Bias in CMIP6 historical U.S. severe convective environments driven by bias in mean-state near-surface moist static energy

Daniel Chavas[1] and Funing LI[2]

[1]Purdue University
[2]PURDUE UNIVERSITY

November 22, 2022

Abstract

This work evaluates how well Coupled Model Intercomparison Project 6 (CMIP6) models reproduce the climatology of North American SCS environments in ERA5 reanalysis and examines what drives biases across models. Biases in Springtime SCS environments vary widely in magnitude and spatial pattern, but most models do well in reproducing the climatological pattern and a few also reproduce the overall magnitude. SCS bias is driven by bias in extreme CAPE. This bias is ultimately found to be driven by bias in mean-state near-surface moist static energy (MSE), indicating that the SCS environments depend strongly on the near-surface mean state. Results are broadly similar to Spring across all seasons, particularly Summer. Biases differ strongly across parent models but weakly across child models of the same parent. These outcomes help identify models well-suited for studying climate effects on SCS environments and also provide a foundation for improving model performance in the future.

Bias in CMIP6 historical U.S. severe convective environments driven by bias in mean-state near-surface moist static energy

Daniel R. Chavas[1], **Funing Li**[1]

[1]Purdue University, Department of Earth, Atmospheric, and Planetary Sciences, West Lafayette, IN

Key Points:

- Most models reproduce severe convective storm (SCS) environment pattern over eastern North America and some also reproduce the magnitude.
- SCS environment bias is driven by mean-state near-surface moist static energy bias.
- Hence, simulating SCS environments well requires simulating the near-surface mean state air properties well.

Corresponding author: Daniel R. Chavas, `drchavas@gmail.com`

Abstract

This work evaluates how well Coupled Model Intercomparison Project 6 (CMIP6) models reproduce the climatology of North American SCS environments in ERA5 reanalysis and examines what drives biases across models. Biases in Springtime SCS environments vary widely in magnitude and spatial pattern, but most models do well in reproducing the climatological pattern and a few also reproduce the overall magnitude. SCS bias is driven by bias in extreme CAPE. This bias is ultimately found to be driven by bias in mean-state near-surface moist static energy (MSE), indicating that the SCS environments depend strongly on the near-surface mean state. Results are broadly similar to Spring across all seasons, particularly Summer. Biases differ strongly across parent models but weakly across child models of the same parent. These outcomes help identify models well-suited for studying climate effects on SCS environments and also provide a foundation for improving model performance in the future.

Plain Language Summary

Climate models are useful tools for studying how severe thunderstorms may change with climate change, Models cannot simulate the storms themselves but can simulate environments that support severe thunderstorms. Using the most recent set of models used to simulate future projections of climate change, we find that some models can simulate the historical climatology of these environments very well, while others do not. We also show that model errors in severe thunderstorm environments arise primarily due to errors in the mean energy content of near-surface air, particularly that associated with water vapor. Hence, simulating these extreme environments well depends strongly on simulating average conditions well.

1 Introduction

An important question of societal significance is how severe convective storms (SCS), including severe thunderstorms and tornadoes, will change in the future under climate change (Ashley, 2007). Climate models are a useful tool for studying how a variety of weather phenomena may change with climate change. SCS events occupy very small spatial scales, though, and as a result climate models cannot directly resolve such events. Instead, it is common to use environmental proxies that indicate whether a given column of air is favorable for supporting SCS activity (Ludlam, 1963; Johns & Doswell III, 1992). Favorability is most commonly defined in terms of two "ingredients": one thermodynamic ingredient given by Convective Available Potential Energy (CAPE), and one kinematic ingredient given by the bulk 0-6 km wind difference (S06; a.k.a. "bulk shear") (Rasmussen & Blanchard, 1998; Brooks et al., 2003; Gensini & Ashley, 2011). High values of the product of the two quantities is a widely used proxy for a favorable environment for SCS activity (hereafter "SCS environments"). These proxies calculated from environmental data have been successfully used to explain spatiotemporal variations in severe thunderstorm activity and their associated hazards using reanalysis data (Taszarek et al., 2020, 2021; Coffer et al., 2020) and in individual global climate model simulations (Li et al., 2021; Hoogewind et al., 2017; Chen et al., 2020).

Global climate models from Coupled Model Intercomparison Project 5 (CMIP5) were found to vary widely in their ability to reproduce the historical climatology of severe convective storm (SCS) environments found in reanalysis over North America (Seeley & Romps, 2015). Recent work has also examined how SCS environments may change in the future across the more recent CMIP6 model suite (Lepore et al., 2021). However, given the wide range of variability in the performance of CMIP5 climate models, it is important to quantify the biases of individual models compared to reanalysis in CMIP6 as well. Moreover, understanding what are the key drivers of those model biases in SCS environments can help identify precisely what underlying aspects of the thermodynamic or kinematic environments are biased, and possibly to provide an avenue for improving models to reduce those biases in the future.

This work has two objectives. First, we examine how well CMIP6 models reproduce the historical SCS climatology in ERA5 reanalysis in a manner similar to Seeley and Romps (2015); we focus on Spring but analyze all four seasons. Second, we combine a simple framework for deconstructing biases with a recent theoretical framework for understanding variations in CAPE (Li & Chavas, 2021) to understand the key drivers of biases across models. Section 2 presents the Data and Methodology. Section 3 presents the results. Finally, Section 4 provides a summary and discussion.

2 Methodology

2.1 Data

For historical climate model data, we use 6-hourly model-level data for 1980-2014 from the 13 historical CMIP6 model simulations (Eyring et al., 2016) with the required data to calculate both CAPE and bulk 0–6km shear (list of models provided in Supplementary Table S1). Results within each model are aggregated into Spring (MAM), Summer (JJA), Fall (SON), and Winter (DJF) seasons.

We compare output from each climate model against the ERA5 reanalysis pressure-level data for the identical period (Hersbach et al., 2020). Li et al. (2020) showed that ERA5 reanalysis representation of severe weather environments and parameters over North America compares well against radiosonde observations, including at the extremes. Hence, here we focus on comparing CMIP6 against ERA5.

