the ML model can be found in Barthel et al. (2023).

In this study, the 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. The ML bias-corrected data is referred to as ML(CLIM), ML(SSP245), and ML(SSP585) in Table 1 and throughout the rest of this manuscript. A comparison between the Group 1 and Group 2 simulations is performed 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. It's important to note that none of the climate simulations for the historical or future scenarios in Group 1 were used in the ML training. As mentioned above, ML training was performed using an EAMv2 simulation with nudging towards ERA5, which is different from the free running histor-

Table 1. List of simulations for present-day and future climate scenarios. The three simulations in Group 1 were conducted with the coupled atmosphere-land components of EAMv2 with default "low-resolution" configuration at $\sim$ 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.

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

ical simulation (CLIM). In addition, 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.

3 Validation of machine learning (ML) bias correction

3.1 Bias correction on historical simulation

We first evaluated the ability of the ML approach to reduce the model biases in EAMv2. 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)(Figs. 1b, e and h) are systematically smaller than those in CLIM (Figs. 1a, d and g) 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 to the wind field (Fig. 1a–b), more significant bias reductions are seen in the temperature (Fig. 1d–e) and humidity fields (Fig. 1g–h). In addition, more promising bias corrections are seen in the near-surface levels relative to the upper model levels in all fields over most regions over the globe (also see Fig B1 in Appendix). This is possibly because of the large variabilities in the fields such as wind in upper model layers (see Fig. 1c), which could limit the skills of ML bias correction model in these regions.

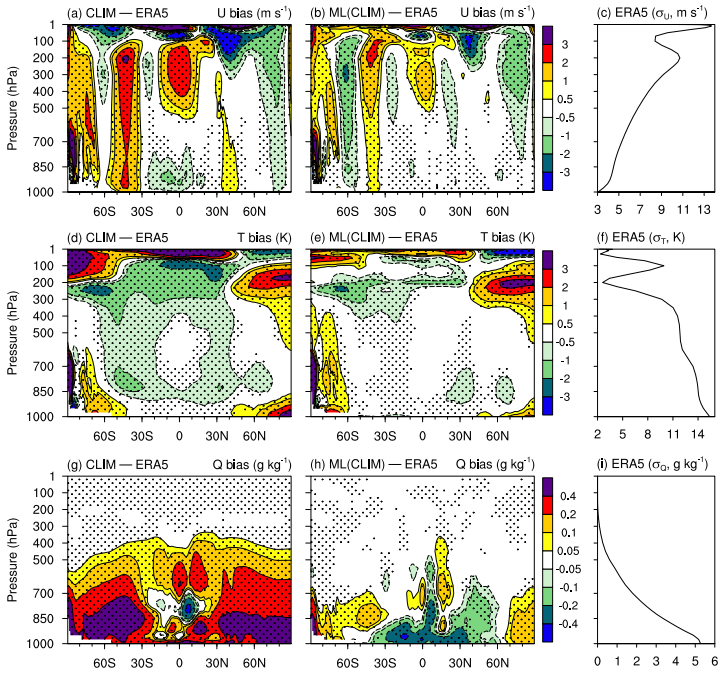


Figure 1. Zonal and annual mean model biases in zonal wind (U, unit: m s^{-1}, panels a–b), temperature (T , unit: K, panels d–e), and water vapor mixing ratio (Q, unit: g kg^{-1}, panels g–h) averaged over the period of 1979-2014 from free-running (i.e., CLIM, first column) and ML bias-corrected (i.e., ML(CLIM), second column) EAMv2 simulations. The biases are derived by comparing monthly mean EAMv2 simulations with ERA5 reanalysis. The dotted region indicates 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 interpolation is used to regrid the EAMv2 data on the hybrid sigma-pressure level to pressure-level for comparison with ERA5 reanalysis. Details of the simulation setups can be found in Table 1.

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 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 (Fig. 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 (Fig. 2c–d). 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.

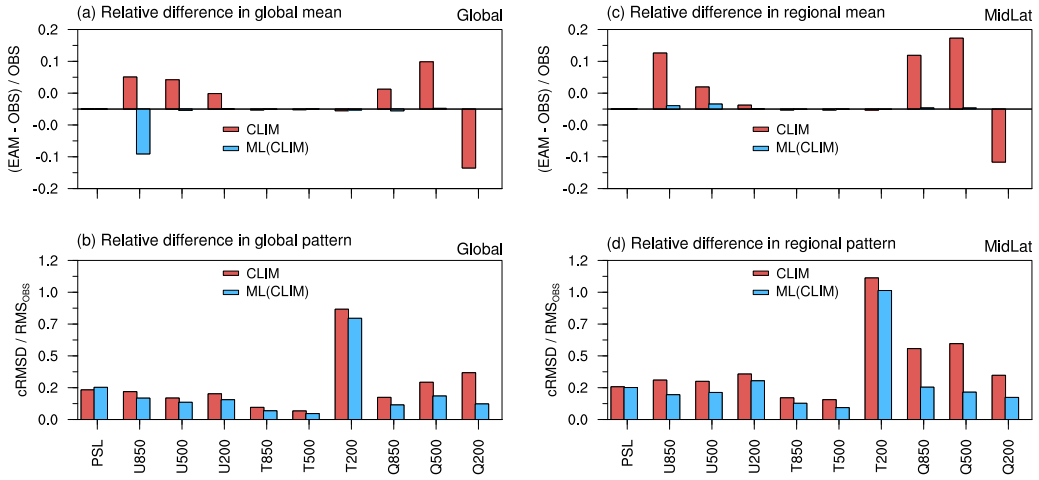


Figure 2. First row: mean biases in selected physical quantities averaged over the globe (panel a) and mid-latitude region ([30-60N] and [30-60S] latitude bands, panel c) from EAMv2 simulations without (i.e "CLIM" in red) and with (i.e."ML(CLIM)" in blue) ML bias correction, normalized by the observed 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) during the period of 1979-2014. The y-axis shows the selected physical quantities, includingsurface 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.

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. Gutiérrez et al., 2019; Teutschbein & Seibert, 2012; Chen et al., 2020, 2021) 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 climate 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 pseudo global warming 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.

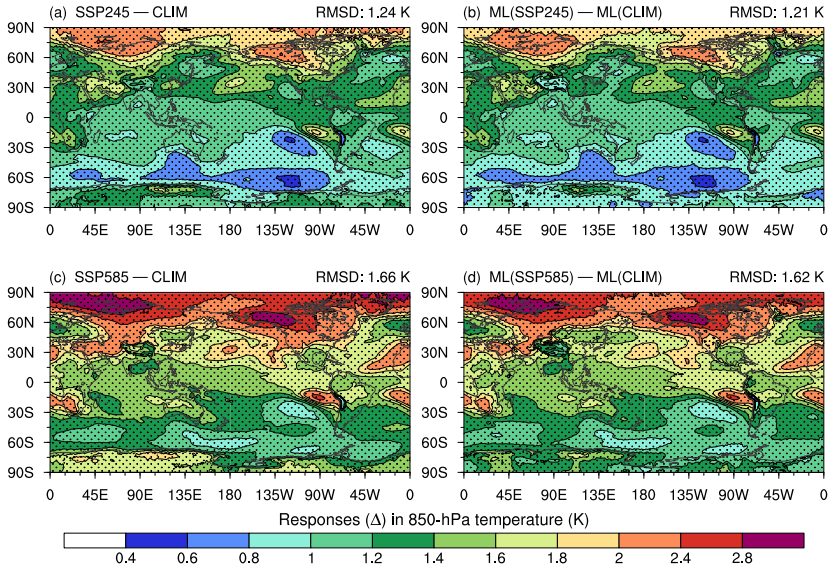


Figure 3. Differences of air temperature at 850-hPa (unit: K) between the present-day and pseudo global warming EAMv2 simulations averaged over the whole simulation period of 1979-2014. Shown are (a) SSP245 – CLIM and (c) SSP585 – CLIM for EAMv2 simulations without ML bias correction, as well as (b) 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 pseudo global warming 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.

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 (Figure 3a and c) are largely consistent with the pseudo global warming perturbations of SST (Figure A1a and c). 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 (Figs. 3a and c), ML bias correction applied to CLIM and the PGW simulations overall does not significantly modify the patterns and magnitude of 850-hPa temperature responses in most regions over the globe (Figs. 3b and d). 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. Fig. 3c versus Fig. 3d). Further analysis indicates that the ML model applies correction to the large cold temperature biases over these regions (see. Fig. 1d–f) 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.

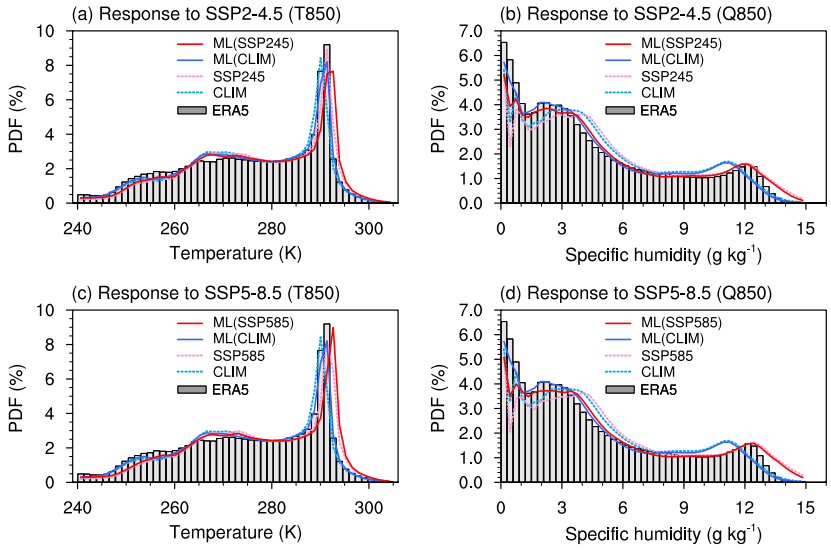


Figure 4. Long-term statistics for monthly mean air temperature (first column, unit: K) and specific humidity (second column, unit: g kg^{-1}) at 850-hPa pressure levels over the whole global domain during the simulation period of 1979-2014. Shown is the comparison among ERA5 reanalysis (grey bars), 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 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.

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 towards the PDF of ERA5 reanalysis data (dashed and solid blue lines versus grey 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 versus blue lines). Therefore, the physical climate change effects, i.e. 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 datasets (dashed blue and red lines,

see also Fig B2), suggesting that the ML bias correction does not significantly interfere with the pseudo global warming-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 (grey bars), compared with the uncorrected free-running simulations (i.e. SSP245 and SSP585, dashed red lines). This implies that the 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 (Figure. 4 b and d). 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 extreme weather events to analyze: atmospheric rivers (ARs), extratropical cyclones (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 B. These three types of events are of interest partly because they represent extreme weather events that 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, on landfall, produce precipitation that can be both beneficial and destructive (Payne et al., 2020). The major features of ARs are reflected by the zonal and meridional moisture transport (model output fields TUQ and TVQ, respectively). As shown in Eq. (C5), TUQ and TVQ are directly linked to large-scale wind and specific humidity, and are expected to be well simulated by EAMv2 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. The occurrence frequency 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 (Fig. 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 (Fig. 5c). The remaining model biases in the composite IVT field are statistically insignificant in most regions over the globe. Following Eq. (C5), such improvements in ARs are obtained because of the effective bias reductions by the ML model in both large-scale wind and humidity fields (Fig. 1h and Fig. 1i). Consistently, the AR annual occurrence frequency also agrees better with

ERA5 reanalysis after the ML bias correction (Fig. 5e versus f). Note that TempestExtremes uses the Laplacian of IVT instead of an IVT threshold for AR tracking. Therefore, biases in large-scale humidity on its own is not responsible for the AR frequency biases in Fig. 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.

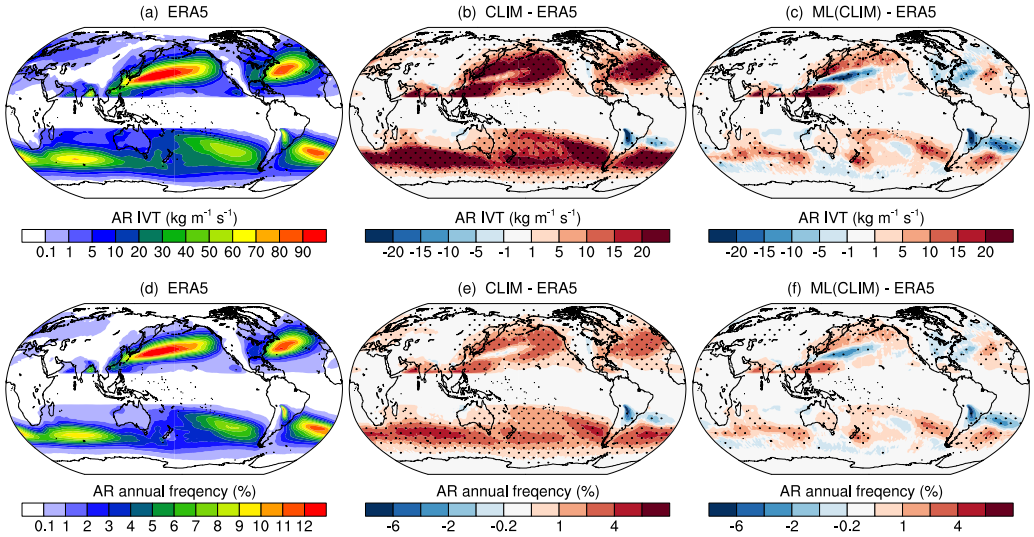


Figure 5. Top row: distribution of the vertically integrated horizontal water vapor transport (IVT, units kg m^{-1} s^{-1}) from ERA5 reanalysis averaged over all identified AR events at each grid point during 1979-2014 (panel a), and the model biases in the EAMv2 simulations without (i.e., CLIM, panel 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-hour) a grid point was part of an AR, divided by the total number of 6-hour 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.

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 pseudo global warming EAMv2 simulations. The differences between SSP245 and CLIM suggest an increase of the IVT (Fig. 6a) and the occurrence of AR events (Fig. 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 sim-

ulations, 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 pseudo global warming 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 B3). 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 pseudo global warming simulations. By eliminating the systematic model biases, the bias-corrected AR environments provide more reliable information for downscaling of ARs for assessing future changes in precipitation and flood hazards associated with ARs.

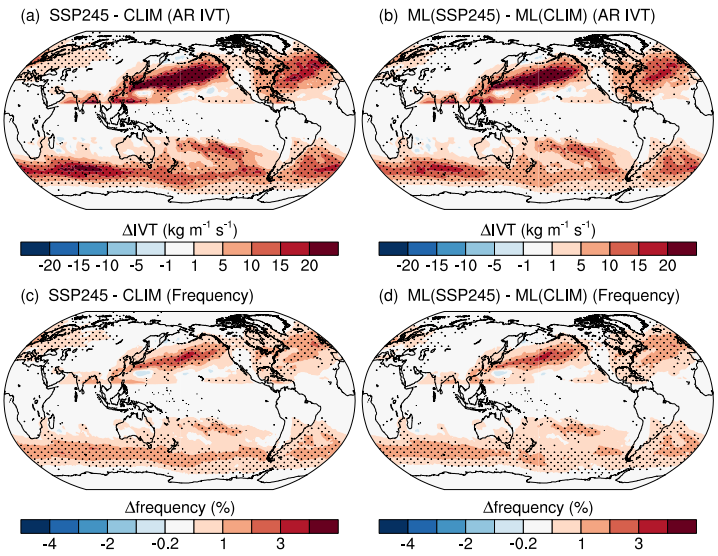


Figure 6. Changes of the vertically integrated horizontal water vapor transport (IVT, units kg m^{-1} s^{-1}, top row) and annual occurrence frequency of ARs (unit: %) in the EAMv2 simulations with imposed climate change perturbations in sea-surface temperature (SST) and sea-ice concentration(SIC) for the SSP2-4.5 scenarios. Shown is 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.

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). The heavy precipitation and strong winds accompanying ETCs are known to cause extreme weather-induced damages in mid-latitude 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

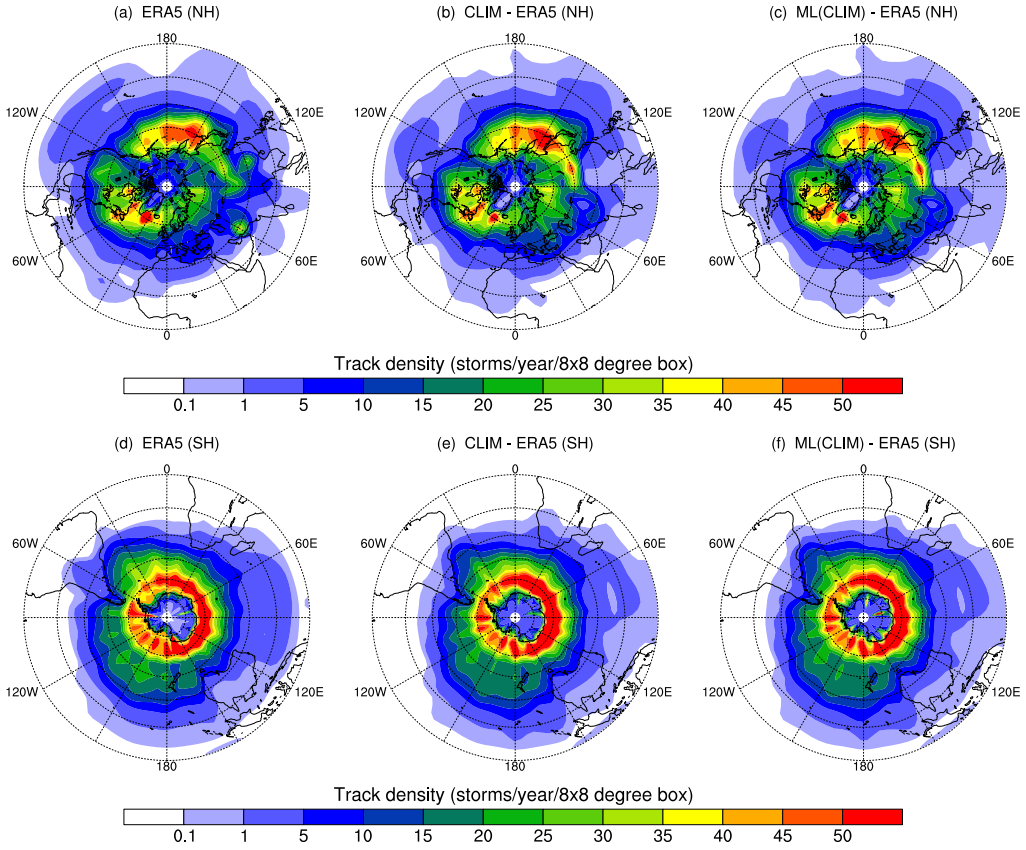


Figure 7. Track density maps for total annual ETCs over Northern Hemisphere (NH, top row) and Southern Hemisphere (SH, bottom row) tracked in the ERA5 reanalysis (panels a and d) and EAMv2 climate simulations without (i.e, CLIM, panels b and e) and with (i.e., ML(CLIM), panels c and f) ML bias correction. The ETC events and composite IVT are tracked with the TempestExtremes with 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-hour) the ETCs passed the 8x8 ^{o} grid box over the globe in each year during 1979-2014. Units are the number of 6-hourly ETC occurrences per $8° \times 8°$ grid box per year.

with TempestExtremes using the sea level pressure (PSL) fields in the two hemispheres (Figure 7). The annual ETC storm tracks over the Northern Hemisphere (NH) in ERA5 reanalysis (Fig 7a) 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°-70° S (Fig. 7d). The low-resolution

EAMv2 model produces a good representation of the observed spatial patterns of ETC track densities in both hemispheres (Fig. 7b and e). Compared with the free-running simulations (i.e. CLIM, Fig. 7b and e), no significant differences in the ETC track densities are seen in simulations with the ML bias correction (i.e. ML(CLIM), Fig. 7c and f). 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 Eq. C1). As shown in Fig. B4, the systematic model biases of PSL in CLIM are less than 2-hPa in most regions over the globe (Fig. B4b). Therefore, we will not expect a strong correction from the ML model in these regions with small PSL biases (Fig. B4c). However, the CLIM simulation indeed reveals large low-pressure biases in the southern ocean region ($50°-70°S$) (Fig. B4b), which is co-located with the highest ETC track density region over Southern Hemisphere (Fig. 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 (Fig. B4c).

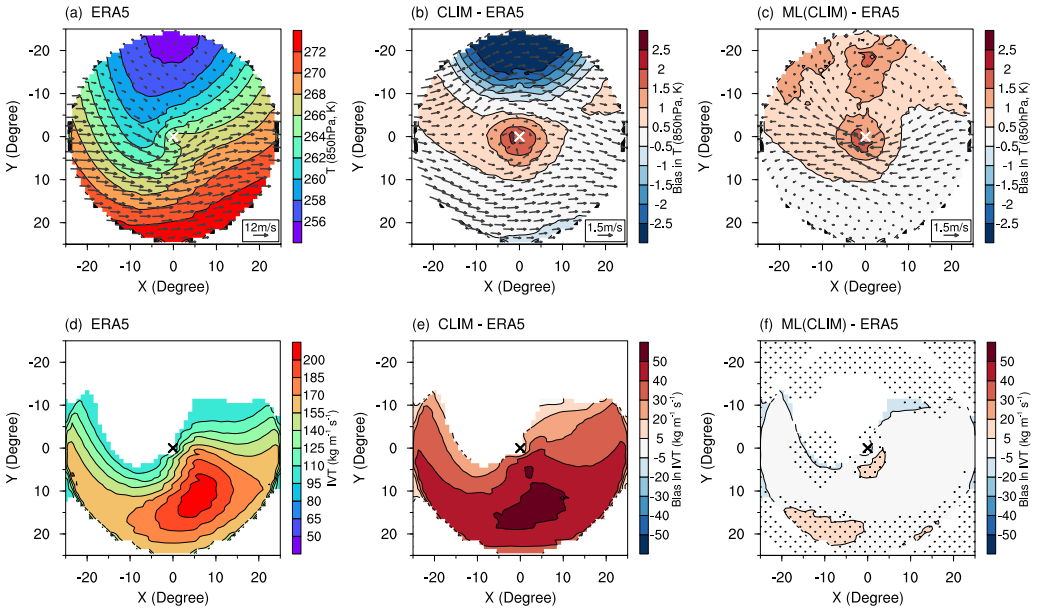


Figure 8. Composites of meteorological quantities centered on ETC storm center of all filtered storms with mean sea level pressure (PSL) less than or equal to 990-hPa in the ERA5 reanalysis (first column) and the differences between EAMv2 simulations and ERA5 reanalysis before (i.e. CLIM, second column) and after (i.e. ML(CLIM), third column) applying ML bias correction. 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; 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.

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 (Fig. 8a–

c) and integrated vapor transport (IVT, Fig. 8d–f). The advection of warm and moist air wrapping cyclonically around the eastern side of the storm center is clearly seen in the ERA5 reanalysis (Fig. 8a and d). 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 (Fig. 8b and e). A sizable cold temperature bias is also pronounced on the north side of the composite storm (Fig. 8b). For ML(CLIM) with the ML bias correction, the biases in the temperature, wind, and vapor transport are clearly reduced in the composite storms (Fig. 8c and f), with more significant improvements in the integrated vapor transport fields (Fig. 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 (Fig. 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 Fig. 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 less sea-ice concentrations leads to a reduction of the storm track density around the Arctic (Fig. 9a). Stronger signals in the ETC track density responses are observed in the simulations with the higher-emission climate change scenario of SSP5-85 (Fig. 9c and Fig. B5c). In addition, the signals of the ETC track density in response to the climate change in simulations with ML bias correction highly agree with the free-running EAMv2 simulations (Fig. 9b versus (Fig. 9a)). Similar conclusions can be drawn for ETC track densities over the Southern Hemisphere (SH, Figure B5)

For the responses of the ETC intensity, Figure 10 shows the changes of composited mean sea level pressure 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 (Fig. 10a) and SSP5-8.5 (Fig. 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 Fig. 11a and c). Meanwhile, there is an enhanced cyclonic circulation in the boundary layer regions (i.e. 850-hPa) due to climate change (vectors in Fig. 11a and c). 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 (i.e. Fig. 10b and d for ML-SSP245, and Fig. 11b and d for ML-SSP585) largely agree with those in the free-running pseudo global warming simulations (i.e. Fig. 10a and c for SSP245; and Fig. 11a and c for SSP585). 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 (Fig.10b and d), compared with those in free-running simulations (Fig.10a and

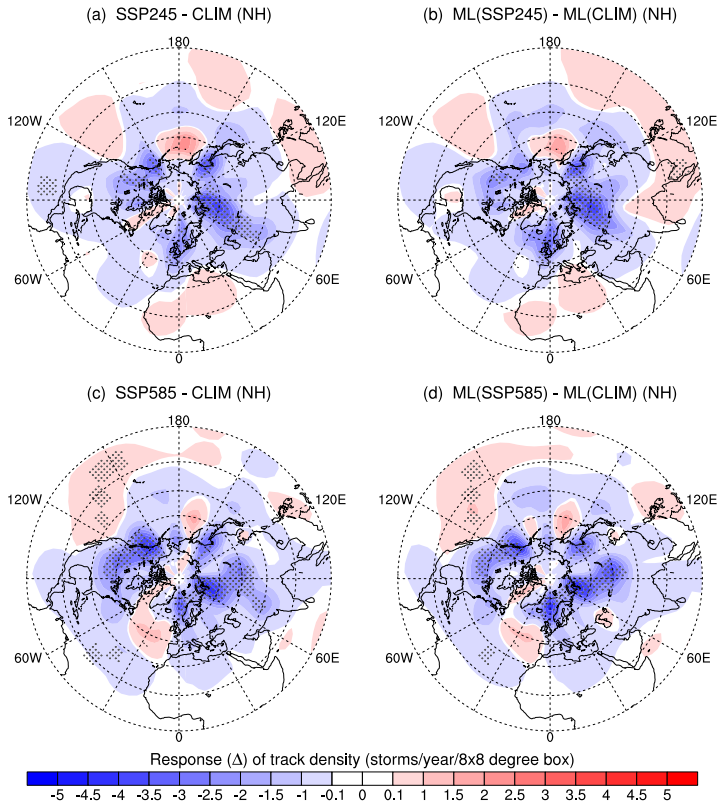


Figure 9. Responses (Δ) of annual ETC track densities over North Hemisphere (NH) to the imposed climate change perturbations in sea-surface temperature (SST) and sea-ice concentration(SIC) for the SSP2-4.5 (top row) and the SSP5-8.5 scenarios (bottom row). Shown is the differences of SSP245 – CLIM (panel 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 during the period of 1979-2014. The annual track density is defined as the total number of time steps (6-hour) the ETCs passed over an 8x8 ^{o} grid box per year. The dotted region indicates that the differences are significant at a 95% significance level.

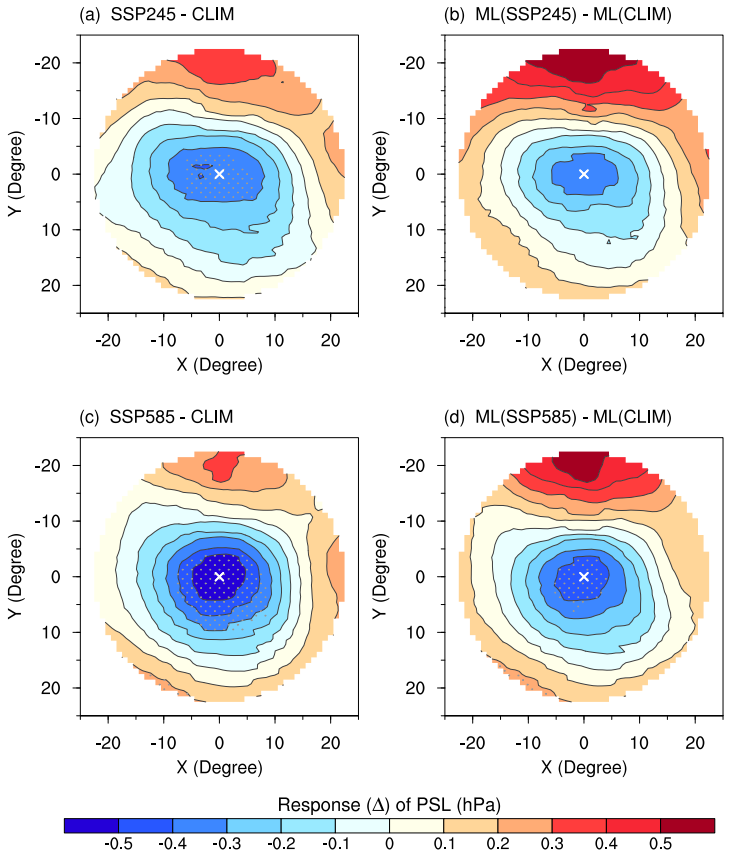


Figure 10. Responses (Δ) of the composite mean sea level pressure (PSL, unit: hPa) centered on ETC storm center of all filtered storms in EAMv2 future climate simulations following SSP2-4.5 (first row) and SSP5-8.5 (second row) emission scenarios. The Δs are derived by subtracting the composite PSL in 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 black cross markers (panels a–d) indicate the center of ETC storms.

c). Accordingly, a weaker change of the temperature and water vapor transport is also evidenced in the simulations with ML bias correction (panels b and d versus panels a and c in Fig. 11). The modifications of the ML bias correction on the climate change signals are likely reasonable. As shown in Figure B1, 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.

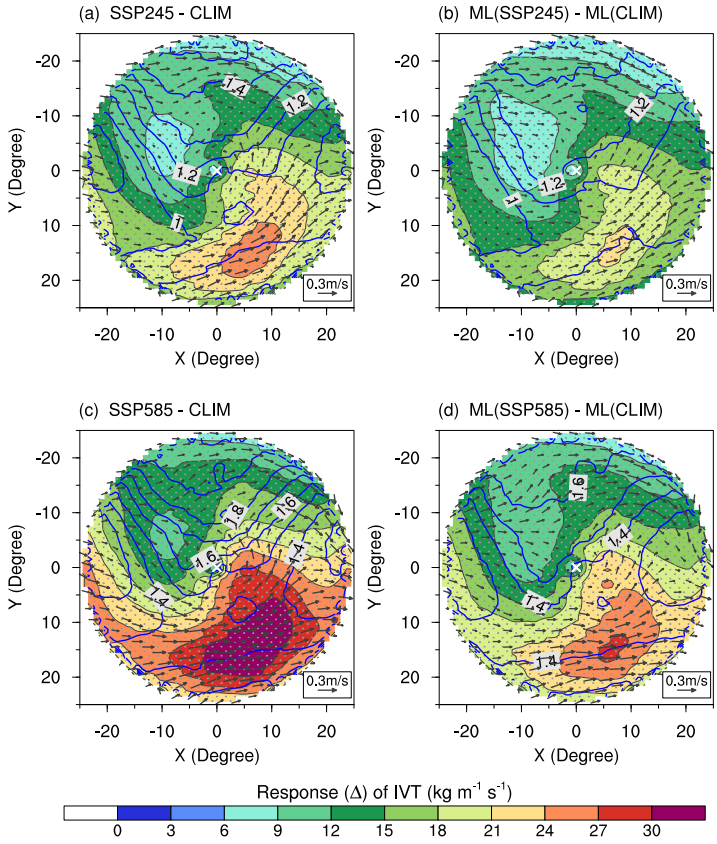


Figure 11. Same as Figure 10 but for responses (Δ) of temperature (blue contours, unit: K) and wind (black vectors) at 850-hPa pressure level as well as the vertically integrated vapor transport (IVT, unit: kg m^{-1} s^{-1}) to climate change following SSP2-4.5 (first row) and SSP5-8.5 (second row) emission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

4.3 Tropical Cyclones (TCs)

Tropical cyclones (TCs) are low-pressure systems that typically form in lower-latitude regions, and are one of the most 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 ~ 100km 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 B6 in Appendix). Therefore, direct evaluation of TempestExtreme-derived metrics for TCs provides limited value to draw conclusions on 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.

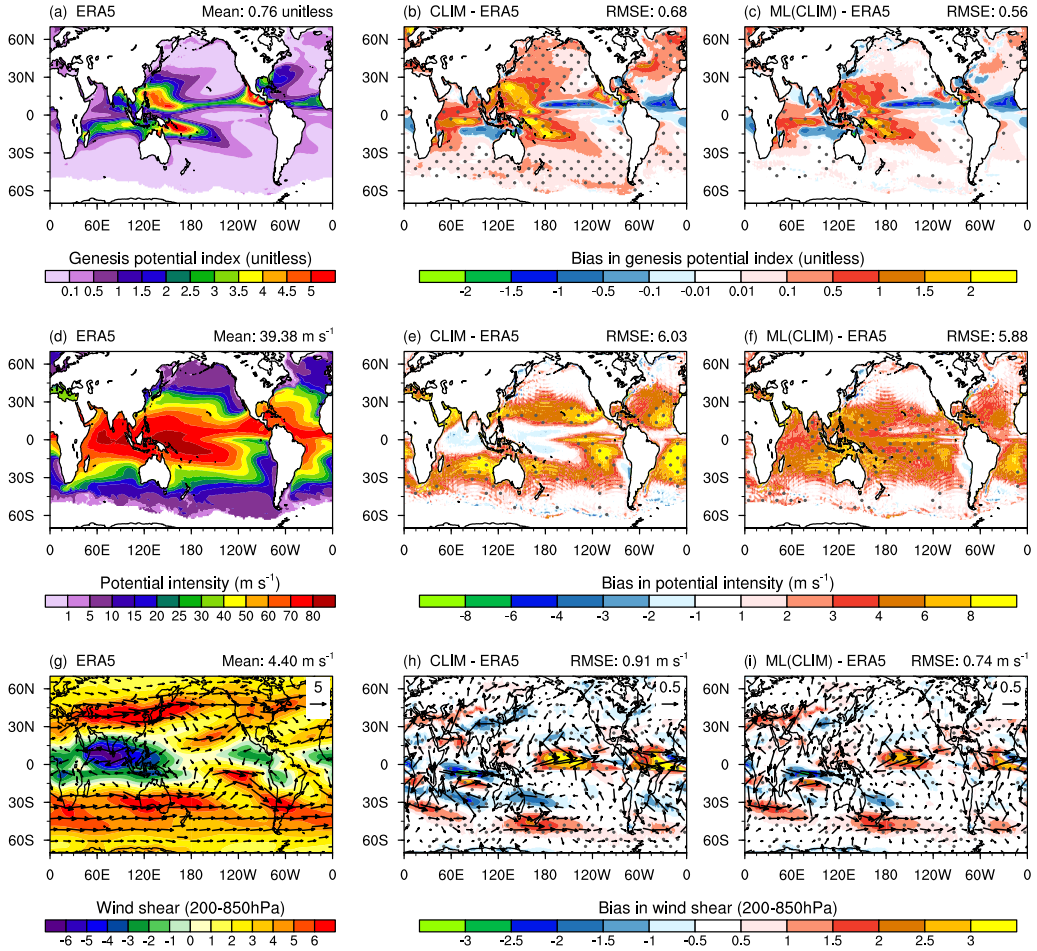


Figure 12. First column: Seasonal mean tropical cyclone Genesis Potential Index (GPI, unitless, panel a), potential intensity (PI, unit: m s^{-1}, panel d), and vertical wind shear between 200 and 850 hPa (unit: m s^{-1}, panel g) from ERA5 reanalysis averaged over the period of 1979-2014; Second column: bias in GPI (panel b), PI (panel e) and 200–850hPa 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 and i). The monthly mean model output from ERA5 reanalysis and EAMv2 simulations are used to calculate the GPI, PI, and vertical wind shear following Eq.(C7). 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. The detailed description of EAMv2 simulations can be found in Table 1.

Figure 12 shows the climatological season 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 Eq.C7 in Appendix). Results in Figure 12 suggest that the differences between CLIM and ERA5 in terms of the GPI, PI and low-level wind shear (Figs. 12 b, e and h) over most of TC basins are systematically reduced when the ML bias correction is applied to the U, V, T and Q fields (Figs. 12 c, f and i). Therefore, ML bias correction improves the representation of TC large-scale storm environment. These improvements are due to the bias reduction in the large-scale U, V, T, and Q fields by the ML model, observed from the Eq.(C7). In addition, EAMv2 CLIM simulation, as shown in Fig. 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 (Fig B6i)) 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 (Fig 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. 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 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 Figure 13b–c). 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 pseudo global warming simulations with and without ML bias correction. As shown in Figures 13d and f, climate change induced changes in the large-scale wind fields lead to a significant increase in the number of TCs over the Atlantic Basin region, especially in the coastal regions over the eastern US. These climate change signals, which have been linked to the warmer SSTs over the eastern tropical Pacific Ocean under warming (Balaguru et al., 2023b) (also seen in Figs. A1a–b), are still seen in the simulations after applying the ML bias correction (Figures 13e and g), 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 g versus Figures 13d and f). 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 g. Overall, the improved representation of large-scale storm

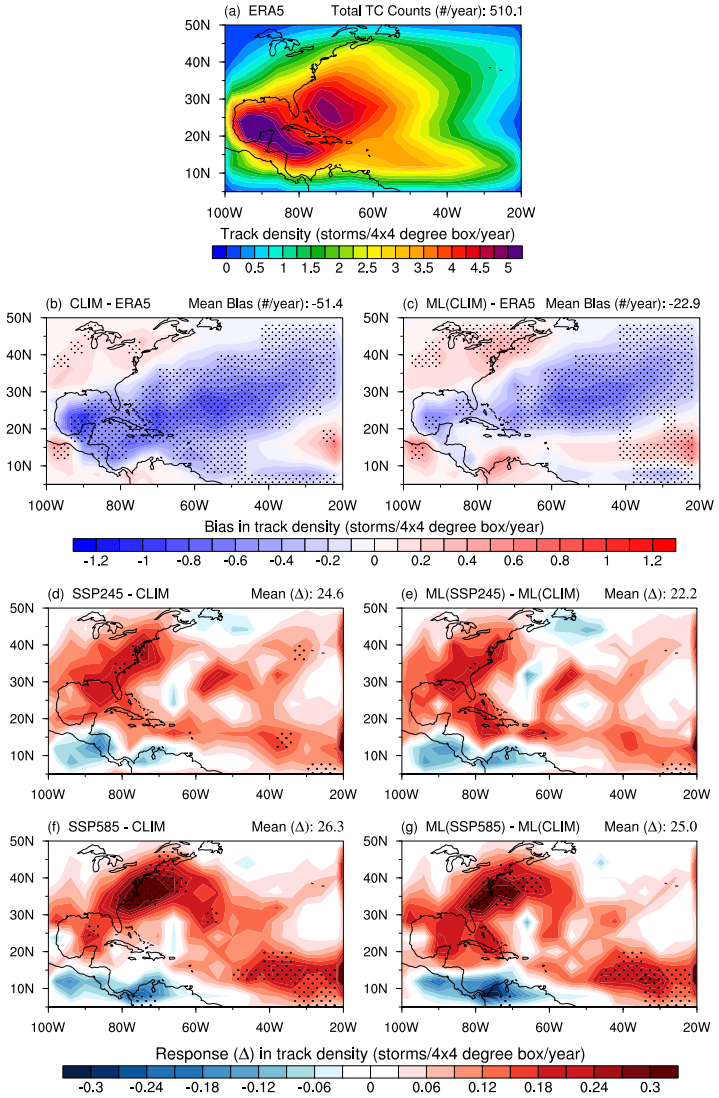


Figure 13. Annual mean TC track density over Atlantic basin from RAFT forecast driven by the large-scale environmental wind fields from ERA5 reanalysis (panel a), and the differences in track densities between RAFT forecasts driven by ERA5 and EAMV2 simulations (panels b–c), and by EAMv2 simulations 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(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 pseudo global warming 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 6hourly 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°x4° grid box per year. The grey dots in panels b and c indicate the differences are significant at a 95% confidence level.

environments (e.g. large-scale wind) by ML bias correction is beneficial for obtaining a more reliable downscaling of extreme weather events 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 extreme events are 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 Figs. 1 and B1). 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 by 10-20% quantitatively (Fig. 2). The same ML bias correction approach is then applied to process the pseudo global warming 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 pseudo global warming 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 Fig. 4).