2.2 Analysis

We define SCS environments by extreme values (99th percentile) of the product of convective available potential energy (CAPE) and 0–6-km bulk vertical wind shear (S06), hereafter 'CAPES06' (Li et al., 2020). CAPE measures conditional instability of an air parcel within an atmospheric column, defined as the vertical integral of buoyancy of a parcel from its level of free convection to its equilibrium level. Here we use the near-surface ($z = 2m$) parcel and calculate CAPE according to (Emanuel (1994) Eq. 6.3.5)

$$CAPE = R_d \int_{p_{EL}}^{p_{LFC}} (T_{\rho,parcel} - T_{\rho,env})\, dln(p) \tag{1}$$

where R_d is the dry gas constant, $T_{\rho,parcel}$ and $T_{\rho,env}$ are the respective density temperatures of the parcel and environment at a given level, p_{LFC} is the pressure at the level of free convection, and p_{EL} is the pressure at the level of neutral buoyancy. We follow Chen et al. (2020) and assume pseudoadiabatic ascent (hence the density temperature is given by the virtual temperature) and neglect the latent heat of freezing. S06 represents lower-tropospheric environmental horizontal vorticity available to generate updraft vertical vorticity, defined as the magnitude of the wind vector difference between 6 km and 10 m above the surface:

$$S06 = |\vec{\mathbf{u}}_{6km} - \vec{\mathbf{u}}_{10m}| \tag{2}$$

We define mean bias in a given quantity x (e.g. CAPES06) as

$$bias = 100 \left(\frac{\Delta x}{x_{ERA5}} \right) \tag{3}$$

where Δ denotes the difference between model and reanalysis ($\Delta x = x_{model} - x_{ERA5}$), and the factor of 100 translates bias to a percent bias in the model relative to reanalysis. Following Li and Chavas (2021), we decompose the bias in CAPES06 of each CMIP6 model relative to ERA5 according to

$$\frac{\Delta CAPES06}{CAPES06} = \frac{\Delta CAPE}{CAPE} + \frac{\Delta S06}{S06} + \frac{\Delta CAPE}{CAPE} \cdot \frac{\Delta S06}{S06} + \epsilon, \tag{4}$$

where we drop the ERA5 subscript in the denominator to simplify the notation. A given term may be multiplied by 100 to express it as a percent bias as in Eq. 3. This decomposition method enables us to quantify conditional bias contributions to CAPES06 from CAPE (first RHS term) and S06 (second); the third term is the conditional bias product term and ϵ is the residual. These biases are 'conditional' because the subset of grid points used for their calculation is conditioned on a particular criteria for CAPES06 (e.g. top 1%). This decomposition is a common approach to understand how changes in the product of two or more quantities depend on changes in each component (e.g., Bony et al. (2004); Emori and Brown (2005); Chen and Chavas (2020)).

As will be shown below, CAPE biases are the dominant driver of biases in SCS environments. Thus, we would like to understand what drives these biases. However, CAPE is a vertically-integrated quantity whose calculation typically requires lifting a hypothetical near-surface parcel through the depth of the troposphere, which is difficult to decompose. Instead, Li and Chavas (2021) demonstrated that variations in CAPE scale closely with variations in a CAPE-like quantity, initially proposed by Agard and Emanuel (2017), that depends only on bulk properties of the thermodynamic profile. This scaling CAPE, which we denote CAPE$_s$, is given by

$$CAPE_s = (M_{sfc} - \overline{D_{FT}}) ln \frac{T_{BT}}{T_{trop}} \tag{5}$$

where M_{sfc} is surface moist static energy, given by $M_{sfc} = c_p T_{v,2m} + g z_{sfc} + L_v q_{2m}$; c_p is the specific heat of air at constant pressure; g is the acceleration due to gravity; L_v is the latent heat of vaporization of water; T_{v2m} is the 2-m virtual air temperature; z_{sfc} is the surface geopotential height; and q_{2m} is the 2-m specific humidity. $\overline{D_{FT}}$ is the mean free tropospheric dry static energy, given by $\overline{D_{FT}} = c_p \overline{T_{vFT}} + g \overline{z_{FT}}$, where $\overline{T_{vFT}}$ and $\overline{z_{FT}}$ are the mean free tropospheric virtual temperature and geopotential height, respectively, each weighted by the natural logarithm of virtual air temperature. T_{BT} and T_{trop} are the virtual temperatures at boundary-layer top and tropopause, respectively. The boundary-layer top is defined as the level where the vertical gradient of relative humidity is the minimum in the lower 2500-m atmosphere (Aryee et al., 2020). The tropopause is defined as the lowest level within the 85-450 hPa layer where the lapse rate decreases to less than 2 K km^{-1} and the average lapse rate between this level and all higher levels over a 100-hPa depth is less than 2 K km^{-1}(WMO/OMM/BMO, 1992); this method is consistent with the function *trop_wmo* in the open-source NCAR Command Language (NCL).

Li and Chavas (2021) showed that CAPE$_s$ scales very closely with CAPE over the North American continent in the MERRA2 reanalysis, even at the extremes. CAPE$_s$ is simpler than CAPE mathematically, as it does not require a vertical integral. Thus, it is much more straightforward to use to decompose the different terms that contribute to variations in CAPE.

Given their close linear scaling, we may relate CAPE$_s$ to CAPE according to

$$CAPE \approx a \cdot CAPE_s + b \tag{6}$$

As the intercept b is in general small, the bias in CAPE approximates the bias in CAPE$_s$:

$$\frac{\Delta CAPE}{CAPE} \approx \frac{\Delta CAPE_s}{CAPE_s} \tag{7}$$

Substituting Eq. 5 into Eq. 7 and simplifying yields an equation for the contributions of each component of CAPE$_s$ to the total bias in CAPE$_s$:

$$\frac{\Delta CAPE_s}{CAPE_s} = \frac{\Delta M_{sfc}}{M_{sfc} - \overline{D_{FT}}} + \frac{-\Delta \overline{D_{FT}}}{M_{sfc} - \overline{D_{FT}}} + \frac{\Delta ln \frac{T_{BT}}{T_{trop}}}{ln \frac{T_{BT}}{T_{trop}}} + \frac{(\Delta M_{sfc} - \Delta \overline{D_{FT}}) \Delta ln \frac{T_{BT}}{T_{trop}}}{(M_{sfc} - \overline{D_{FT}}) ln \frac{T_{BT}}{T_{trop}}} + \epsilon \tag{8}$$

The right hand side represents contributions to bias in CAPE$_s$ from biases in M_{sfc} (first term), $\overline{D_{FT}}$ (second term), log-temperature term (third term); the fourth term is the bias product term and ϵ is the residual.

Finally, for the purpose of our analysis below, we further decompose the M_{sfc} bias term linearly into biases in near-surface sensible heat (temperature) and latent heat (moisture):

$$\frac{\Delta M_{sfc}}{M_{sfc} - \overline{D_{FT}}} = \frac{c_p \Delta T_{v2m}}{M_{sfc} - \overline{D_{FT}}} + \frac{L_v \Delta q_{2m}}{M_{sfc} - \overline{D_{FT}}} \tag{9}$$

We will use the above equations to understand how biases in near-surface temperature or moisture can drive biases in CAPE. In our analysis below, all biases include the multiplicative factor 100 to translate them to percent biases relative to reanalysis.

3 Results

In the presentation and discussion of our results below, we focus on Spring as the principal season for SCS activity. We then describe notable differences in other seasons (Supplementary figures), as severe weather is also common in other seasons, particularly in the northern Great Plains Summer and in the southeast US Winter (Hoogewind et al., 2017; Long et al., 2018).

3.1 Comparison of CMIP6 models against ERA5

We first evaluate CMIP6 model performance against ERA5 in reproducing the climatological Springtime spatial distribution of SCS environments defined by extreme CAPES06 (Fig. 1). We define 'extreme' as the 99th percentile at each gridpoint for a given model or ERA5 on their original grid based on the 6-hourly time series (00, 06, 12, 18 UTC) for 1980–2014, which is the finest temporal resolution available in CMIP6 models. The result across all gridpoints are then linearly interpolated to 1x1-deg grids using python package scipy.interpolate.griddata. Performance across models is summarized in Fig. 1b in terms of explained variance (r^2, x-axis) and mean bias (y-axis; [%]), similar to Seeley and Romps (2015) (their Fig. 2); values are listed in Supplementary Table S2. For this quantitative calculation and all subsequent calculations below, we use the subset of all land grid points within a domain that spans much of the United States east of the Rocky Mountains (pink box in Fig. 1), and grid points are excluded that have a seasonal ERA5 value for extreme (99th percentile) CAPE below 150 J/kg in order to avoid locations where extreme CAPE is very low and hence severe weather is simply not expected to occur at all. The latter criterion only removes roughly the northern half of our domain of interest in Winter and a few grid northern grid points near the Great Lakes in Spring, and it has no effect on other seasons. Mean bias is calculated from this subset of gridpoints by calculating the mean difference between model and ERA5 and then dividing by the mean in ERA5 (i.e. $100 * \frac{\overline{\Delta x}}{\overline{x}_{ERA5}}$), where the mean is weighted by cosine of latitude to account for variations in surface area with latitude.

Compared to ERA5 reanalysis, the CMIP6 historical simulations exhibit a wide range of biases in both magnitude and spatial distribution of extreme CAPES06 (Fig.

1a), similar to that found in CMIP5 (Seeley & Romps, 2015). The majority of the models do reasonably well in reproducing the basic spatial distribution found in ERA5, with CAPES06 largest over eastern Mexico and southern Texas and decreasing moving northwards through the Great Plains and eastward toward the Ohio Valley. Pattern correlations yield r^2 values that range from 0.26 to 0.94, though the majority of models (9/13) have r^2 values greater than 0.75 (Fig. 1b). A smaller number of models reproduce the overall magnitude well, with only 6/13 having bias magnitudes less than 40%. All models are biased high (positive bias), indicating they all overestimate the magnitude of CAPES06 over the central and eastern U.S. Pattern correlation and mean bias are correlated with one another such that models that better reproduce the magnitude also tend to better reproduce the spatial pattern.

Model performance differs across parent models (MPI, CNRM, MIROC) but is quite similar across child models within a given parent model. For example, all three MPI models reproduce both the pattern (high pattern r^2) and magnitude (low mean bias) very well (Fig. 1b). The two CNRM models also reproduce the magnitude with only slightly reduced performance in reproducing the spatial distribution. Meanwhile, the two MIROC models capture the spatial distribution reasonably well but with a larger overestimation of the magnitude. These outcomes indicate that the structure of the bias is driven by the deeper architecture of the parent model and is not sensitive to subtler changes among child models, including grid resolution (Supplementary Table S1). Overall, the MPI and CNRM model groups perform very well in reproducing both the magnitude and spatial distribution of severe weather environments found in ERA5.

The spatial distribution of bias in extreme CAPES06 closely mirrors that of extreme CAPE alone across models (Fig. 1c), especially for the high-biased models. This result suggests that bias in CAPES06 is primarily driven by bias in CAPE. To test this quantitatively, we use Eq. 4 to decompose bias in CAPES06 into contributions from CAPE and S06 (Fig. 2). At each gridpoint, we first extract the top 1% of cases of CAPES06 (and its associated values of CAPE and S06) in a given season from a model or ERA5. From this subsample, we then calculate the median of CAPES06 (hence this is the 99.5th percentile of CAPES06) and of CAPE and S06 associated with this CAPES06 subsample (hence this is *not* the 99.5th percentiles of CAPE or S06). Finally we follow the same method as was done for CAPES06 in Fig. 1b to calculate the mean bias for each quantity. The result is a single domain-wide measure of model mean bias in each quantity that is input into Eq. 4 to decompose CAPES06 bias into contributions from conditional CAPE (grey bar) and S06 (yellow bar) in Fig. 1d. We also repeat this process for the top 1% of cases of CAPE (i.e. unconditional bias in CAPE; black bar) for comparison with conditional bias in CAPE.

Variance in conditional CAPE bias (grey bar) across models accounts for 94% of the variance in CAPES06 bias (red bar) across models (Fig. 1d). The conditional S06 bias is relatively small, which indicates that shear bias does not play a significant role. The bias product and the residual terms are negligible. CAPE bias is uniformly positive, reflecting the fact that CMIP6 models systematically overestimate CAPE over the eastern half of the U.S. (Fig. 1a) similar to CAPES06 above. Results are quantitatively similar across all other seasons (Summer Fig. S1; Fall Fig. S2; Winter Fig. S3). In Fall, a few low-bias models exhibit a non-negligible S06 bias that largely offsets the CAPE bias. Finally, conditional CAPE bias (grey bar) covaries closely with unconditional CAPE bias (black bar; $r^2 = 0.99$), with the latter typically slightly overestimating the former. Indeed, nearly all of the the top 1% of CAPE cases within our domain of interest are associated with the top 10% of CAPES06 cases. This outcome indicates that the top 1% of CAPE cases, while not an identical subset to that of CAPES06, are still sampled from extreme SCS environments. As a result, we may understand bias in CAPE conditioned on extreme CAPES06 via bias in unconditional extreme CAPE, which we examine next.

3.2 Decomposing variation in CAPE across models using CAPE$_s$

Given that CAPE bias is the dominant contributor to CAPES06 bias, we next investigate the drivers of bias in extreme CAPE via variations in extreme CAPE$_s$. As noted above, CAPE$_s$ has been shown in MERRA2 reanalysis to scale very closely with CAPE over the North American continent and is much more straightforward mathematically to understand its variability Li and Chavas (2021).

In ERA5, CAPE$_s$ is also found to scale very closely with CAPE over the North American continent (Fig. 2a-c; $r^2 = 0.99$). This result is consistent across seasons as well, particularly Summer (Fig. S4a-c; $r^2 = 0.98$) and Fall (Fig. S5a-c; $r^2 = 0.99$), with only a slight decrease in explained variance in Winter (Fig. S6a-c; $r^2 = 0.94$). Hence CAPE$_s$ captures not only the spatial pattern of CAPE but also its seasonal cycle.