This study further demonstrated the value of the employed ML bias correction in the assessment of extreme weather events in the low-resolution EAM model. We used the model state of U, V, T, and Q with and without ML bias correction to derive the long-term statistics, and evaluated the skills of bias correction in improving the EAM's representation of the extreme weather events (e.g. occurrence frequency, intensity, and storm environment, etc.) in both present-day and pseudo global warming scenarios, with a focus on atmospheric rivers (ARs), extratropical cyclones (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 datasets than those without bias correction, leading to a better representation of the occurrence frequency and the strength of ARs in the EAM model (see Sect.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 Sect.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 Sect.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 ex-

603 treme weather events for low-resolution climate models by providing more realistic large-
604 scale storm environment information.

Appendix A Supplementary material for Section 2

A1 Additional tables and figures for Section 2.2

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 psedo global warming simulations in Table. 1. 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.

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

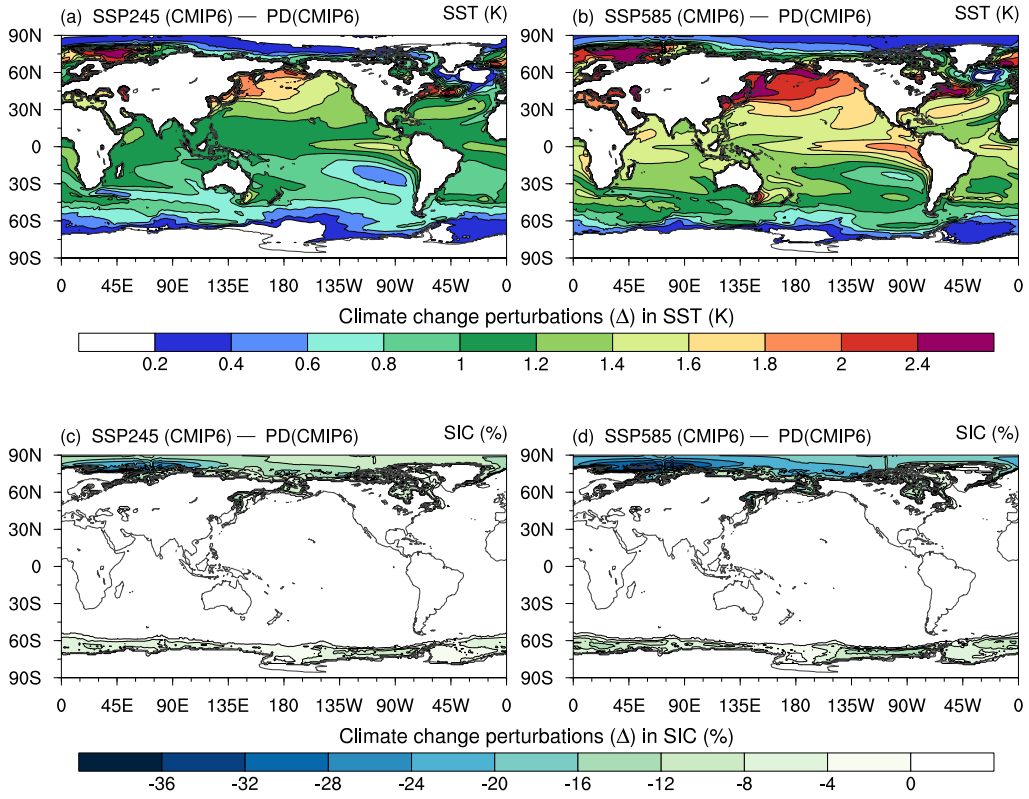


Figure A1. Changes of annual mean sea surface temperature (SST, unit: K, panels a and c) and sea-ice concentration (SIC, unit: %, panels b and d) in response to the forcing pathways of SSP2-4.5 (top row) and SSP5-8.5 (bottom row) from CMIP6 coupled model simulations. Shown is the multi-model ensemble 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).

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 et al. (2023). 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. Afterwards, 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.

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

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

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 time statistics of u by solving the imperfect model (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 affect 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 towards the reference data,

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

Here $\mathcal{P}$ is an operator which projects the reference solution onto the coarse grid. The constant τ is a user defined parameter which 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

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

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

$$a_k \equiv \sqrt{\frac{\mathcal{E}_{k,v}}{\mathcal{E}_{k,v_\tau}}} \tag{A4}$$

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

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

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

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

After the correction operator, $\mathcal{G}$, is trained on the spectrally-corrected nudged data, during testing it is applied the free run coarse model trajectory $v(t)$. The resulting corrected trajectory constitutes our ML prediction and is then used to compute statistics and other properties of interest. We refer the interested reader to Barthel et al. (2023) for a more detailed discussion of the mathematical framework and network architecture.

Appendix B Additional Tables and Figures

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

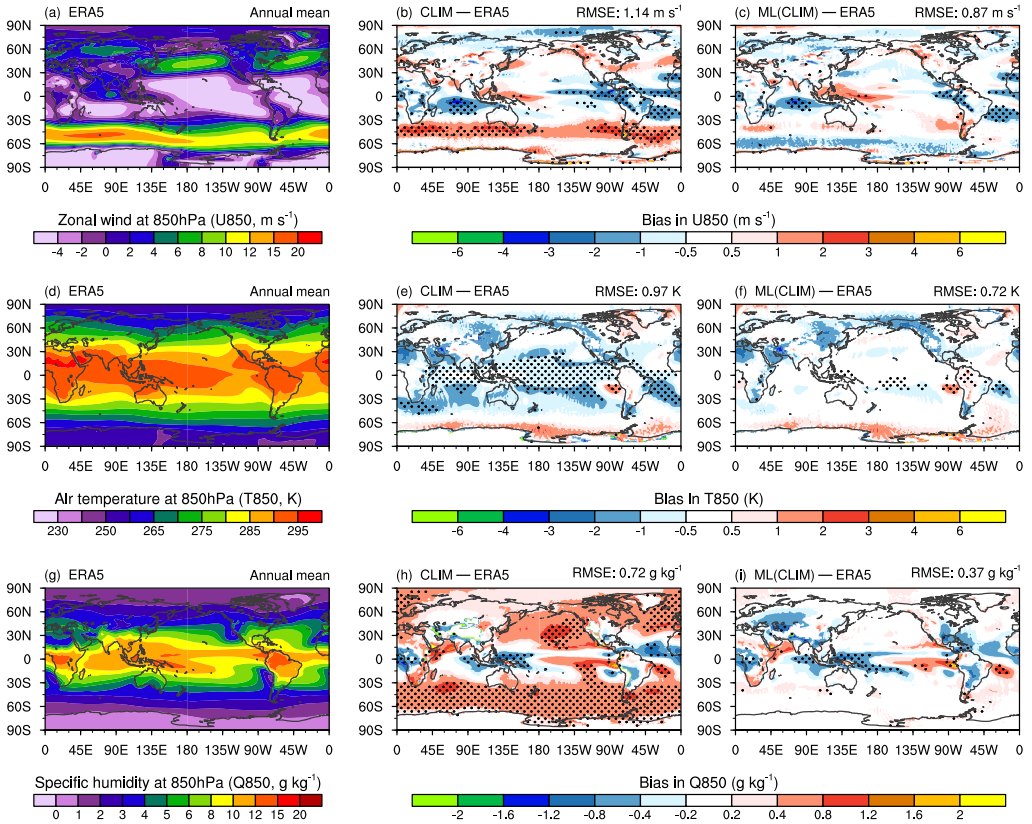


Figure B1. First column: Distribution of annual mean zonal velocity (U850, unit: m s^{-1}, panel a), temperature (T850, unit: K, panel d) and specific humidity (Q850, unit: g kg^{-1}, panel g) at 850-hPa pressure level from ERA5 reanalysis averaged over 1979-2014; Second and Third columns: the same as 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 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.

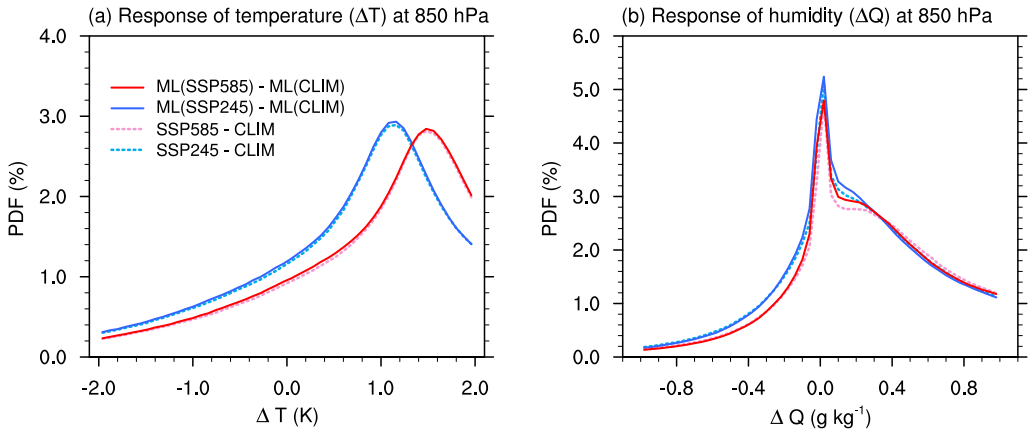


Figure B2. Probability Density Function (PDF) of monthly mean temperature (ΔT, unit: K, panel a) and humidity (ΔQ, unit: K, panel b) differences between present-day and pseudo global warming simulations conducted with EAMv2. Shown is the SSP245 - CLIM (dashed blue line) and SSP585 - CLIM (dashed red 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 on simulations can be found in Table 1.

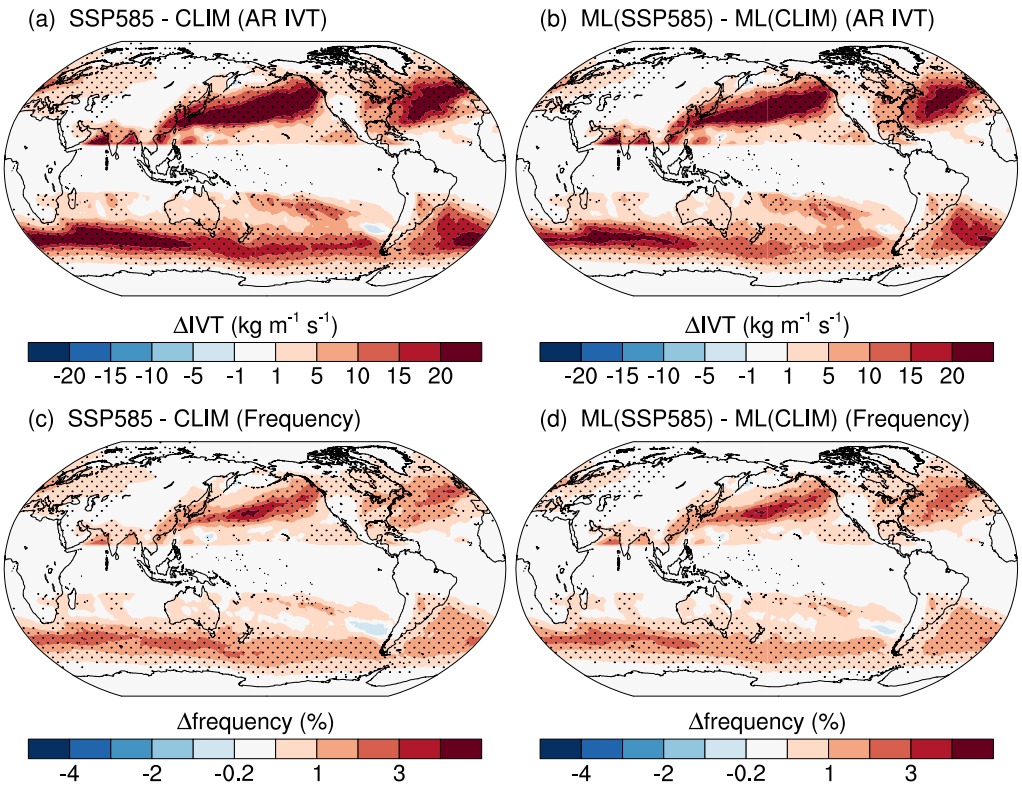


Figure B3. Same as Figure 6, but for the pseudo global warming simulations conducted with EAMv2 using imposed climate change perturbations in SST and SIC derived from SSP5-8.5 future scenarios.

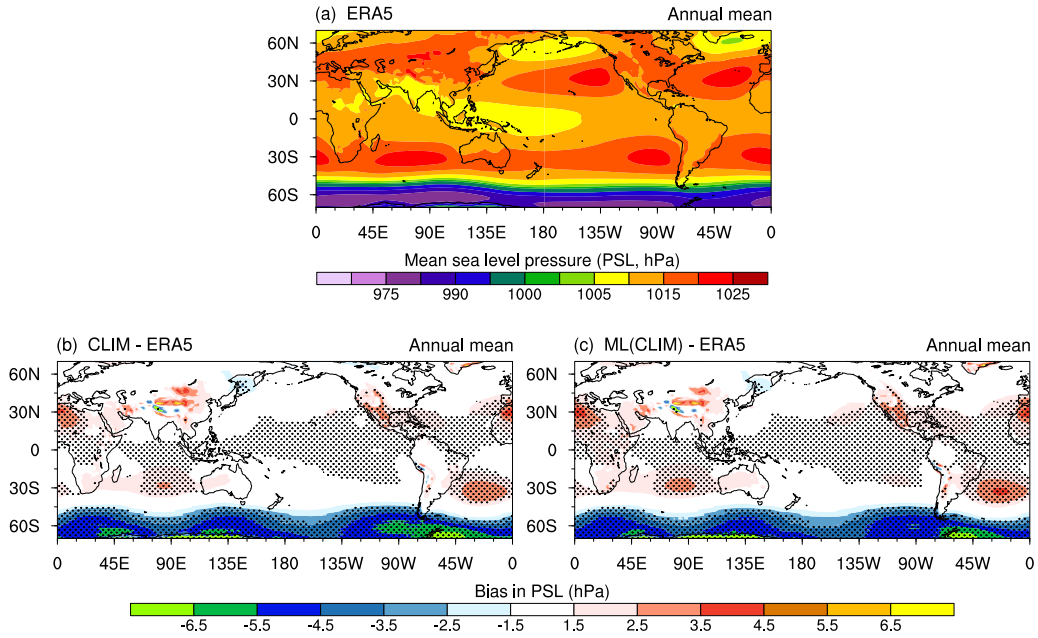


Figure B4. Horizontal distribution of mean sea level pressure averaged over the period of 1979-2014 from ERA5 reanalysis (panel a), and mean model biases in EAMv2 free-running simulations without (i.e CLIM, panel b) and with (i.e, ML(CLIM), panel c) ML bias correction. The dotted region in panels b and c indicate the differences between the model and ERA5 reanalysis are significant at a 95% significance level.

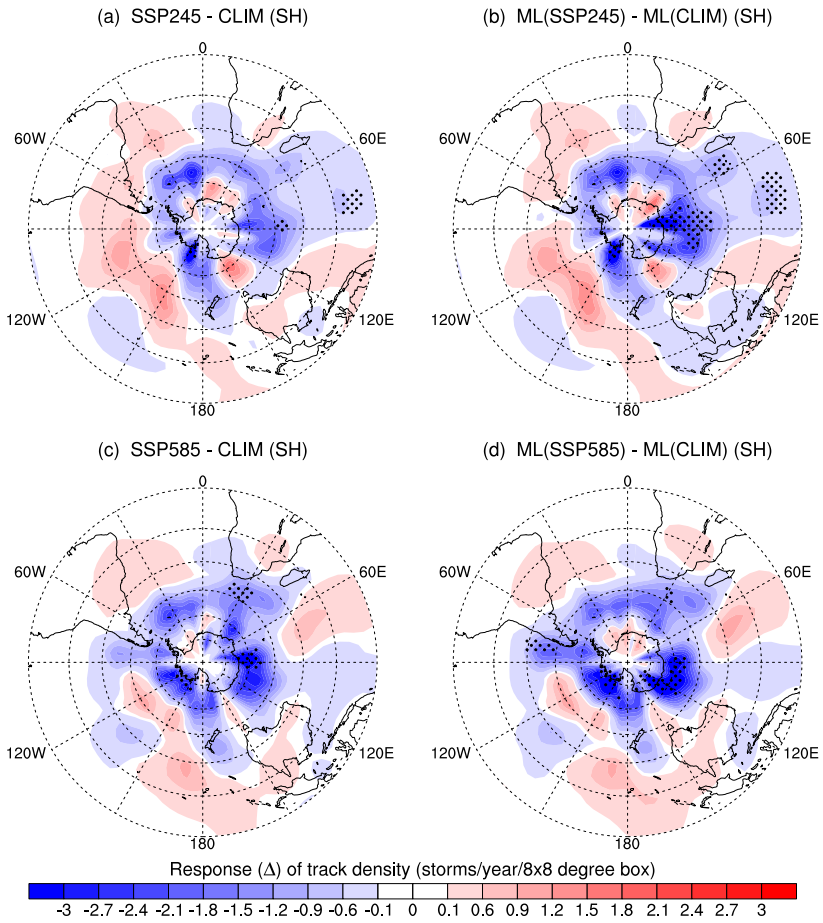


Figure B5. Same as Figure 9 but for the responses of ETC track density to imposted climate change perturbations in sea surface temperature (SST) and sea-ice concentration (SIC) from SSP2-4.5 (top row) and SSP5-8.5 (bottom row) over Southern Hemisphere.

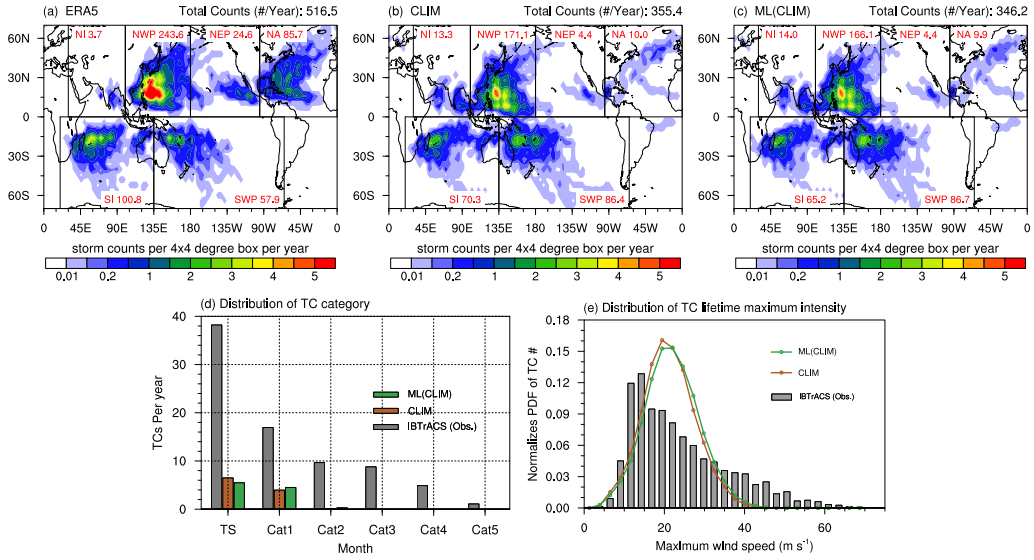


Figure B6. Top row: track density maps for tropical cyclones (TCs) tracked in ERA5 reanalysis (panel a) and EAMv2 free-running simulations without (i.e., CLIM, panel b) and with (i.e., ML(CLIM), panel c) ML bias correction. The 6-hourly sea level pressure (PSL) from 1979 to 2014 are used to track the 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°x4° 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-meter maximum wind speed (e) in IBTrACS observations (grey 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) correspond to the Saffir-Simpson wind scale: TS, tropical storm (17.5–32 m s^{-1}); Cat1, Category 1 (33–42 m s^{-1}); Cat2, Category 2 (43–49 m s^{-1}); Cat3, Category 3 (50–58 m s^{-1}); Cat4, Category 4 (59–69 m s^{-1}); Cat5, Category 5 (> 69 m s^{-1}). The IBTrACS refers to 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 1979–2014.

Appendix C Additional notes for Section 4

Sect. 4 discussed three extreme weather events, including atmospheric rivers (ARs), extratropical cyclones (ETCs), and tropical cyclones (TCs), that are discussed in Sect. 4. Table C1 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 events 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 the integrated water vapor transport (TUQ, TVQ). In this paper, the same algorithm and parameter setups described in P. A. Ullrich et al. (2021) were used by TempestExtremes for feature tracking of each extreme weather events in EAMv2 simulations. Here, both TCs and ETCs are defined as low-pressure weather systems, but TCs are distinguished from ETCs with their unique warm core structure. The average temperature over 200- and 500-hPa pressure levels at each grid point was used in TempestExtreme to identify the TC warm cores, and to separate TCs and ETCs in the feature tracking.