Across CMIP6 models, extreme CAPE$_s$ scales closely with extreme CAPE over the eastern U.S as well (Fig. 2d), with pattern r^2 values exceeding 0.85 and all but one model exceeding 0.9. The linear scaling factor a between CAPE and CAPE$_s$ lies between 0.5–0.65 across models, with most clustered between 0.525–0.6, indicating that the linear relationship is quite similar across models. The results indicate that variations in CAPE$_s$ may be used to understand variations in CAPE within and across models. Across models, the spatial-mean values of extreme CAPE and CAPE$_s$ also covary strongly (Fig. 2e; $r^2 = 0.94$), indicating that the magnitude of bias in CAPE is very well captured by that of CAPE$_s$. In other seasons, results are similar though relationships are slightly weaker, with pattern r^2 values exceeding 0.8 for Summer (Fig. S4), 0.85 for Fall (Fig. S5), and 0.75 for Winter (Fig. S6).

Next, we compare biases in CAPE and CAPE$_s$ and then decompose CAPE$_s$ (Eq. 8) to understand what drives bias in CAPE. We use the same approach (median of top 1%) as was used to decompose CAPES06 in Fig. 1d. Following from Eq. 7, mean bias in extreme CAPE across models covaries closely with mean bias in extreme CAPE$_s$ (Fig. 3a; $r^2 = 0.93$), especially for high-biased models. This CAPE$_s$ bias is primarily driven by bias in surface moist static energy, M$_{sfc}$ (Fig. 3b; $r^2 = 0.88$). Notably, there are a couple of models (e.g. CNRM) that have relatively small bias in $CAPE_s$ but large errors in M$_{sfc}$ and D$_{FT}$ that largely offset one another.

We further decompose variations in M$_{sfc}$ bias into its contributions from temperature (sensible heat; black bar) and moisture (latent heat; grey bar) using Eq. 9 (Fig. 3c). For models that exhibit high bias in M$_{sfc}$, this bias is driven principally by moist bias. Meanwhile, for models that exhibit low bias in M$_{sfc}$, this bias is driven principally by cold bias. Temperature and moisture biases covary strongly with one another and have the same sign within a given model in all but one model, indicating that their effects act in concert to drive bias in M$_{sfc}$.

Finally, we examine whether this bias in SCS environments can be directly tied to bias in the mean state over the central and eastern U.S. Across models, mean bias in M$_{sfc}$ associated with extreme CAPE$_s$ environments is found to scale closely with mean bias in mean-state M$_{sfc}$ (Fig. 4a; $r^2 = 0.95$). Moreover, the finding of a strong positive covariance between moist and warm biases in SCS environments extends to the mean state as well (Fig. 4b; $r^2 = 0.95$), as models tend to be either biased warm and moist or biased cool and dry, with preference for larger warm and moist bias (and hence high M_{sfc} bias). Hence, while our focus is on bias in extreme environments associated with severe convective storms, this bias is intimately linked to bias in the properties of mean state near-surface air. This finding suggests that a model's ability to correctly reproduce these relatively rare but impactful environments depends on its ability to correctly reproduce the low-level background state.

For other seasons, bias in CAPE across models covaries closely with bias in CAPE$_s$ in Summer, Fall, and Winter as well (Fig. S7-S9, respectively). However, the drivers of bias become more complex, as M_{sfc} explains less variance in CAPE$_s$ in Summer (58%), Fall (60%), and Winter (68%). Bias in D$_{FT}$ becomes more significant and occasionally dominant in driving bias in $CAPE_s$ in some models, indicating a greater role for variations in the temperature structure (mean temperature and depth) of the free troposphere outside of Spring. However, in Fall and Winter, biases in M$_{sfc}$ (low bias; cold, dry) and D$_{FT}$ (high bias; warm, deep) are often large and oppose each other. Finally, bias in M_{sfc} associated with extreme CAPE$_s$ also scales closely with bias in mean-state M$_{sfc}$ across seasons (Fig. S10-12), with a similar very close relationship in Fall (Fig. S11; $r^2 = 0.93$) and a somewhat weaker relationship in Summer (Fig. S10; $r^2 = 0.57$) and Winter (Fig. S12; $r^2 = 0.51$).

4 Discussion

This work has demonstrated that most CMIP6 models can reproduce the spatial structure of the historical SCS environment climatology over central and eastern North America, and some of those models (MPI and CNRM) reproduce the overall magnitude, too, though always with a high bias. Bias in SCS environments is driven by bias in extreme CAPE. The latter is driven principally by near-surface moist static energy bias in the mean state, particularly in Spring. Biases differ strongly across parent models but weakly across child models of the same parent, suggesting that the underlying cause of the bias lies in the deeper architecture of the parent model rather than subtle variations among child models, including grid resolution.

Our results help identify models that can more faithfully reproduce the spatial structure and amplitude of the climatology of SCS environments over North America and hence may be better suited for studying how these environments may change with climate. Linking bias in extreme SCS environments (i.e. from the tail of the distribution of CAPE*S06) to bias in the near-surface mean state provides an avenue to understand the physics that generate this model bias in the context of the climate system as a whole via e.g. energy budgets. Our results suggest that land surface properties in a model may play an important role in its ability to reproduce not only the mean state but also the extremes. Finally, the distinct behavior of parent vs. child model could be used to identify specific aspects of a model's architecture (e.g. dynamical core, physics parameterizations) that drive bias in the mean state and, in turn, SCS environments.

5 Open Research

The data used to generate the figures in the manuscript are available at `https://doi.org/10.4231/42ZJ-A891` (Chavas & Li, 2022). 6-hourly pressure-level ERA5 reanalysis data were accessed from `https://doi.org/10.5065/BH6N-5N20` (European Centre for Medium-Range Weather Forecasts, 2019). 6-hourly CMIP6 model historical experiment data were accessed from `https://esgf-node.llnl.gov/search/cmip6`; model information is detailed in Table S1. Analyses were performed on the NCAR Cheyenne and Casper supercomputers (Computational and Information Systems Laboratory, 2019) as well as on computational resources provided by Purdue Rosen Center for Advanced Computing (RCAC) (McCartney et al., 2014).

Acknowledgments

DC and FL were funded by NSF grant 1648681, NASA grant 19-EARTH20-0216, and a grant from the Purdue Covid Disruption Fund. The authors thank the National Science Foundation for their support of NCAR Cheyenne and Purdue University for their support of the Rosen Center for Advanced Computing (RCAC).

References

Agard, V., & Emanuel, K. (2017). Clausius–clapeyron scaling of peak cape in continental convective storm environments. *Journal of the Atmospheric Sciences*, *74*(9), 3043–3054.

Aryee, J., Amekudzi, L., Preko, K., Atiah, W., & Danuor, S. (2020). Estimation of planetary boundary layer height from radiosonde profiles over West Africa during the AMMA field campaign: Intercomparison of different methods. *Scientific African*, *7*, e00228.