Table C1. List of extreme weather events and model variables used for feature tracking by the TempestExtremes package. The EAM model output used to derive the feature quantities is listed in the fourth column. See context in Appendix B for details.

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

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

$$PSL = \begin{cases} P_S, & \text{if } Z_s \leq 1e^{-4}\ m \\ P_S\ e^{\frac{Z_s}{R_d T_s}\left(1-\frac{x}{2}+\frac{x^2}{3}\right)}, & \text{otherwise} \end{cases} \tag{C1}$$

where,

$$x = \begin{cases} 0, & \text{if } T_s > 290.5\ K \text{ and } T_0 > 290.5\ K \\ \frac{R_d}{g}\frac{(290.5-T_s)}{T_s}, & \text{if } T_s \leq 290.5\ K \text{ and } T_0 > 290.5\ K \\ \frac{R_d}{g}\frac{\Gamma_0 Z_s}{T_s}, & \text{otherwise} \end{cases} \tag{C2}$$

$$T_0 = T_b\left[1+\frac{R_d\Gamma_0}{g}\left(\frac{P_S}{P_b}-1\right)\right]+\Gamma_0 Z_s \tag{C3}$$

$$T_s = T_{bot}\left[1+\frac{R_d\Gamma_0}{g}\left(\frac{P_S}{P_b}-1\right)\right], \text{ where} \tag{C4}$$

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

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

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

$$TUQ = \frac{1}{g}\int_{P_s}^{P_t} UQ\, dP \tag{C5}$$

$$TVQ = \frac{1}{g}\int_{P_s}^{P_t} VQ\, dP \tag{C6}$$

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 hPa.

Finally, the cyclone genesis potential index (GPI), 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):

$$GPI = 10^5 \eta^{\frac{3}{2}} \left(\frac{H_{700}}{50}\right)^3 \left(\frac{V_{pot}}{70}\right)^3 (1 + 0.1V_{shear})^{-2} \tag{C7}$$

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}.

Open Research Section

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/tempestextremes` (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.10106705` (S. Zhang & Charalampopoulos, 2023).

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.

References

Angélil, O., Perkins-Kirkpatrick, S., Alexander, L. V., Stone, D., Donat, M. G., Wehner, M., ... Christidis, N. (2016). Comparing regional precipitation and temperature extremes in climate model and reanalysis products. *Weather and Climate Extremes*, *13*, 35-43. doi: 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., ... Mametjanov, A. (2020). Characterizing tropical cyclones in the energy exascale earth system model version 1. *Journal of Advances in Modeling Earth Systems*, *12*(8), e2019MS002024. (e2019MS002024 10.1029/2019MS002024) doi: https://doi.org/10.1029/2019MS002024

Balaguru, K., Xu, W., Chang, C.-C., Leung, L. R., Judi, D. R., Hagos, S. M., ... Ting, M. (2023a). Increased u.s. coastal hurricane risk under climate change. *Science Advances*, *9*(14), eadf0259. doi: 10.1126/sciadv.adf0259

Balaguru, K., Xu, W., Chang, C.-C., Leung, L. R., Judi, D. R., Hagos, S. M., ... Ting, M. (2023b). Increased us coastal hurricane risk under climate change. *Science advances*, *9*(14), eadf0259.

Barthel, B., Charalampopoulos, A.-T., Zhang, S., Harrop, B., yung Ruby Leung, L., & Sapsis, T. (2023). *Machine learning correction operators for under-resolved climate models using nudged simulations.*

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*, 1847–1856. doi: 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. Retrieved from `https://journals.ametsoc.org/view/journals/clim/20/19/jcli4282.1.xml` doi: https://doi.org/10.1175/JCLI4282.1

Charalampopoulos, A.-T., Zhang, S., Harrop, B., yung Ruby Leung, L., & Sapsis, T. (2023). *Statistics of extreme events in coarse-scale climate simulations via machine learning correction operators trained on nudged datasets.* doi: 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. doi: 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. doi: 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. doi: 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*, 125–143. doi: https://doi.org/10.1007/s40641-015-0013-7

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., ... Booth, B. (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). United Kingdom: Cambridge University Press.

Dai, P., & Nie, J. (2022). Robust expansion of extreme midlatitude storms under global warming. *Geophysical Research Letters*, *49*(10), e2022GL099007. (e2022GL099007 2022GL099007) doi: https://doi.org/10.1029/2022GL099007

Dennis, J. M., Edwards, J., Evans, K. J., Guba, O., Lauritzen, P. H., Mirin, A. A., ... Worley, P. H. (2012). Cam-se: A scalable spectral element dynamical core for the community atmosphere model,. *Int. J. High Perform.*, *26*, 74–89. doi: 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) doi: https://doi.org/10.1016/j.gloplacha.2006.11.030

Dixon, K. W., Lanzante, J. R., Nath, M. J., Hayhoe, K., Stoner, A., Radhakrishnan, A., ... Gaitán, C. F. (2016). Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? *Climatic Change*, *135*, 395–408. doi: 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*, 381–395. doi: https://doi.org/10.1007/s10584-013-0954-6

E3SM Project, D. (2021, sep). *Energy exascale earth system model v2.0.* [Computer Software] `https://doi.org/10.11578/E3SM/dc.20210927.1`. Retrieved from `https://doi.org/10.11578/E3SM/dc.20210927.1` doi: 10.11578/E3SM/dc.20210927.1

Emanuel, K. A. (2000). A statistical analysis of tropical cyclone intensity. *Monthly Weather Review*, *128*(4). doi: 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. doi: 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. doi: 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 intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. Retrieved from `https://www.geosci-model-dev.net/9/1937/2016/` doi: 10.5194/gmd-9-1937-2016

Feng, L., Smith, S. J., Braun, C., Crippa, M., Gidden, M. J., Hoesly, R., ... van der Werf, G. R. (2020). The generation of gridded emissions data for cmip6. *Geoscientific Model Development*, *13*(2), 461–482. Retrieved from `https://www.geosci-model-dev.net/13/461/2020/` doi: 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. Retrieved from `https://nhess.copernicus.org/articles/9/405/2009/` doi: 10.5194/nhess-9-405-2009

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., ... Rummukainen, M. (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, UK: Cambridge University Press. doi: 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. Retrieved from `https://esd.copernicus.org/articles/11/537/2020/` doi: 10.5194/esd-11-537-2020

Fu, C., Wang, S., Xiong, Z., Gutowski, W. J., Lee, D.-K., McGregor, J. L., ... Suh, M.-S. (2005). Regional climate model intercomparison project for asia. *Bulletin of the American Meteorological Society*, *86*(2), 257 - 266. Retrieved from `https://journals.ametsoc.org/view/journals/bams/86/2/bams-86-2-257.xml` doi: 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. doi: 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., ... Williams, D. N. (1999). An overview of the results of the atmospheric model intercomparison project (amip i). *Bulletin of the American Meteorological Society*, *80*(1), 29-56. Retrieved from `https://doi.org/10.1175/1520-0477(1999)080<0029:AOOTRO>2.0.CO;2` doi: 10.1175/1520-0477(1999)080⟨0029:AOOTRO⟩2.0.CO;2

Gettelman, 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*, 1268–1287. doi: 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. Retrieved from `https://journals.ametsoc.org/view/journals/clim/7/3/1520-0442_1994_007_0375_rccsot_2_0_co_2.xml` doi: 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., ... Zhu, Q. (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. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001603` doi: 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*, 3540–3551. doi: 10.1175/1520-0469(2002)059⟨3540:APBMFB⟩2.0.CO;2

Golaz, J.-C., Van Roekel, L. P., Zheng, X., Roberts, A. F., Wolfe, J. D., Lin, W., ... Bader, D. C. (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. doi: 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. doi: 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. Retrieved from https://hess.copernicus.org/articles/16/3383/2012/ doi: 10.5194/hess-16-3383-2012

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., ... Pagé, C. (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. doi: 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., ... Zarzycki, C. (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. Retrieved from https://journals.ametsoc.org/view/journals/bams/101/5/bams-d-19-0113.1.xml doi: https://doi.org/10.1175/BAMS-D-19-0113.1

Han, L., Chen, M., Chen, K., Chen, H., Zhang, Y., Lu, B., ... Qin, R. (2021). A deep learning method for bias correction of ecmwf 24–240 h forecasts. *Adv. Atmos. Sci.*, *38*, 1444–1459. doi: 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). doi: 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-cslam. *Journal of Advances in Modeling Earth Systems*, *11*(7), 1894-1916. doi: https://doi.org/10.1029/2019MS001684

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñz Sabater, J., ... Thépaut, J.-N. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: https://doi.org/10.1002/qj.3803

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., ... Zhang, Q. (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. Retrieved from https://www.geosci-model-dev.net/11/369/2018/ doi: 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. doi: https://doi.org/10.1175/1520-0469(2002)059⟨1041:NPOTNH⟩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: Atmospheres*, *113*, D13103. doi: 10.1029/2008JD009944

Kanada, S., Takemi, T., Kato, M., Yamasaki, S., Fudeyasu, H., Tsuboki, K., ... Takayabu, I. (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. doi: 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*, 181-190. doi: 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.

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. Retrieved from https://gmd.copernicus.org/articles/15/5461/2022/ doi: 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. (Observed and Projected (Longer-term) Changes in Weather and Climate Extremes) doi: 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., ... Wu, L. (2019). Tropical cyclones and climate change assessment: Part i: Detection and attribution. *Bulletin of the American Meteorological Society*, *100*(10), 1987 - 2007. doi: 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., ... Wu, L. (2020). Tropical cyclones and climate change assessment: Part ii: Projected response to anthropogenic warming. *Bulletin of the American Meteorological Society*, *101*(3), E303 - E322. doi: 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., ... Villarini, G. (2013). Dynamical downscaling projections of twenty-first-century atlantic hurricane activity: Cmip3 and cmip5 model-based scenarios. *Journal of Climate*, *26*(17), 6591 - 6617. doi: https://doi.org/10.1175/JCLI-D-12-00539.1

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*, 3519–3539. doi: 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., ... Tribbia, J. J. (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. doi: 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. doi: 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.

Liu, X., Ma, P.-L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., ... Rasch, P. (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*, 505–522. doi: doi.org/10.5194/gmd-9-505-2016

Lucas-Picher, P., Arg ueso, D., Brisson, E., Tramblay, Y., Berg, P., Lemonsu, A., ... Caillaud, C. (2021). Convection-permitting modeling with regional climate models: Latest developments and next steps. *WIREs Climate Change*, *12*(6), e731. doi: https://doi.org/10.1002/wcc.731

Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D. N., ... Wurpts, A. (2020). Impact forecasting to support emergency management of natural hazards. *Reviews of Geophysics*, *58*(4), e2020RG000704. doi: 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: Atmospheres*, *102*, 16,663-16,682. doi: 10.1029/97JD00237

Moghim, S., & Bras, R. L. (2017). Bias correction of climate modeled temperature and precipitation using artificial neural networks. *J. Hydrometeor.*, *18*, 1867–1884.

Moon, H., Gudmundsson, L., & Seneviratne, S. I. (2018). Drought persistence errors in global climate models. *Journal of Geophysical Research: Atmospheres*, *123*(7), 3483-3496. doi: https://doi.org/10.1002/2017JD027577

Mori, N., Shimura, T., Yoshida, K., Mizuta, R., Okada, Y., Fujita, M., ... Nakakita, E. (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. doi: 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*, 3642–3659. doi: 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. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013GL058055` doi: 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. doi: 10.1073/pnas.1800357115

Oleson, K. W., Lawrence, M., Bonan, B., Drewniak, B. A., Huang, M., Koven, D., ... Yang, Z. (2013). *Technical description of version 4.5 of the community land model (CLM)* (NCAR Technical Note NCAR/TN-503+STR, No. 420pp.) NCAR, USA. doi: 10.5065/D6RR1W7M

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., ... Sanderson, B. M. (2016). The scenario model intercomparison project (scenariomip) for cmip6. *Geoscientific Model Development*, *9*(9), 3461–3482. Retrieved from `https://gmd.copernicus.org/articles/9/3461/2016/` doi: 10.5194/gmd-9-3461-2016

Payne, A. E., Demory, M.-E., Leung, L. R., Ramos, A. M., Shields, C. A., Rutz, J. J., ... Ralph, F. M. (2020). Responses and impacts of atmospheric rivers to climate change. *Nature Reviews Earth & Environment*, *1*(3), 143–157.

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. doi: 10.1175/1520-0442(2002)015⟨1609:AIISAS⟩2.0.CO;2

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., ... others (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* (p. 109–230). Cambridge University Press. doi: 10.1017/CBO9781139177245.006

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., ... Zhou, B. (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* (p. 1513–1766). Cambridge University Press. doi: 10.1017/9781009157896.013

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. Retrieved from `https://wcd.copernicus.org/articles/1/1/2020/` doi: 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. doi: 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. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001831` doi: 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. doi: 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. doi: 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). Retrieved from `https://api.semanticscholar.org/CorpusID:131237602` doi: doi:10.5065/D6HX19NH

Ullrich, P. (2022, jan). *Tempestextremes v2.2.1.* [Computer Software] `https://github.com/ClimateGlobalChange/tempestextremes/releases/tag/v2.2.1`. Retrieved from `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.

Vaittinada Ayar, P., Vrac, M., & Mailhot, A. (2021). Ensemble bias correction of climate simulations: preserving internal variability. *Scientific Reports*, *11*(1), 3098. doi: 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. doi: 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. Retrieved from `https://nhess.copernicus.org/articles/12/2769/2012/` doi: 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 Dyn.*, *59*, 3451–3468. doi: https://doi.org/10.1007/s00382-022-06277-2

Wang, H., Easter, R. C., Zhang, R., Ma, P.-L., Singh, B., Zhang, K., ... Yoon, J.-H. (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. doi: https://doi.org/10.1029/2019MS001851

Wang, Y., Liu, X., Hoose, C., & Wang, B. (2014). Different contact angledistributions for heterogeneous ice nucleation in the communityatmospheric model version 5,. *Atmospheric Chemistry and Physics*, *14*, 10411–10430. doi: 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. doi: https://doi.org/10.1029/2018GL079220

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. doi: 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.

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. doi: 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. Retrieved from `https://journals.ametsoc.org/view/journals/clim/25/18/jcli-d-12-00005.1.xml` doi: https://doi.org/10.1175/JCLI-D-12-00005.1

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. doi: https://doi.org/10.1175/JCLI-D-12-00573.1

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. *Atmos. Ocean*, *33*, 407–446. doi: 10.1080/07055900.1995.9649539

Zhang, S., & Charalampopoulos, A.-T. (2023, November). *Analysis scripts and dataset for Zhang et. al. (2023)[Dataset].* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.10106706` doi: 10.5281/zenodo.10106706

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. Retrieved from `https://gmd.copernicus.org/articles/15/6787/2022/` doi: 10.5194/gmd-15-6787-2022

Figure 1.

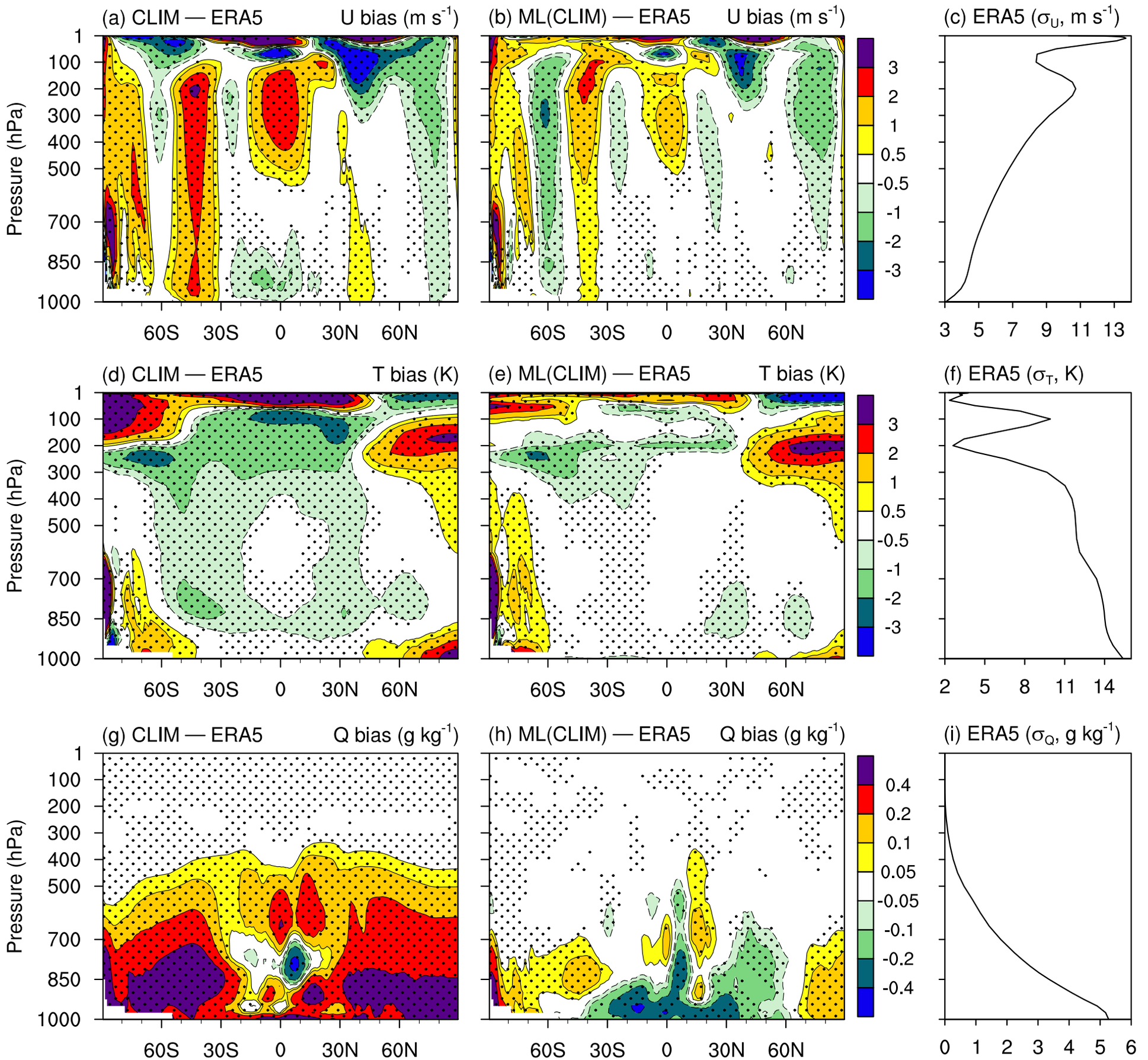

(a) CLIM — ERA5
U bias (m s-1)
(b) ML(CLIM) — ERA5
U bias (m s-1)
(c) ERA5 (σU, m s-1)
(d) CLIM — ERA5
T bias (K)
(e) ML(CLIM) — ERA5
T bias (K)
(f) ERA5 (σT, K)
(g) CLIM — ERA5
Q bias (g kg-1)
(h) ML(CLIM) — ERA5
Q bias (g kg-1)
(i) ERA5 (σQ, g kg-1)
Pressure (hPa)
1
100
200
300
400
500
700
850
1000
60S
30S
0
30N
60N
3
2
1
0.5
-0.5
-1
-2
-3
3 5 7 9 11 13
2 5 8 11 14
0.4
0.2
0.1
0.05
-0.05
-0.1
-0.2
-0.4
0 1 2 3 4 5 6

Figure 2.

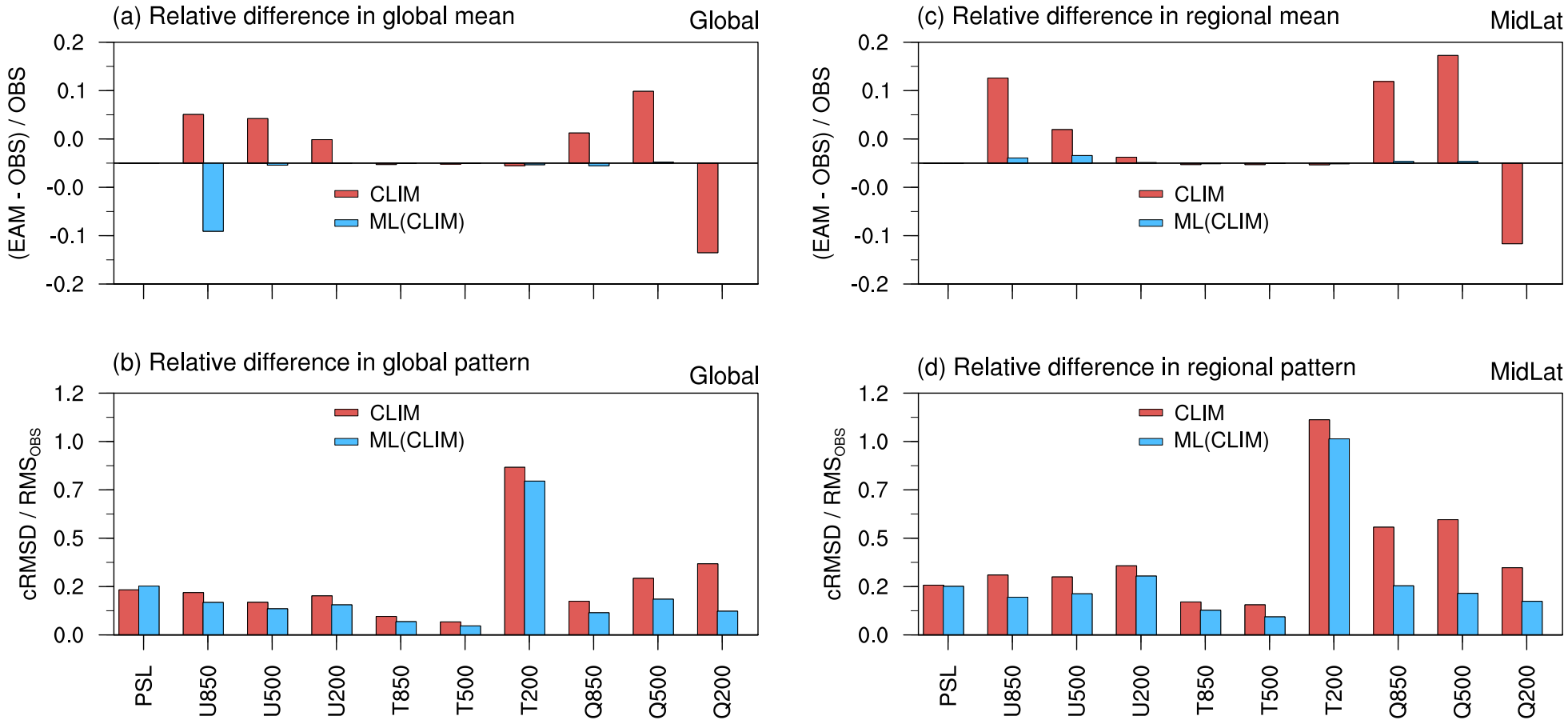

(a) Relative difference in global mean
Global
(EAM - OBS) / OBS
0.2
0.1
0.0
-0.1
-0.2
CLIM
ML(CLIM)
(c) Relative difference in regional mean
MidLat
(EAM - OBS) / OBS
0.2
0.1
0.0
-0.1
-0.2
CLIM
ML(CLIM)
(b) Relative difference in global pattern
Global
cRMSD / RMS$_{OBS}$
1.2
1.0
0.7
0.5
0.2
0.0
CLIM
ML(CLIM)
PSL
U850
U500
U200
T850
T500
T200
Q850
Q500
Q200
(d) Relative difference in regional pattern
MidLat
cRMSD / RMS$_{OBS}$
1.2
1.0
0.7
0.5
0.2
0.0
CLIM
ML(CLIM)
PSL
U850
U500
U200
T850
T500
T200
Q850
Q500
Q200

Figure 3.

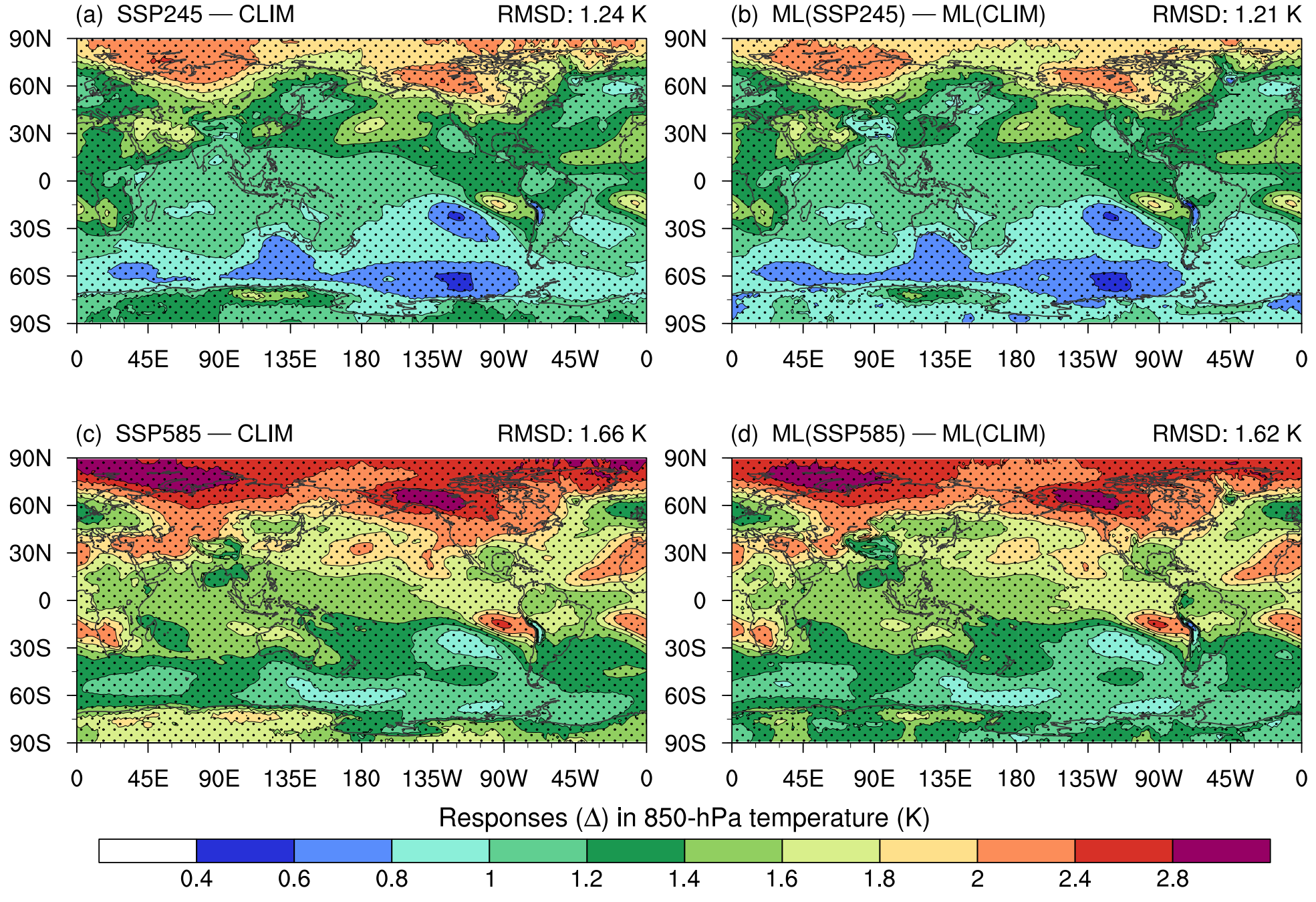

(a) SSP245 — CLIM
RMSD: 1.24 K
(b) ML(SSP245) — ML(CLIM)
RMSD: 1.21 K
(c) SSP585 — CLIM
RMSD: 1.66 K
(d) ML(SSP585) — ML(CLIM)
RMSD: 1.62 K
90N
60N
30N
0
30S
60S
90S
0
45E
90E
135E
180
135W
90W
45W
0
Responses (Δ) in 850-hPa temperature (K)
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
2.4
2.8

Figure 4.

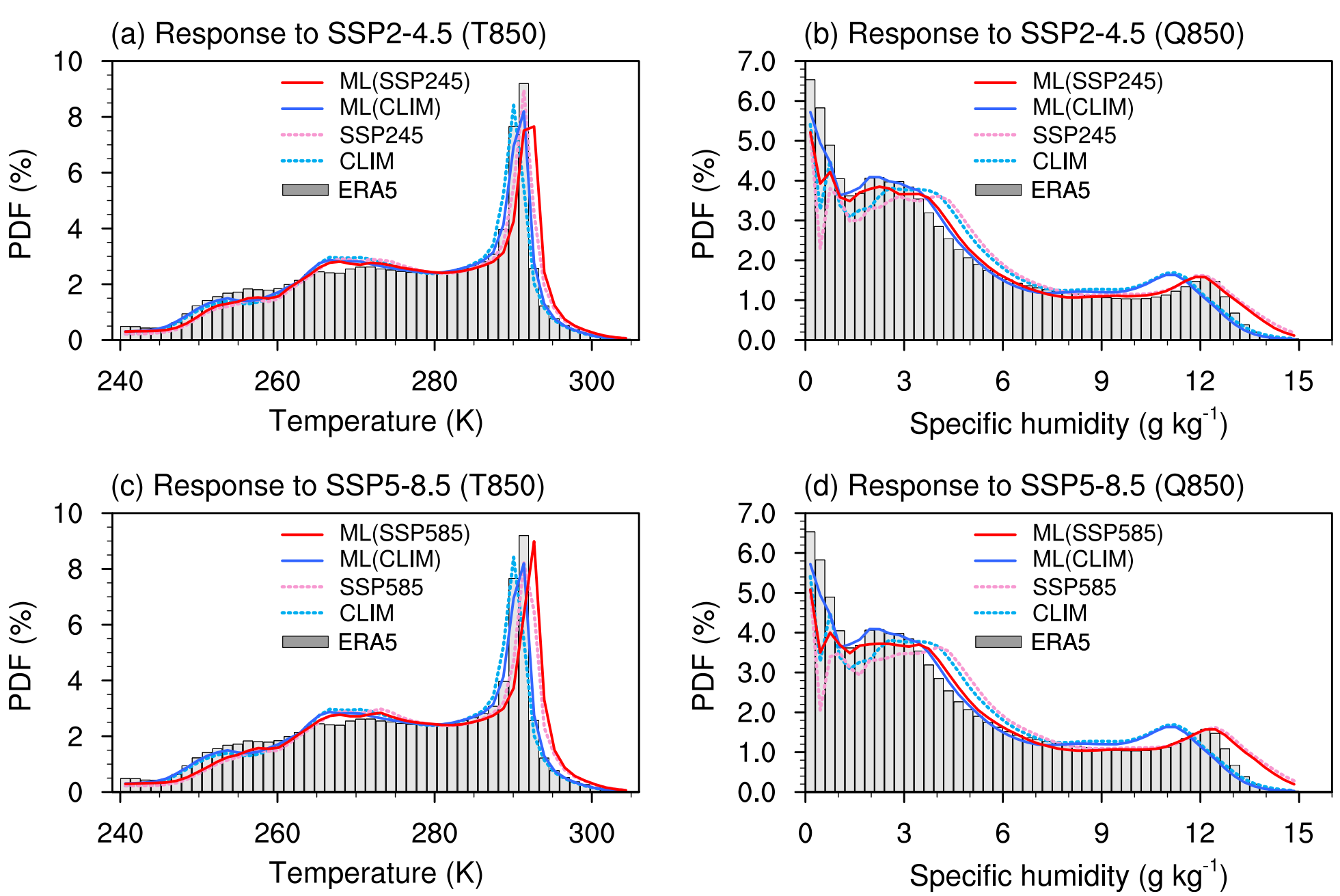

(a) Response to SSP2-4.5 (T850)
ML(SSP245)
ML(CLIM)
SSP245
CLIM
ERA5
PDF (%)
Temperature (K)
(b) Response to SSP2-4.5 (Q850)
Specific humidity (g kg-1)
(c) Response to SSP5-8.5 (T850)
ML(SSP585)
SSP585
(d) Response to SSP5-8.5 (Q850)

Figure 5.

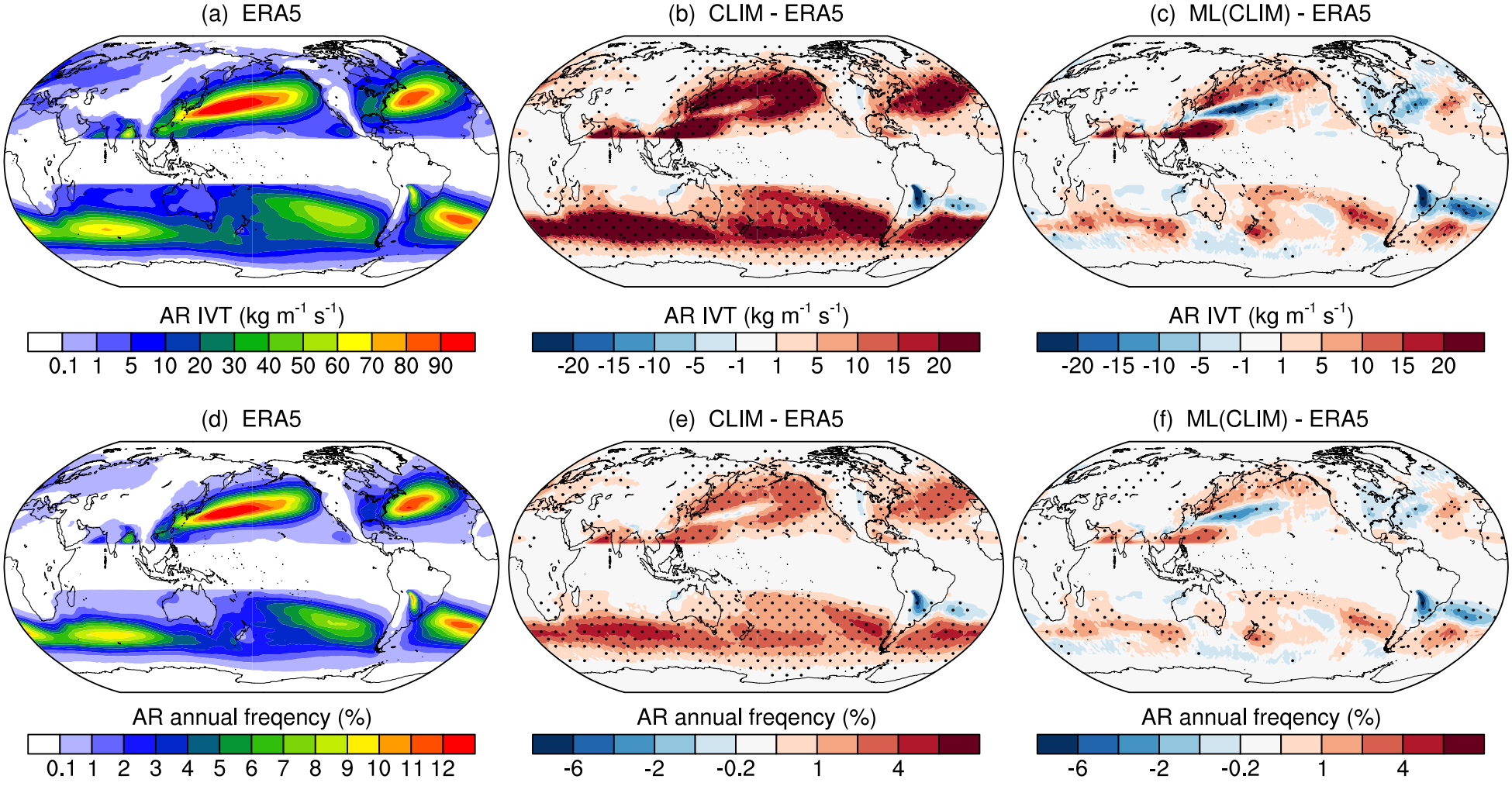

(a) ERA5
(b) CLIM - ERA5
(c) ML(CLIM) - ERA5
AR IVT (kg m-1 s-1)
0.1 1 5 10 20 30 40 50 60 70 80 90
AR IVT (kg m-1 s-1)
-20 -15 -10 -5 -1 1 5 10 15 20
AR IVT (kg m-1 s-1)
-20 -15 -10 -5 -1 1 5 10 15 20
(d) ERA5
(e) CLIM - ERA5
(f) ML(CLIM) - ERA5
AR annual freqency (%)
0.1 1 2 3 4 5 6 7 8 9 10 11 12
AR annual freqency (%)
-6 -2 -0.2 1 4
AR annual freqency (%)
-6 -2 -0.2 1 4

Figure 6.

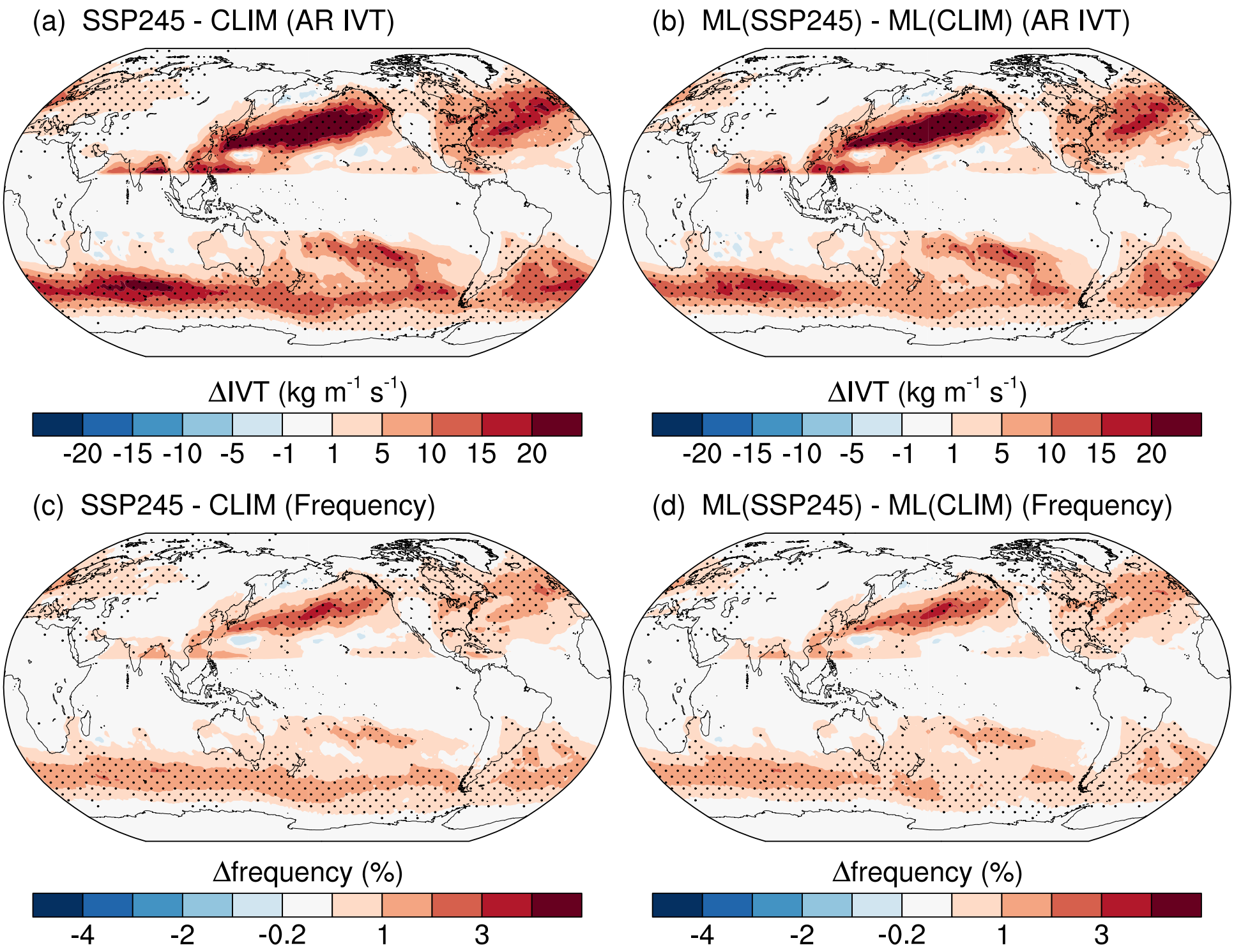

(a) SSP245 - CLIM (AR IVT)
(b) ML(SSP245) - ML(CLIM) (AR IVT)
ΔIVT (kg m-1 s-1)
-20 -15 -10 -5 -1 1 5 10 15 20
ΔIVT (kg m-1 s-1)
-20 -15 -10 -5 -1 1 5 10 15 20
(c) SSP245 - CLIM (Frequency)
(d) ML(SSP245) - ML(CLIM) (Frequency)
Δfrequency (%)
-4 -2 -0.2 1 3
Δfrequency (%)
-4 -2 -0.2 1 3

Figure 7.