Ashley, W. S. (2007). Spatial and temporal analysis of tornado fatalities in the united states: 1880–2005. *Weather and forecasting*, *22*(6), 1214–1228.

Bony, S., Dufresne, J. L., Le Treut, H., Morcrette, J. J., & Senior, C. (2004). On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, *22*(2-3), 71–86. doi: https://doi.org/10.1007/s00382-003-0369-6

Brooks, H. E., Lee, J. W., & Craven, J. P. (2003). The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmospheric Research*, *67*, 79–94. doi: https://doi.org/10.1016/S0169-8095(03)00045-0

Chavas, D. R., & Li, F. (2022). *Data: Bias in CMIP6 historical U.S. severe convective environments driven by bias in mean-state near-surface moist static energy.* Retrieved from `https://purr.purdue.edu/publications/3977/1` doi: 10.4231/42ZJ-A891

Chen, J., & Chavas, D. R. (2020). The transient responses of an axisymmetric tropical cyclone to instantaneous surface roughening and drying. *Journal of the Atmospheric Sciences*, *77*(8), 2807–2834. doi: https://doi.org/10.1175/JAS-D-19-0320.1

Chen, J., Dai, A., Zhang, Y., & Rasmussen, K. L. (2020). Changes in Convective Available Potential Energy and Convective Inhibition under global warming. *Journal of Climate*, *33*(6), 2025–2050. doi: https://doi.org/10.1175/JCLI-D-19-0461.1

Coffer, B. E., Taszarek, M., & Parker, M. D. (2020). Near-ground wind profiles of tornadic and nontornadic environments in the united states and europe from era5 reanalyses. *Weather and Forecasting*, *35*(6), 2621–2638.

Computational and Information Systems Laboratory. (2019). *Cheyenne: HPE/SGI ICE XA System (University Community Computing).* Boulder, CO: National Center for Atmospheric Research.

Emanuel, K. (1994). *Atmospheric convection.* Oxford University Press, USA. Retrieved from `http://books.google.com/books?id=VdaBBHEGAcMC`

Emori, S., & Brown, S. J. (2005). Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophysical Research Letters*, *32*(17), 1–5. doi: https://doi.org/10.1029/2005GL023272

European Centre for Medium-Range Weather Forecasts. (2019). *ERA5 Reanalysis (0.25 Degree Latitude-Longitude Grid).* Boulder CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. Retrieved from `https://doi.org/10.5065/BH6N-5N20`

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://gmd.copernicus.org/articles/9/1937/2016/` doi: 10.5194/gmd-9-1937-2016

Gensini, V. A., & Ashley, W. S. (2011). Climatology of potentially severe convective environments from the north american regional reanalysis. *E-Journal of Severe Storms Meteorology*, *6*(8).

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … others (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal

Meteorological Society, *146* (730), 1999–2049.

Hoogewind, K. A., Baldwin, M. E., & Trapp, R. J. (2017). The impact of climate change on hazardous convective weather in the united states: insight from high-resolution dynamical downscaling. *Journal of Climate*, *30* (24), 10081–10100.

Johns, R. H., & Doswell III, C. A. (1992). Severe local storms forecasting. *Weather and Forecasting*, *7* (4), 588–612. doi: https://doi.org/10.1175/1520-0434(1992)007⟨0588:SLSF⟩2.0.CO;2

Lepore, C., Abernathey, R., Henderson, N., Allen, J. T., & Tippett, M. K. (2021). Future global convective environments in cmip6 models. *Earth's Future*, *9* (12), e2021EF002277.

Li, F., & Chavas, D. R. (2021). Midlatitude continental cape is predictable from large-scale environmental parameters. *Geophysical Research Letters*, *48* (8), e2020GL091799.

Li, F., Chavas, D. R., Reed, K. A., & Dawson II, D. T. (2020). Climatology of severe local storm environments and synoptic-scale features over north america in era5 reanalysis and cam6 simulation. *Journal of Climate*, *33* (19), 8339–8365.

Li, F., Chavas, D. R., Reed, K. A., Rosenbloom, N., & Dawson II, D. T. (2021). The role of elevated terrain and the gulf of mexico in the production of severe local storm environments over north america. *Journal of Climate*, *34* (19), 7799–7819.

Long, J. A., Stoy, P. C., & Gerken, T. (2018). Tornado seasonality in the southeastern united states. *Weather and climate extremes*, *20*, 81–91.

Ludlam, F. (1963). Severe local storms: A review. In *Severe local storms. meteorological monographs, vol 5.* (pp. 1–32). American Meteorological Society, Boston, MA. doi: https://doi.org/10.1007/978-1-940033-56-3_1

McCartney, G., Hacker, T., & Yang, B. (2014). Empowering Faculty: A Campus Cyberinfrastructure Strategy for Research Communities. *Educause Review*. Retrieved from `https://er.educause.edu/articles/2014/7/empowering-faculty-a-campus-cyberinfrastructure-strategy-for-research-communities`

Rasmussen, E. N., & Blanchard, D. O. (1998). A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Weather and Forecasting*, *13* (4), 1148–1164. doi: https://doi.org/10.1175/1520-0434(1998)013⟨1148:abcosd⟩2.0.co;2

Seeley, J. T., & Romps, D. M. (2015). The effect of global warming on severe thunderstorms in the united states. *Journal of Climate*, *28* (6), 2443–2458.

Taszarek, M., Allen, J. T., Brooks, H. E., Pilguj, N., & Czernecki, B. (2021). Differing trends in united states and european severe thunderstorm environments in a warming climate. *Bulletin of the American Meteorological society*, *102* (2), E296–E322.

Taszarek, M., Allen, J. T., Púčik, T., Hoogewind, K. A., & Brooks, H. E. (2020). Severe convective storms across europe and the united states. part ii: Era5 environments associated with lightning, large hail, severe wind, and tornadoes. *Journal of Climate*, *33* (23), 10263–10286.

WMO/OMM/BMO. (1992). *International meteorological vocabulary* (Tech. Rep. No. 182). Geneva, Switzerland: Secretariat of the World Meteorological Organization.

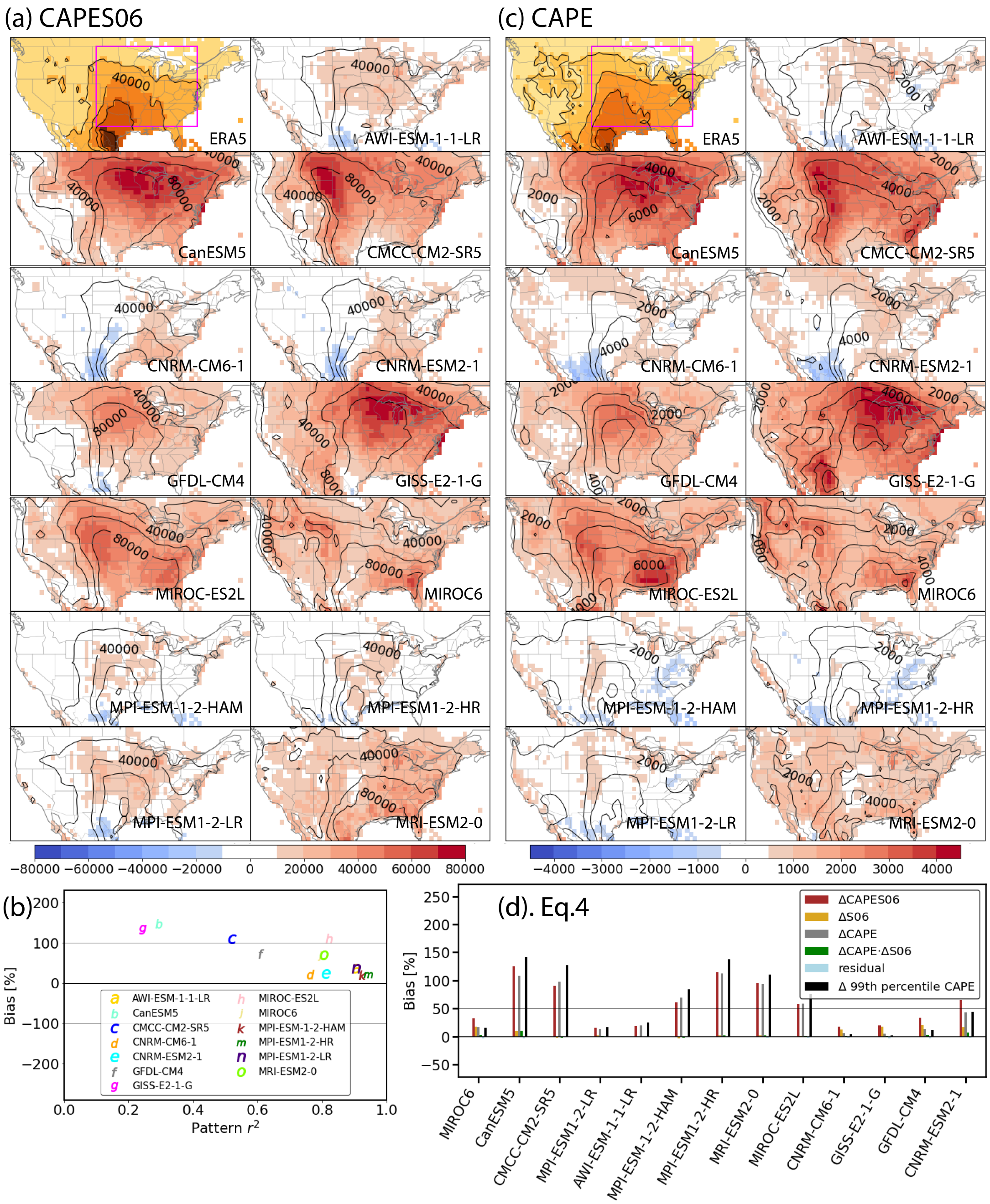


Figure 1. (a) Comparison of Spring (MAM) CMIP6 historical model simulations against ERA5 in reproducing the severe weather environment climatology, defined as the 99th percentile CAPE*S06. Top-left panel: ERA5 distribution. Other panels: 13 CMIP6 model differences from ERA5 (color) and absolute values (contour). Pink box indicates land region used for all subsequent analyses and mean bias calculations. (b) Pattern r^2 and mean bias ([%]; Eq. 3). (c) as (a) but for 99th percentile CAPE only. (d) Mean bias of extreme CAPES06 over the central and eastern U.S. across CMIP6 models relative to ERA5 (red bar), and the conditional bias contribution from CAPE (grey bar) and S06 (yellow bar) given by Eq. 4. Black bar represents mean unconditional bias in the 99th percentile of CAPE. Period is 1980–2014. All calculations in (b) and (d) are from land gridpoints in the pink box whose Spring 99th percentile CAPE exceeds 150 J kg^{-1} (see text for details). Results for Summer (JJA), Fall (SON), and Winter (DJF) shown in Supplementary Figures S1–S3.

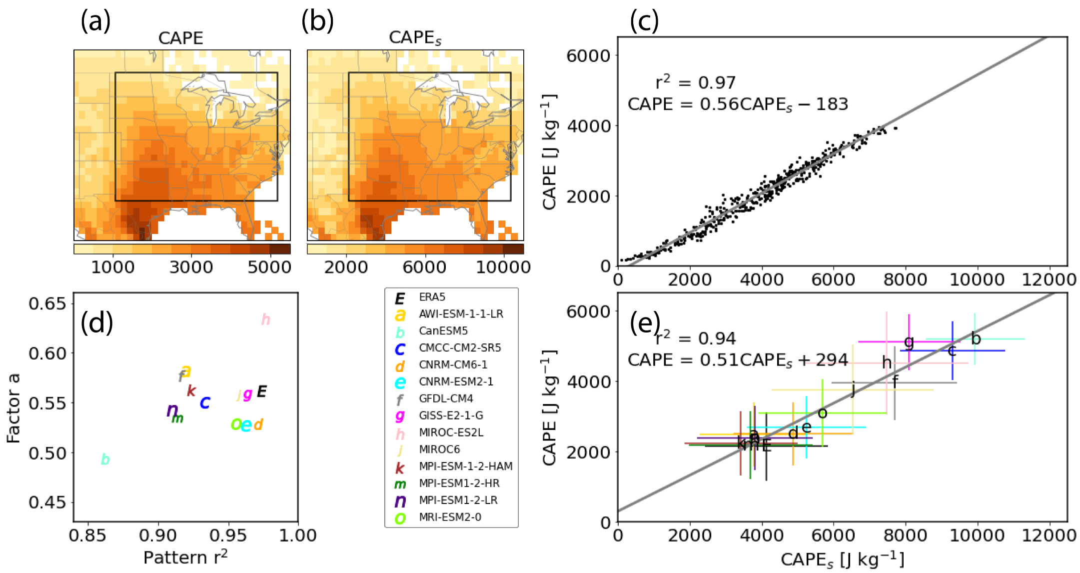


Figure 2. Comparison of extreme CAPE against extreme CAPE$_s$ in Spring. ERA5: (a) 99th percentiles of CAPE; (b) as in (a) but for CAPE$_s$; (c) CAPE$_s$ vs. CAPE with linear regression fit (solid line). CMIP6 model vs. ERA5: (d) scatter plot of pattern r^2 between extreme CAPE and CAPE$_s$ (x-axis) and linear regression slope (y-axis; a in Eq. 6) within ERA5 ('E') and within each model; (e) spatial-mean 99th percentile CAPE vs. CAPE$_s$ over the central and eastern U.S. within ERA5 ('E') and within each model, with error bars indicating one standard deviation. Results for Summer (JJA), Fall (SON), and Winter (DJF) shown in Supplementary Figures S4–S6.