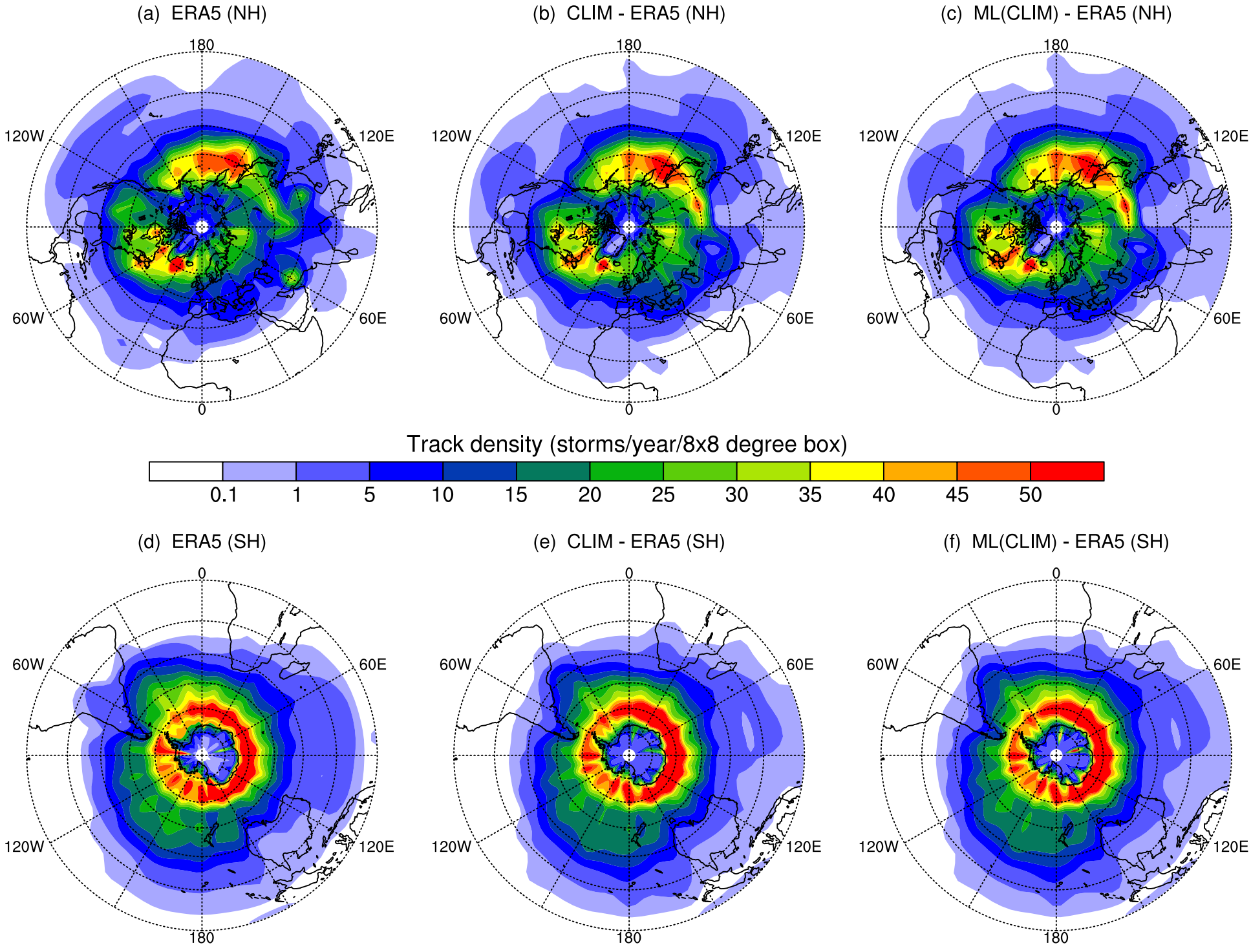

(a) ERA5 (NH)
(b) CLIM - ERA5 (NH)
(c) ML(CLIM) - ERA5 (NH)
180
120W
120E
60W
60E
0
Track density (storms/year/8x8 degree box)
0.1
1
5
10
15
20
25
30
35
40
45
50
(d) ERA5 (SH)
(e) CLIM - ERA5 (SH)
(f) ML(CLIM) - ERA5 (SH)
0
60W
60E
120W
120E
180
Track density (storms/year/8x8 degree box)
0.1
1
5
10
15
20
25
30
35
40
45
50

Figure 8.

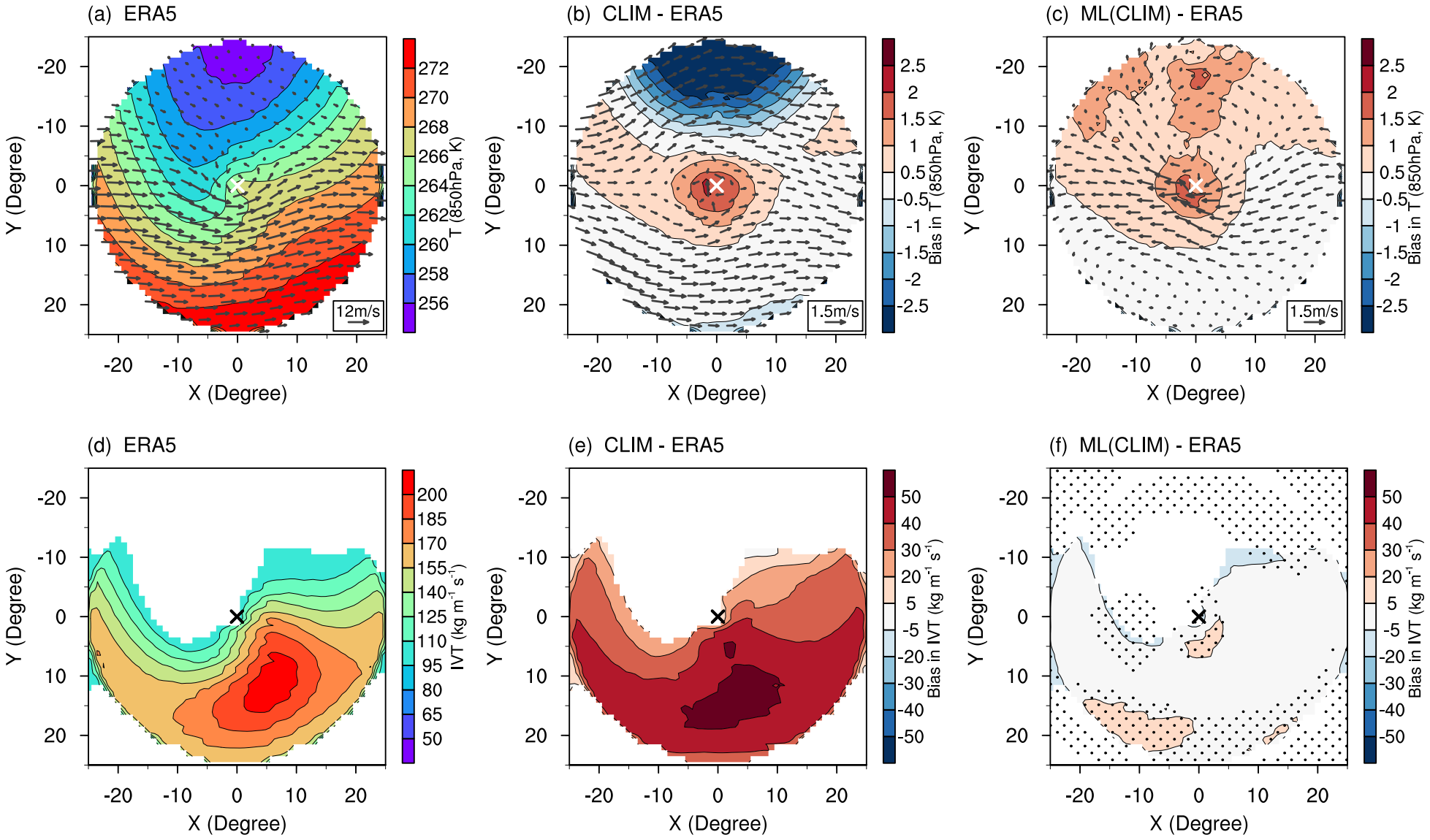

(a) ERA5
(b) CLIM - ERA5
(c) ML(CLIM) - ERA5
(d) ERA5
(e) CLIM - ERA5
(f) ML(CLIM) - ERA5
Y (Degree)
X (Degree)
T (850hPa, K)
Bias in T (850hPa, K)
IVT (kg m-1 s-1)
Bias in IVT (kg m-1 s-1)
12m/s
1.5m/s

Figure 9.

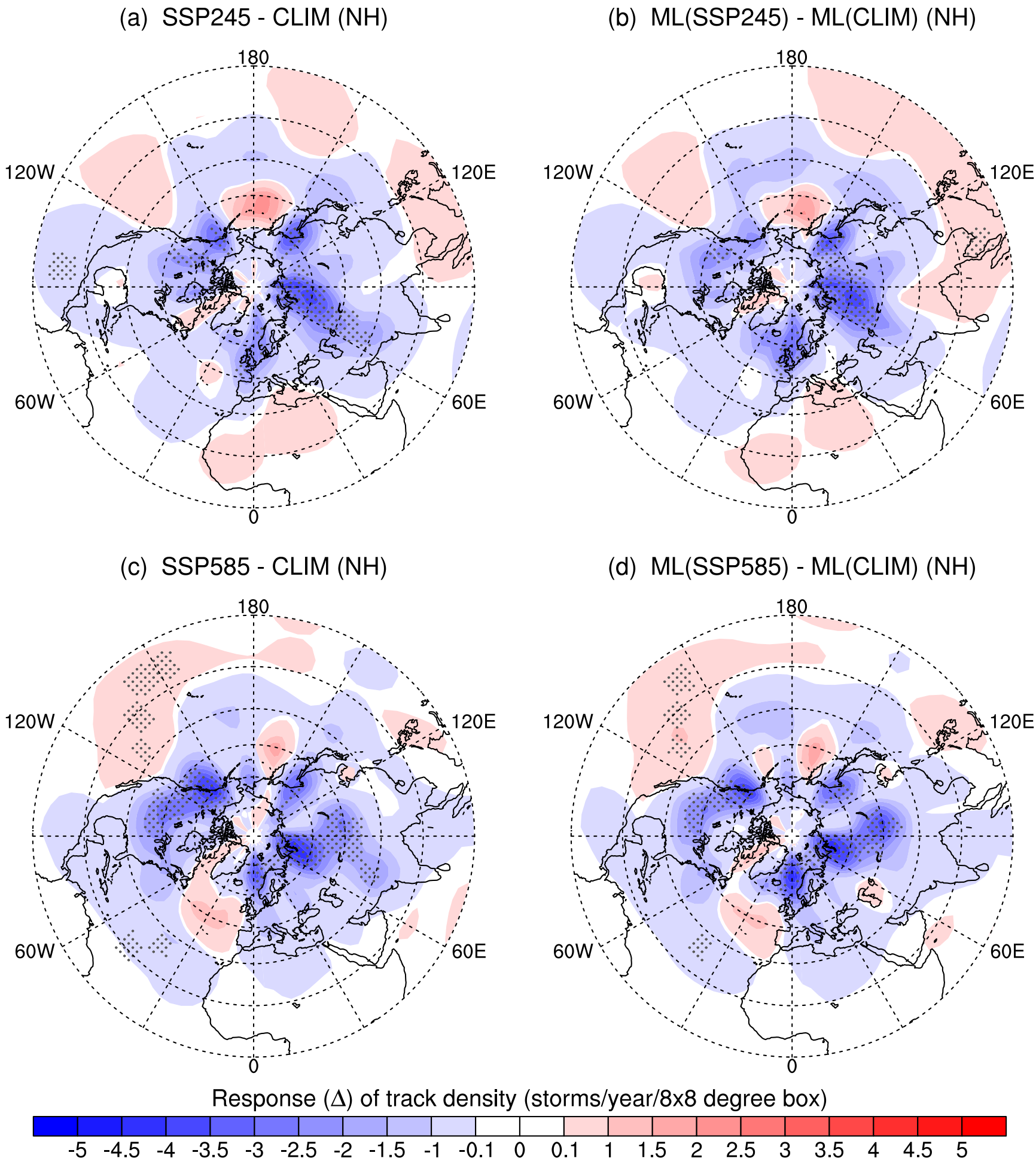

(a) SSP245 - CLIM (NH)
(b) ML(SSP245) - ML(CLIM) (NH)
(c) SSP585 - CLIM (NH)
(d) ML(SSP585) - ML(CLIM) (NH)
180
120W
120E
60W
60E
0
Response (Δ) of track density (storms/year/8x8 degree box)
-5 -4.5 -4 -3.5 -3 -2.5 -2 -1.5 -1 -0.1 0 0.1 1 1.5 2 2.5 3 3.5 4 4.5 5

Figure 10.

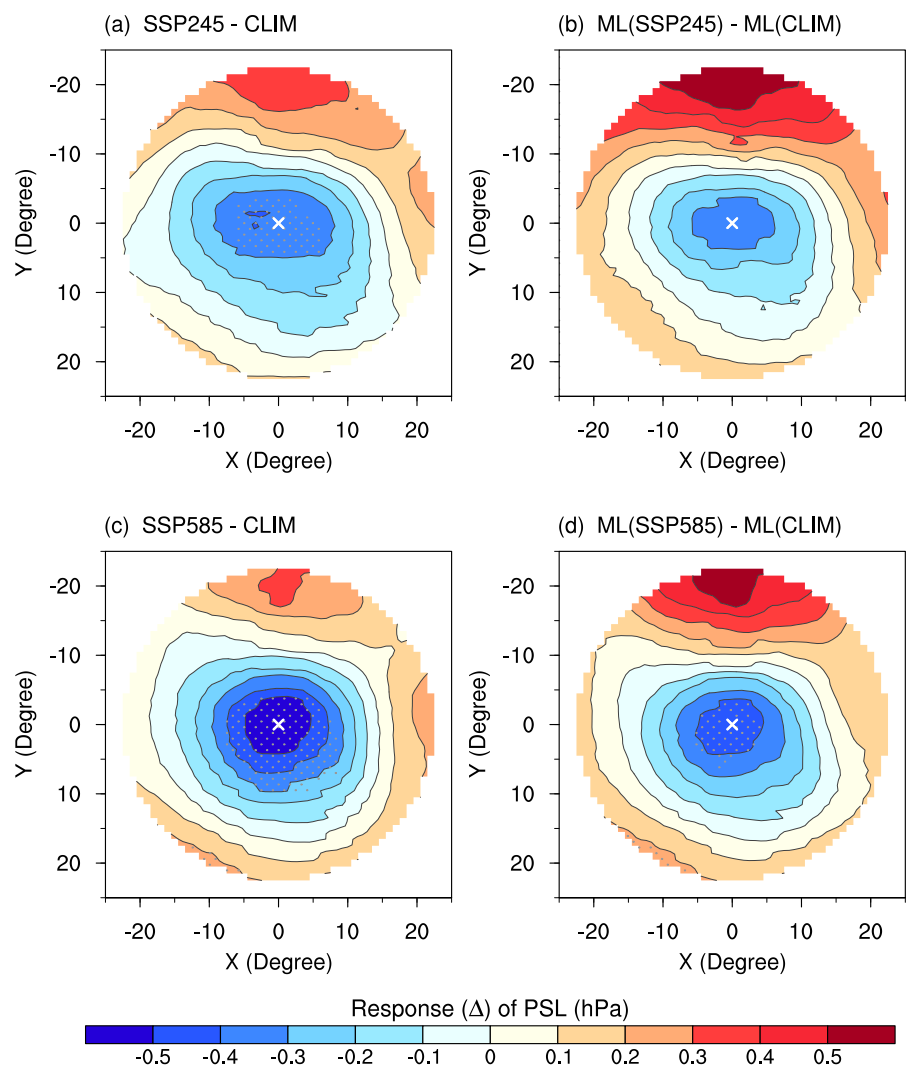

(a) SSP245 - CLIM
(b) ML(SSP245) - ML(CLIM)
(c) SSP585 - CLIM
(d) ML(SSP585) - ML(CLIM)
Y (Degree)
X (Degree)
-20
-10
0
10
20
Response (Δ) of PSL (hPa)
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5

Figure 11.

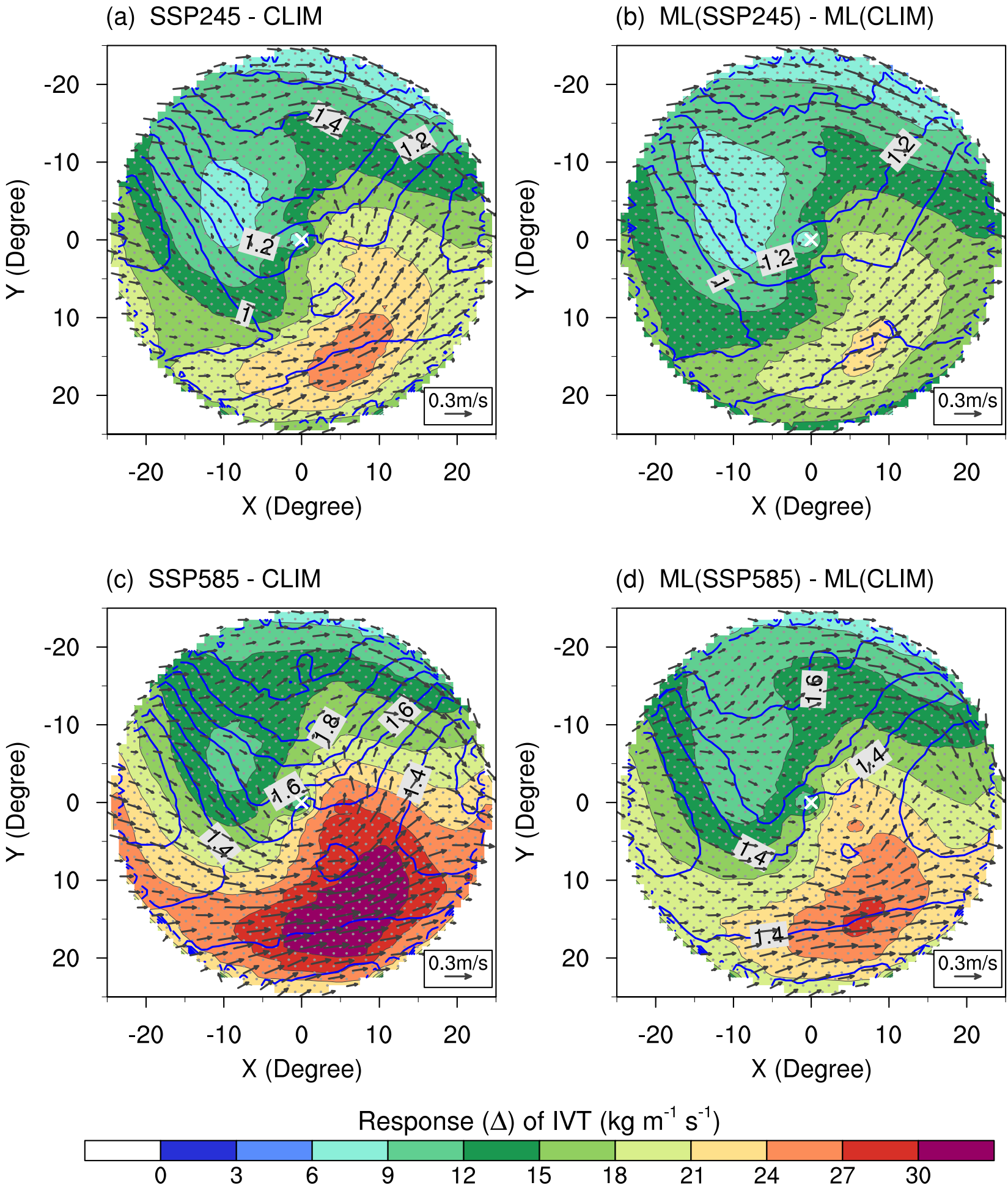

(a) SSP245 - CLIM
(b) ML(SSP245) - ML(CLIM)
(c) SSP585 - CLIM
(d) ML(SSP585) - ML(CLIM)
Y (Degree)
X (Degree)
-20
-10
0
10
20
0.3m/s
1
1.2
1.4
1.6
1.8
Response (Δ) of IVT (kg m-1 s-1)
0
3
6
9
12
15
18
21
24
27
30

Figure 12.

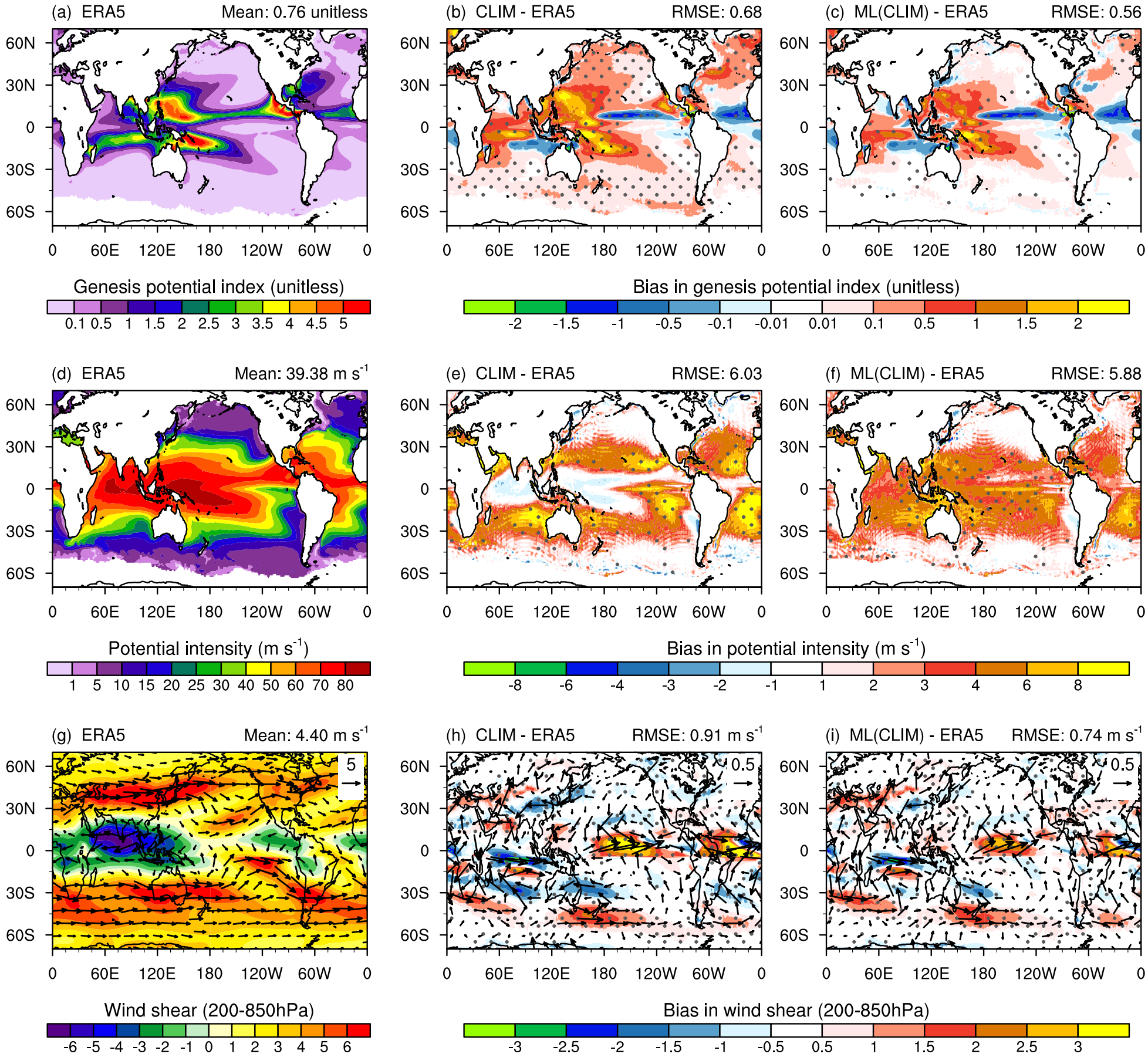

(a) ERA5
Mean: 0.76 unitless
(b) CLIM - ERA5
RMSE: 0.68
(c) ML(CLIM) - ERA5
RMSE: 0.56
60N
30N
0
30S
60S
0
60E
120E
180
120W
60W
0
Genesis potential index (unitless)
0.1 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
Bias in genesis potential index (unitless)
-2 -1.5 -1 -0.5 -0.1 -0.01 0.01 0.1 0.5 1 1.5 2
(d) ERA5
Mean: 39.38 m s-1
(e) CLIM - ERA5
RMSE: 6.03
(f) ML(CLIM) - ERA5
RMSE: 5.88
Potential intensity (m s-1)
1 5 10 15 20 25 30 40 50 60 70 80
Bias in potential intensity (m s-1)
-8 -6 -4 -3 -2 -1 1 2 3 4 6 8
(g) ERA5
Mean: 4.40 m s-1
(h) CLIM - ERA5
RMSE: 0.91 m s-1
(i) ML(CLIM) - ERA5
RMSE: 0.74 m s-1
5
0.5
0.5
Wind shear (200-850hPa)
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6
Bias in wind shear (200-850hPa)
-3 -2.5 -2 -1.5 -1 -0.5 0.5 1 1.5 2 2.5 3

Figure 13.

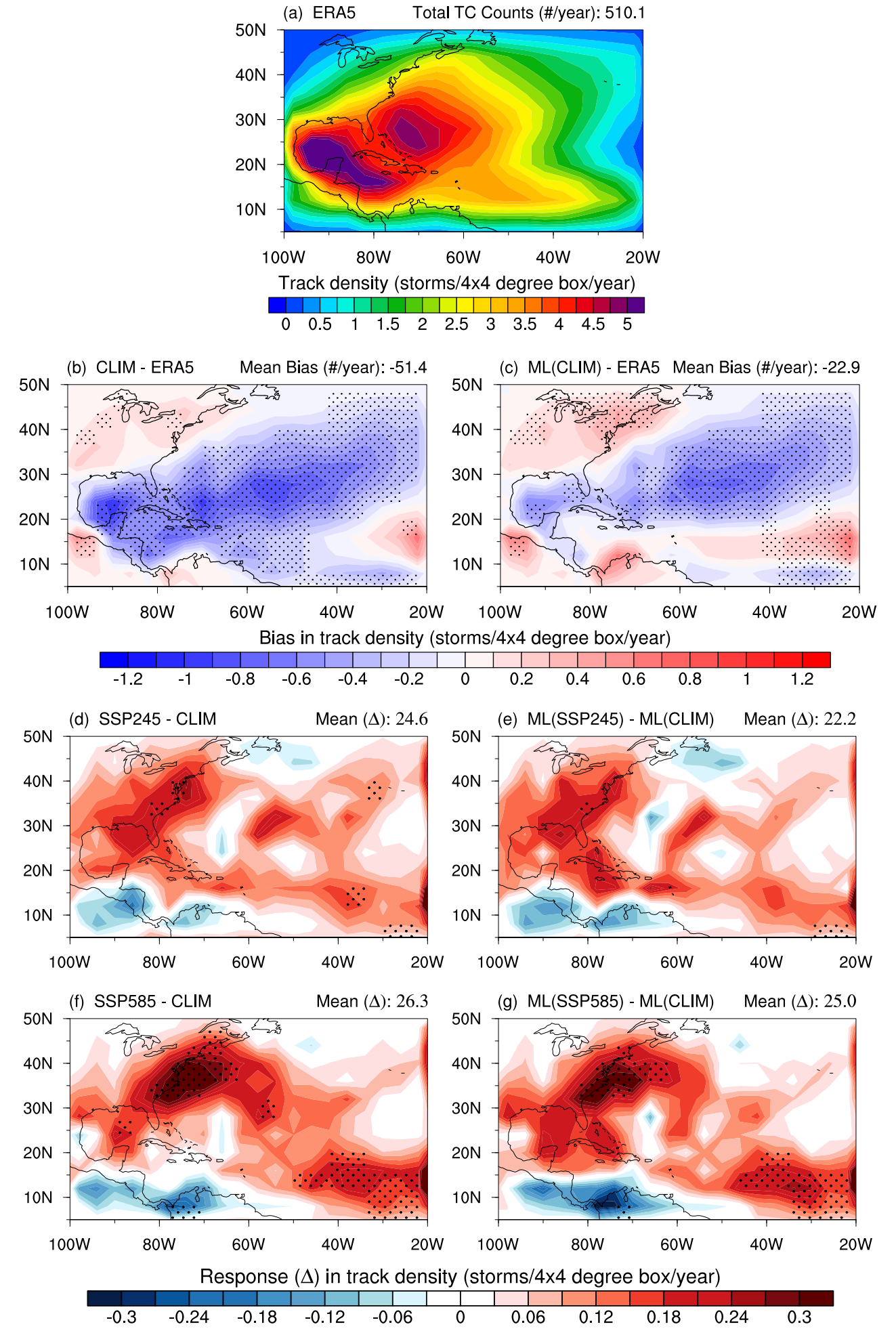

(a) ERA5
Total TC Counts (#/year): 510.1
50N
40N
30N
20N
10N
100W
80W
60W
40W
20W
Track density (storms/4x4 degree box/year)
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
(b) CLIM - ERA5
Mean Bias (#/year): -51.4
(c) ML(CLIM) - ERA5
Mean Bias (#/year): -22.9
Bias in track density (storms/4x4 degree box/year)
-1.2 -1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1 1.2
(d) SSP245 - CLIM
Mean (Δ): 24.6
(e) ML(SSP245) - ML(CLIM)
Mean (Δ): 22.2
(f) SSP585 - CLIM
Mean (Δ): 26.3
(g) ML(SSP585) - ML(CLIM)
Mean (Δ): 25.0
Response (Δ) in track density (storms/4x4 degree box/year)
-0.3 -0.24 -0.18 -0.12 -0.06 0 0.06 0.12 0.18 0.24 0.3

Figure A1.

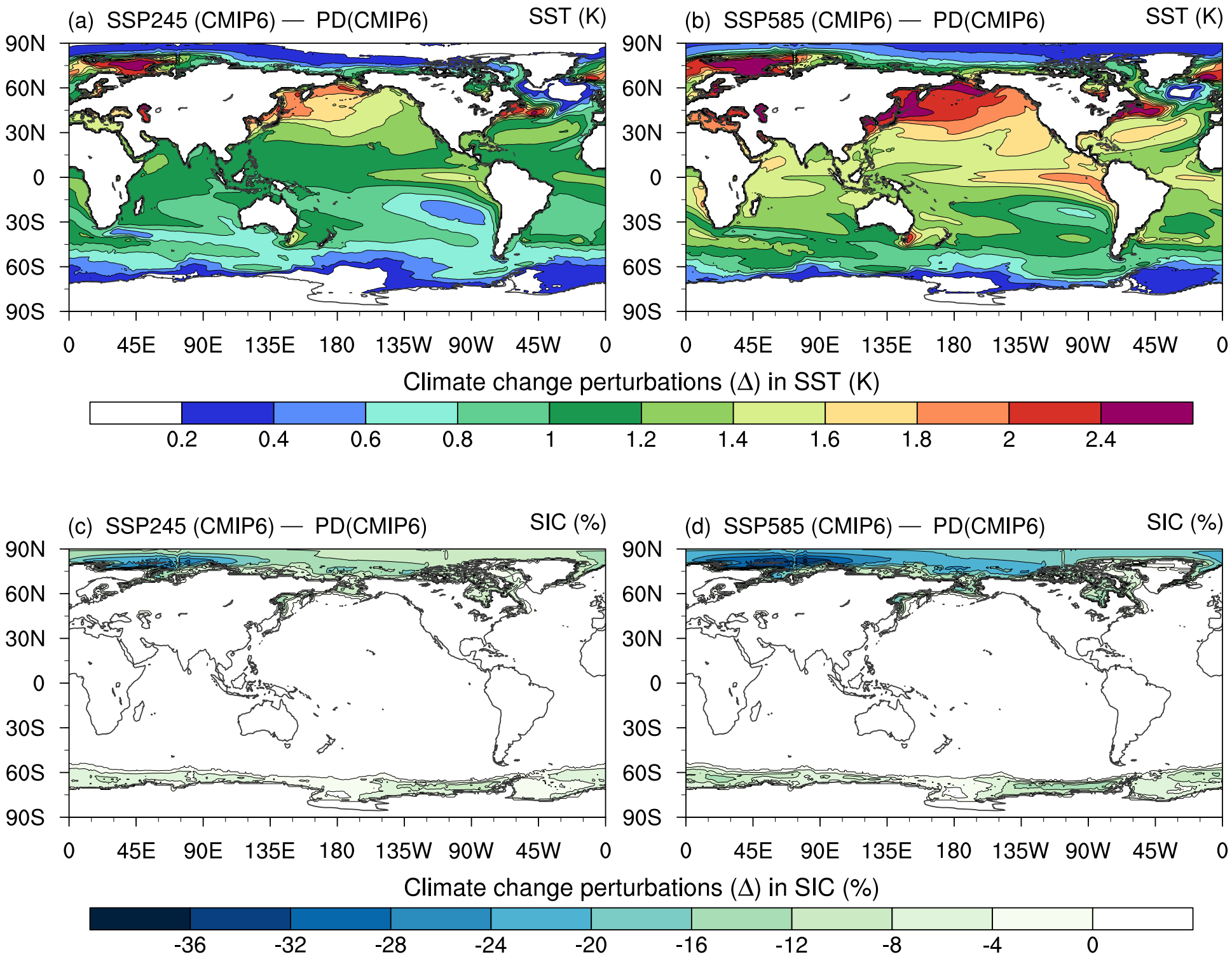

(a) SSP245 (CMIP6) — PD(CMIP6)
SST (K)
(b) SSP585 (CMIP6) — PD(CMIP6)
SST (K)
90N
60N
30N
0
30S
60S
90S
0
45E
90E
135E
180
135W
90W
45W
0
Climate change perturbations (Δ) in SST (K)
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
2.4
(c) SSP245 (CMIP6) — PD(CMIP6)
SIC (%)
(d) SSP585 (CMIP6) — PD(CMIP6)
SIC (%)
Climate change perturbations (Δ) in SIC (%)
-36
-32
-28
-24
-20
-16
-12
-8
-4
0

Figure B1.

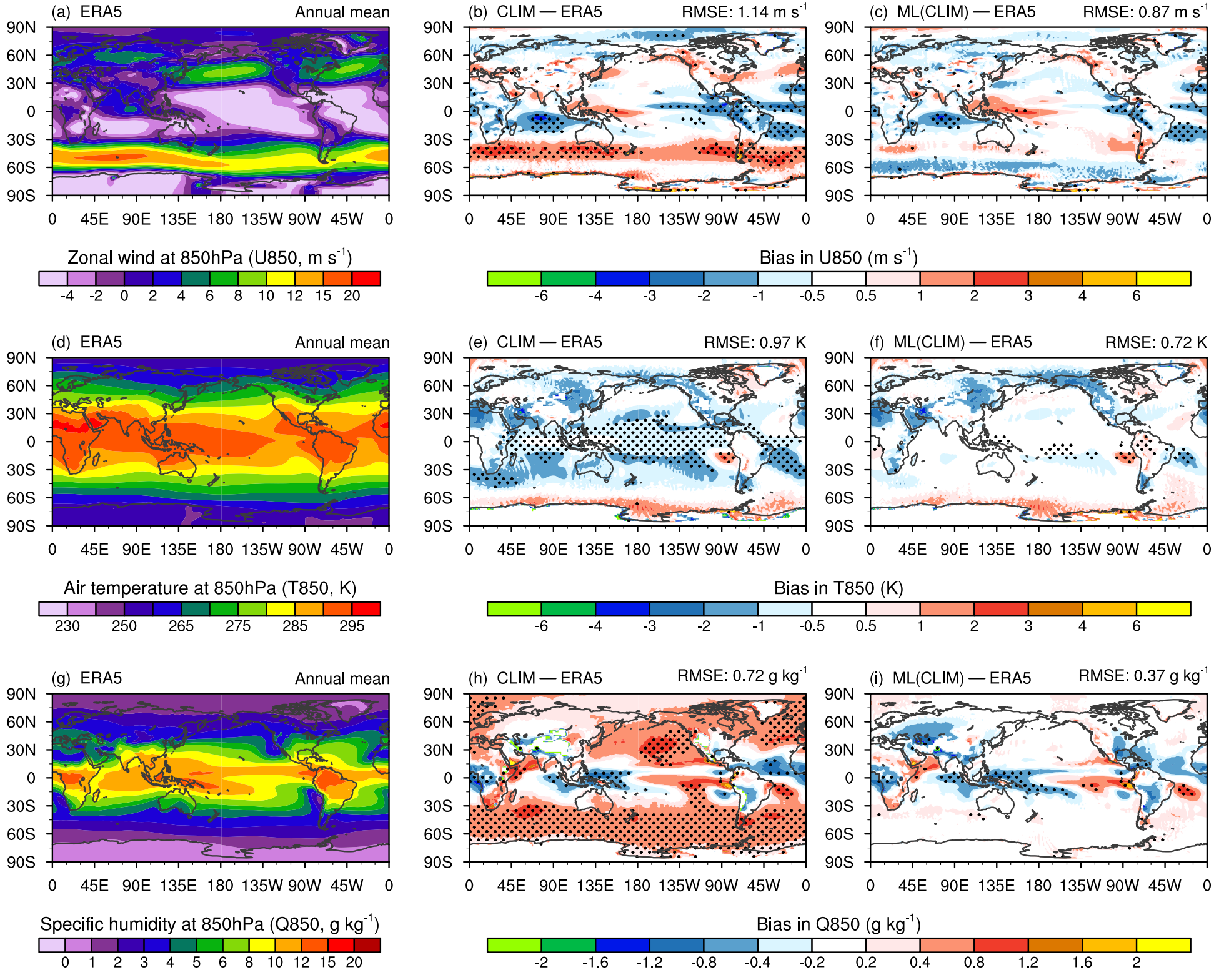

(a) ERA5
Annual mean
(b) CLIM — ERA5
RMSE: 1.14 m s-1
(c) ML(CLIM) — ERA5
RMSE: 0.87 m s-1
90N
60N
30N
0
30S
60S
90S
0 45E 90E 135E 180 135W 90W 45W 0
Zonal wind at 850hPa (U850, m s-1)
-4 -2 0 2 4 6 8 10 12 15 20
Bias in U850 (m s-1)
-6 -4 -3 -2 -1 -0.5 0.5 1 2 3 4 6
(d) ERA5
Annual mean
(e) CLIM — ERA5
RMSE: 0.97 K
(f) ML(CLIM) — ERA5
RMSE: 0.72 K
Air temperature at 850hPa (T850, K)
230 250 265 275 285 295
Bias in T850 (K)
-6 -4 -3 -2 -1 -0.5 0.5 1 2 3 4 6
(g) ERA5
Annual mean
(h) CLIM — ERA5
RMSE: 0.72 g kg-1
(i) ML(CLIM) — ERA5
RMSE: 0.37 g kg-1
Specific humidity at 850hPa (Q850, g kg-1)
0 1 2 3 4 5 6 8 10 12 15 20
Bias in Q850 (g kg-1)
-2 -1.6 -1.2 -0.8 -0.4 -0.2 0.2 0.4 0.8 1.2 1.6 2

Figure B2.