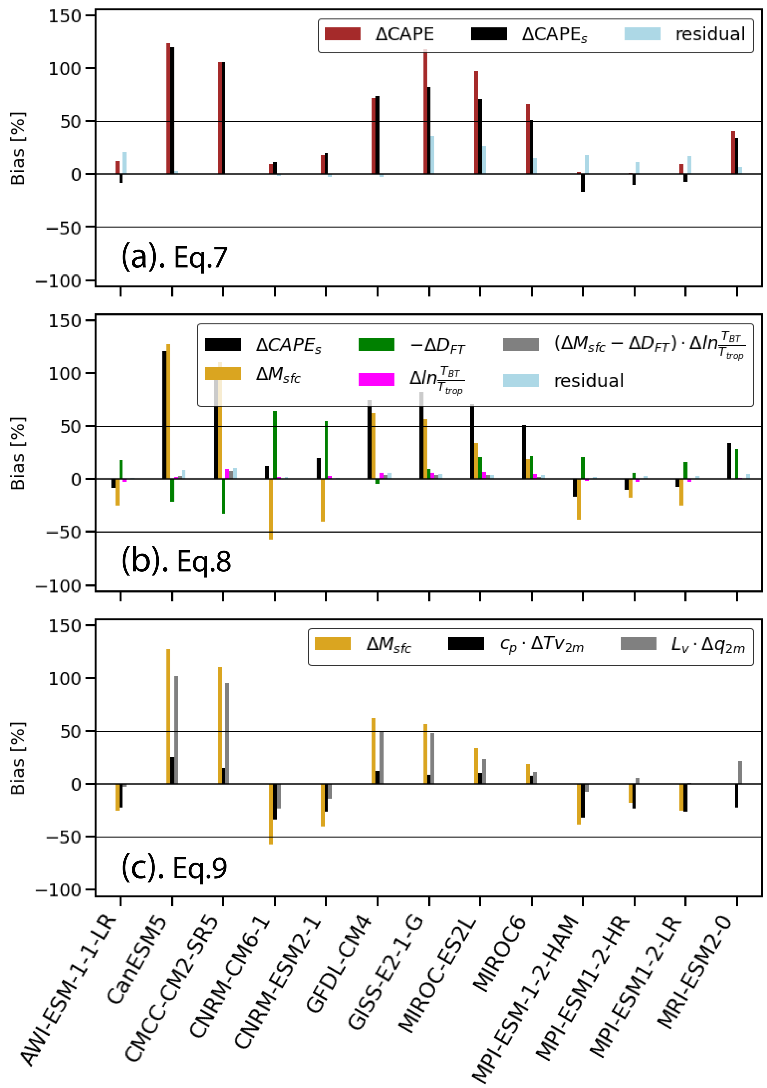


Figure 3. (a) Mean bias over the central and eastern U.S. in the median of top 1% cases of CAPE (red bar) vs. CAPE$_s$ (black bar) for each model in Spring (Eq. 7). (b) Decomposing bias in CAPE$_s$ (top 1%) into conditional bias contributions from M_{sfc} (yellow bar), D_{FT} (green bar), and other terms (Eq. 8). (c) Linearly decomposing conditional bias in M$_{sfc}$ into bias contributions from sensible heat (black bar) and latent heat (grey bar). Results for Summer (JJA), Fall (SON), and Winter (DJF) shown in Supplementary Figures S7–S9.

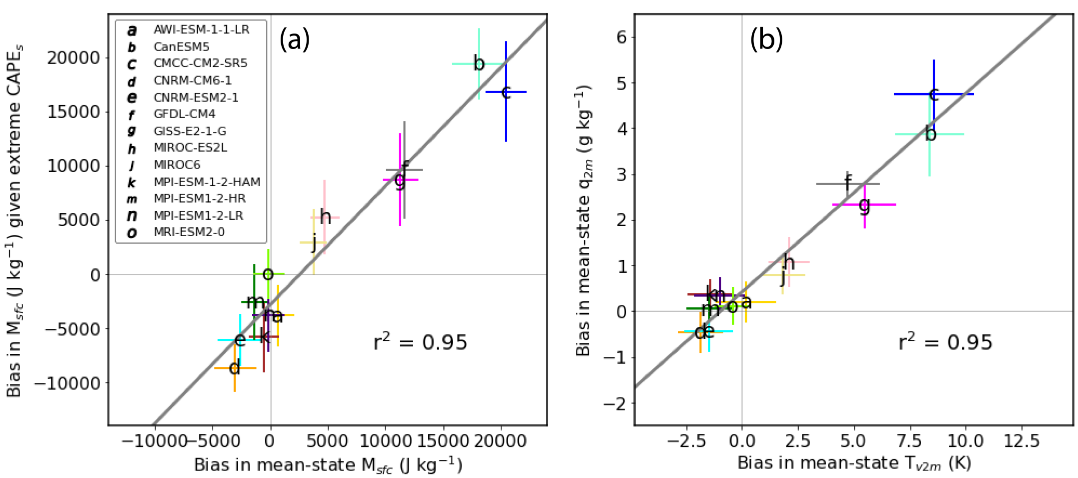


Figure 4. (a) Mean bias in M_{sfc} from the top 1% of $CAPE_s$ environments (Fig. 3) vs. mean bias in mean-state M_{sfc}. (b) Mean-state biases in q_{2m} vs. T_{v2m}. Solid line indicates linear regression fit and cross bars indicate one standard deviation in each quantity. Results for Summer (JJA), Fall (SON), and Winter (DJF) shown in Supplementary Figures S10–S12.

Figure 1.