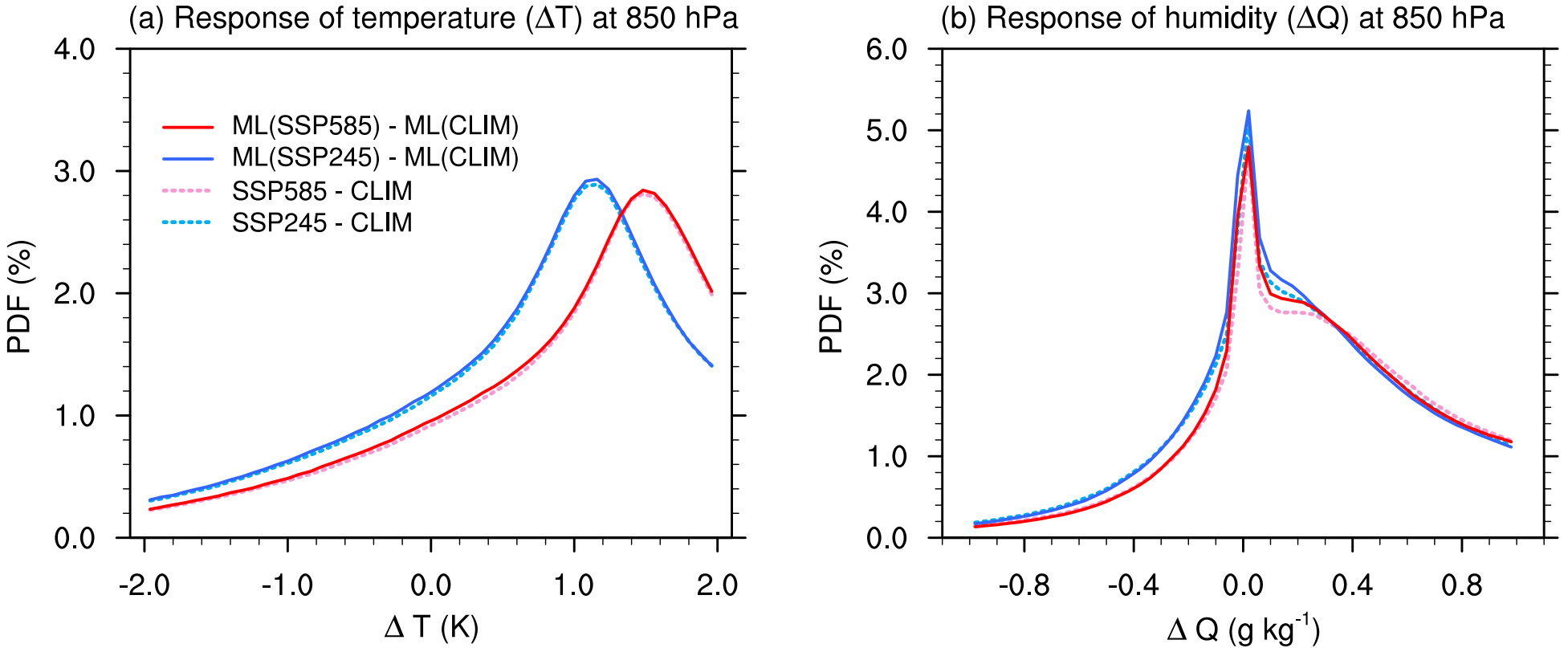

(a) Response of temperature (ΔT) at 850 hPa
ML(SSP585) - ML(CLIM)
ML(SSP245) - ML(CLIM)
SSP585 - CLIM
SSP245 - CLIM
PDF (%)
Δ T (K)
(b) Response of humidity (ΔQ) at 850 hPa
PDF (%)
Δ Q (g kg-1)

Figure B3.

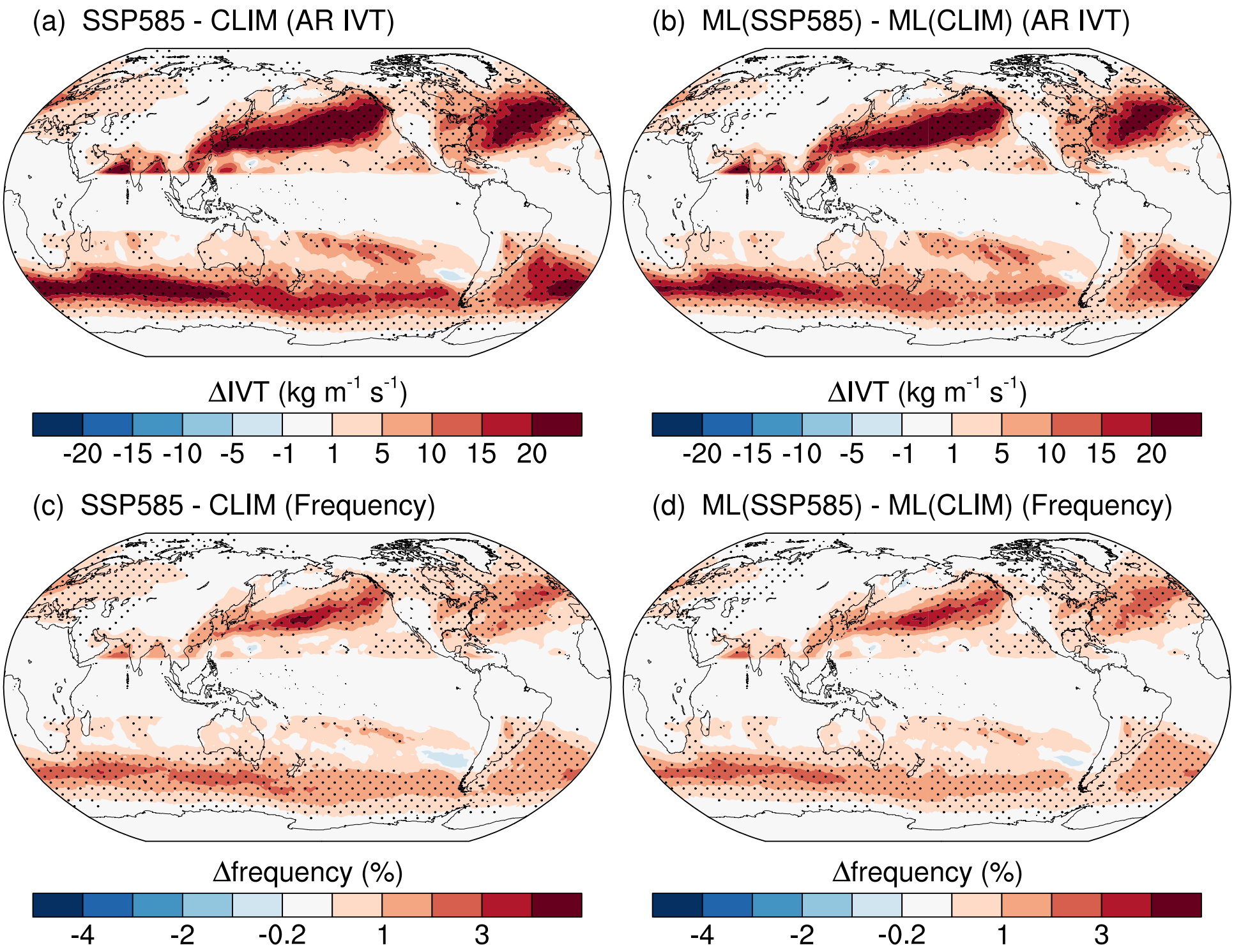

(a) SSP585 - CLIM (AR IVT)
(b) ML(SSP585) - ML(CLIM) (AR IVT)
ΔIVT (kg m-1 s-1)
-20 -15 -10 -5 -1 1 5 10 15 20
ΔIVT (kg m-1 s-1)
-20 -15 -10 -5 -1 1 5 10 15 20
(c) SSP585 - CLIM (Frequency)
(d) ML(SSP585) - ML(CLIM) (Frequency)
Δfrequency (%)
-4 -2 -0.2 1 3
Δfrequency (%)
-4 -2 -0.2 1 3

Figure B4.

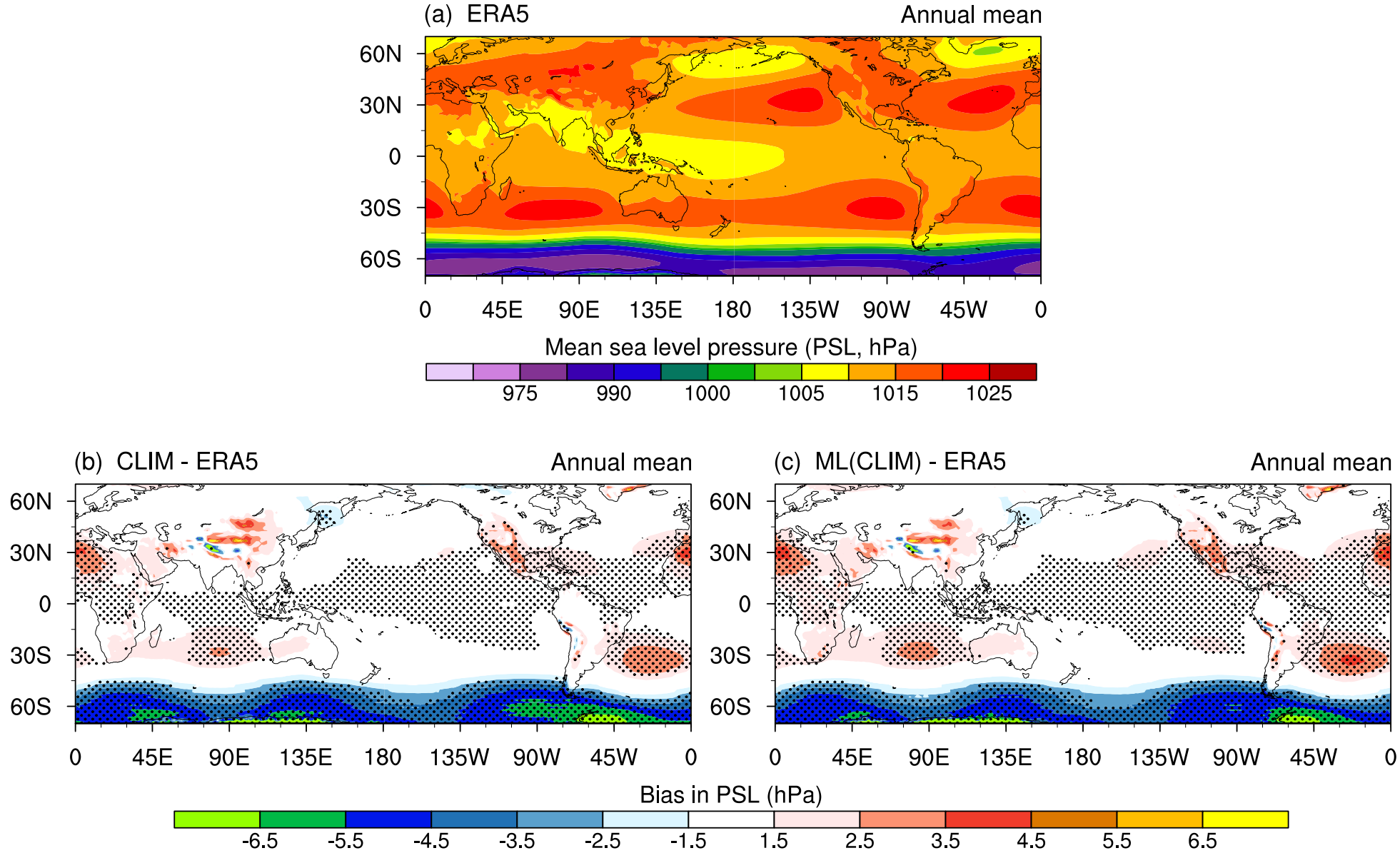

(a) ERA5
Annual mean
60N
30N
0
30S
60S
0
45E
90E
135E
180
135W
90W
45W
0
Mean sea level pressure (PSL, hPa)
975
990
1000
1005
1015
1025
(b) CLIM - ERA5
Annual mean
(c) ML(CLIM) - ERA5
Annual mean
Bias in PSL (hPa)
-6.5
-5.5
-4.5
-3.5
-2.5
-1.5
1.5
2.5
3.5
4.5
5.5
6.5

Figure B5.

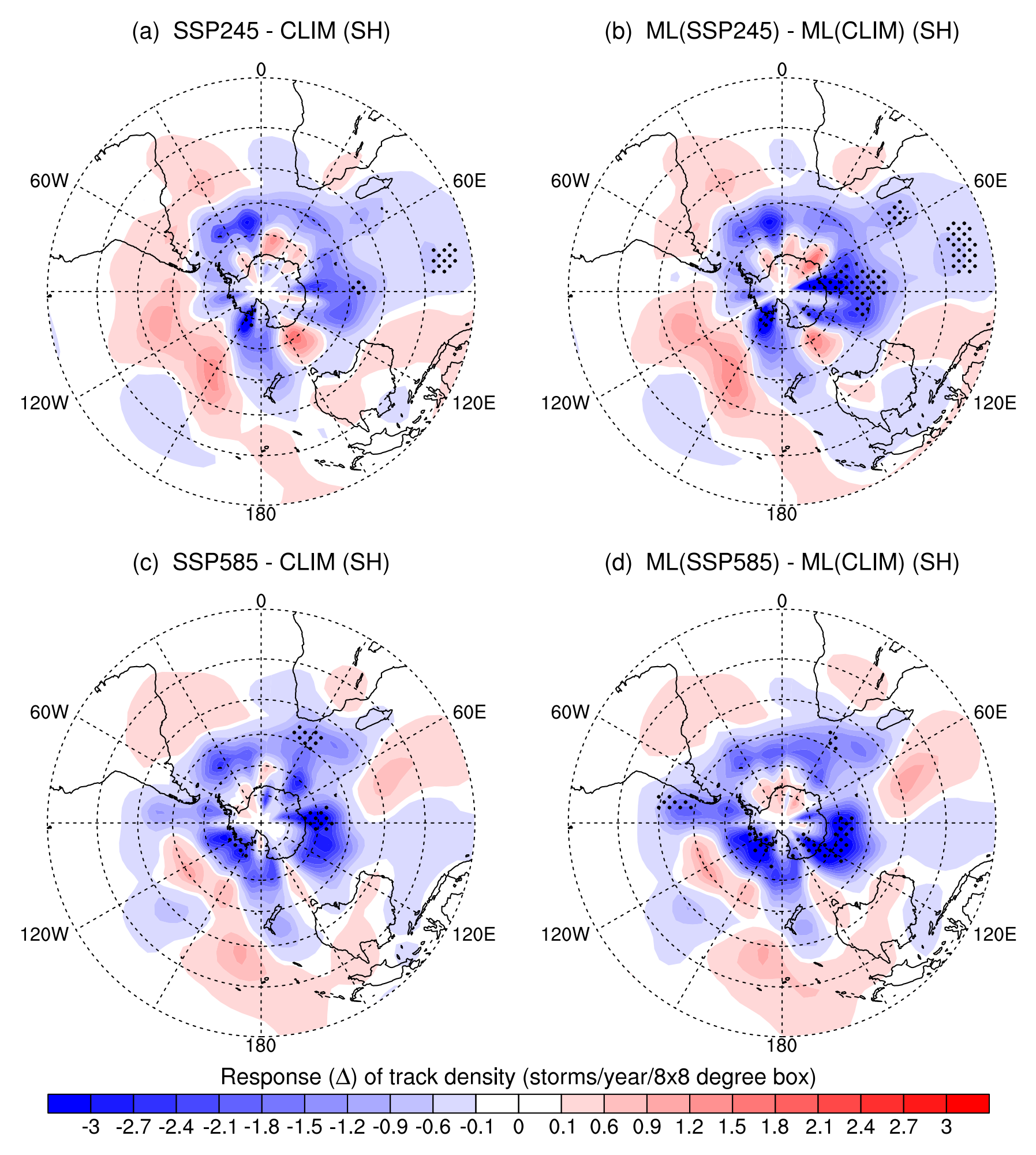
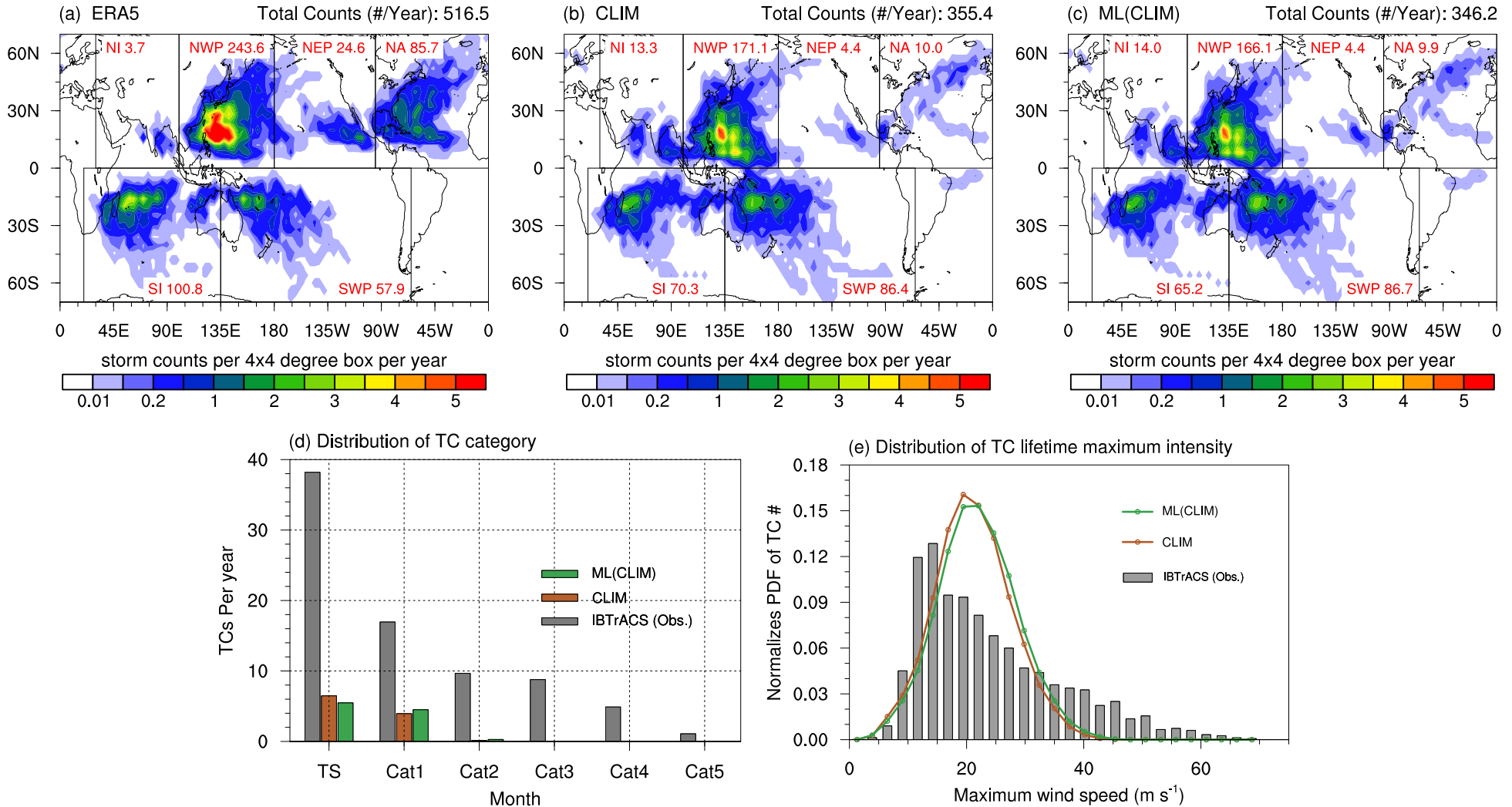

(a) SSP245 - CLIM (SH)
(b) ML(SSP245) - ML(CLIM) (SH)
(c) SSP585 - CLIM (SH)
(d) ML(SSP585) - ML(CLIM) (SH)
0
60W
60E
120W
120E
180
Response (Δ) of track density (storms/year/8x8 degree box)
-3 -2.7 -2.4 -2.1 -1.8 -1.5 -1.2 -0.9 -0.6 -0.1 0 0.1 0.6 0.9 1.2 1.5 1.8 2.1 2.4 2.7 3

Figure B6.


(a) ERA5
Total Counts (#/Year): 516.5
NI 3.7
NWP 243.6
NEP 24.6
NA 85.7
SI 100.8
SWP 57.9
(b) CLIM
Total Counts (#/Year): 355.4
NI 13.3
NWP 171.1
NEP 4.4
NA 10.0
SI 70.3
SWP 86.4
(c) ML(CLIM)
Total Counts (#/Year): 346.2
NI 14.0
NWP 166.1
NEP 4.4
NA 9.9
SI 65.2
SWP 86.7
60N
30N
0
30S
60S
0 45E 90E 135E 180 135W 90W 45W 0
storm counts per 4x4 degree box per year
0.01 0.2 1 2 3 4 5
(d) Distribution of TC category
TCs Per year
ML(CLIM)
CLIM
IBTrACS (Obs.)
TS Cat1 Cat2 Cat3 Cat4 Cat5
Month
(e) Distribution of TC lifetime maximum intensity
Normalizes PDF of TC #
Maximum wind speed (m s-1)