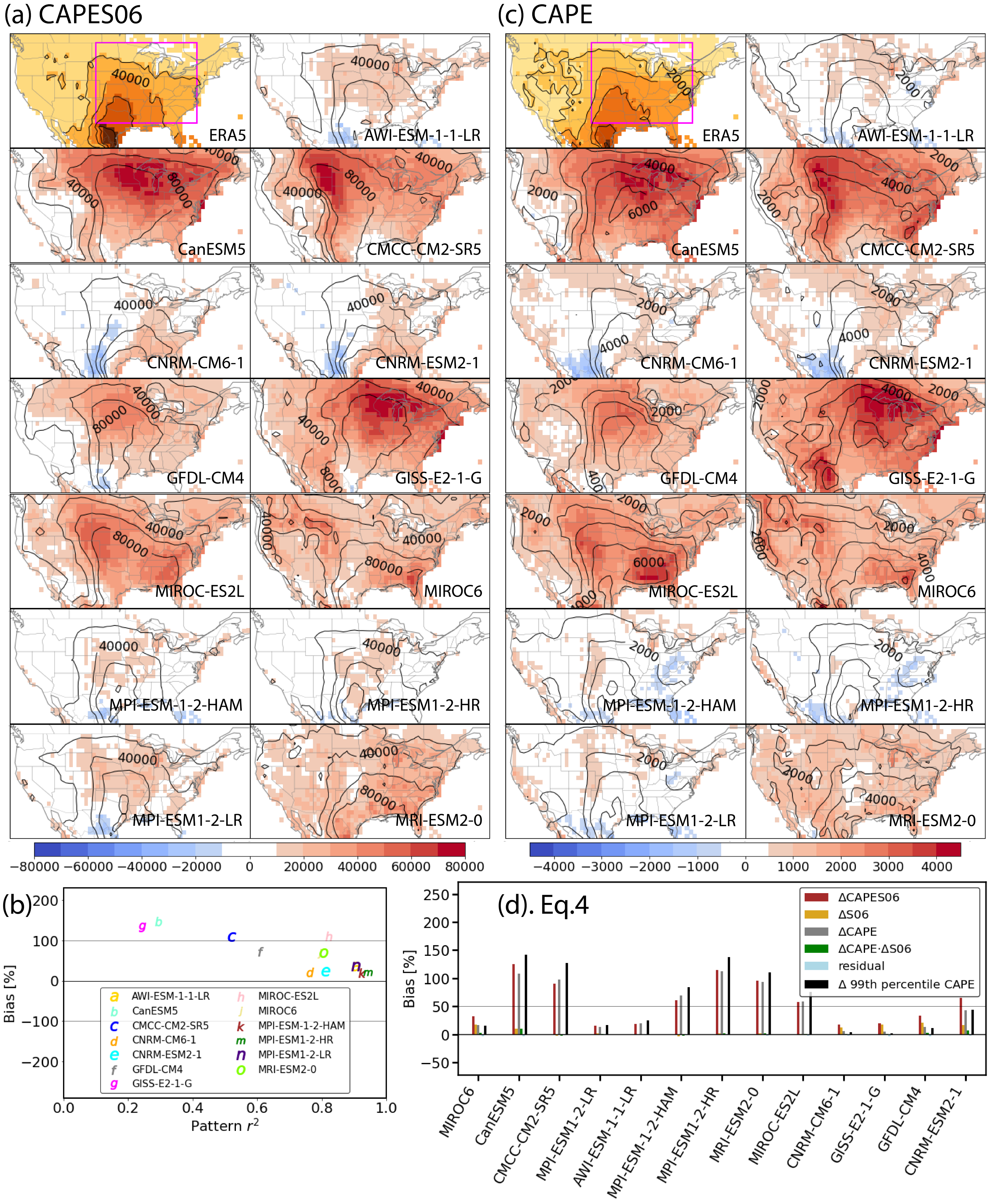
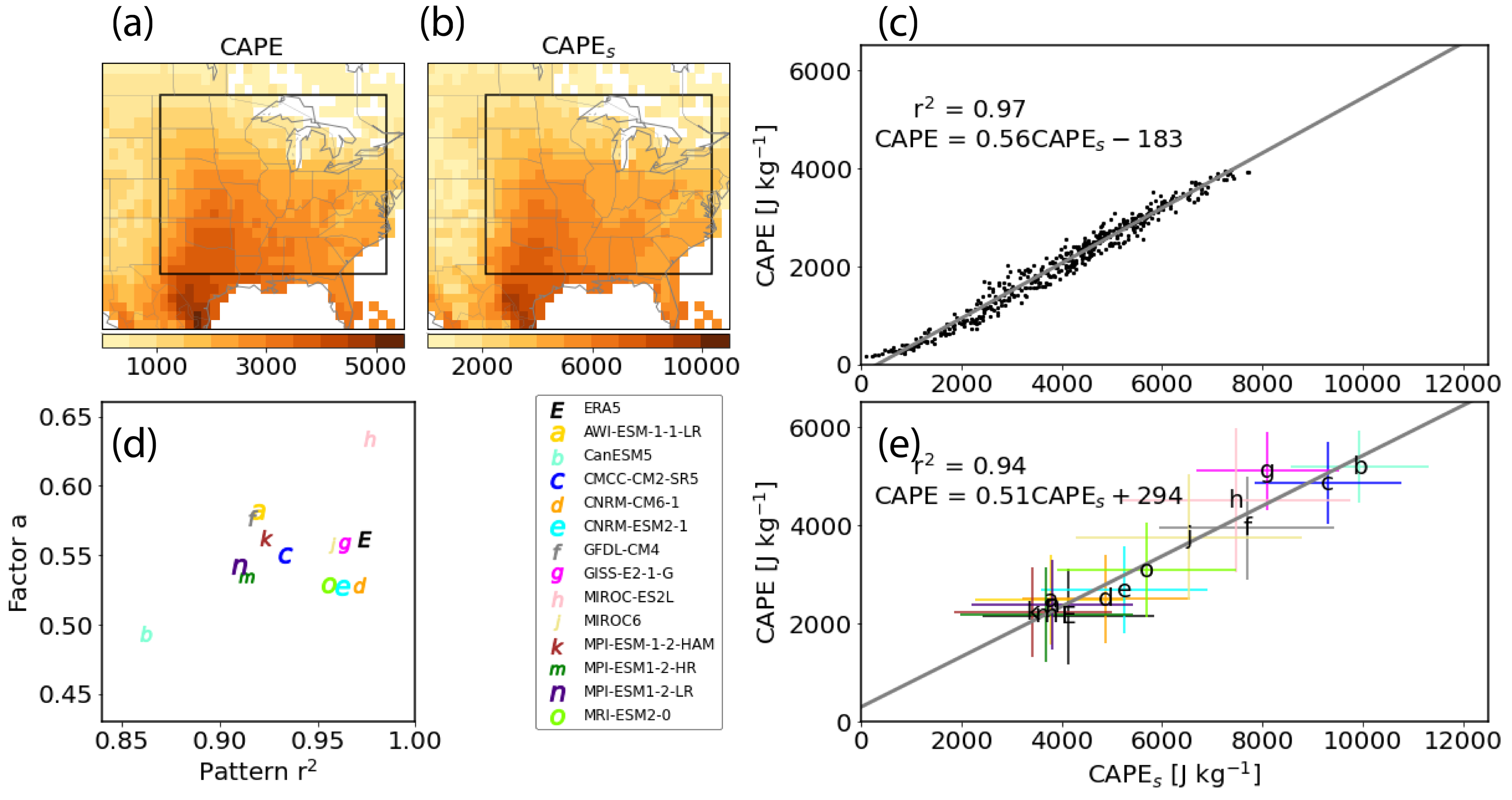
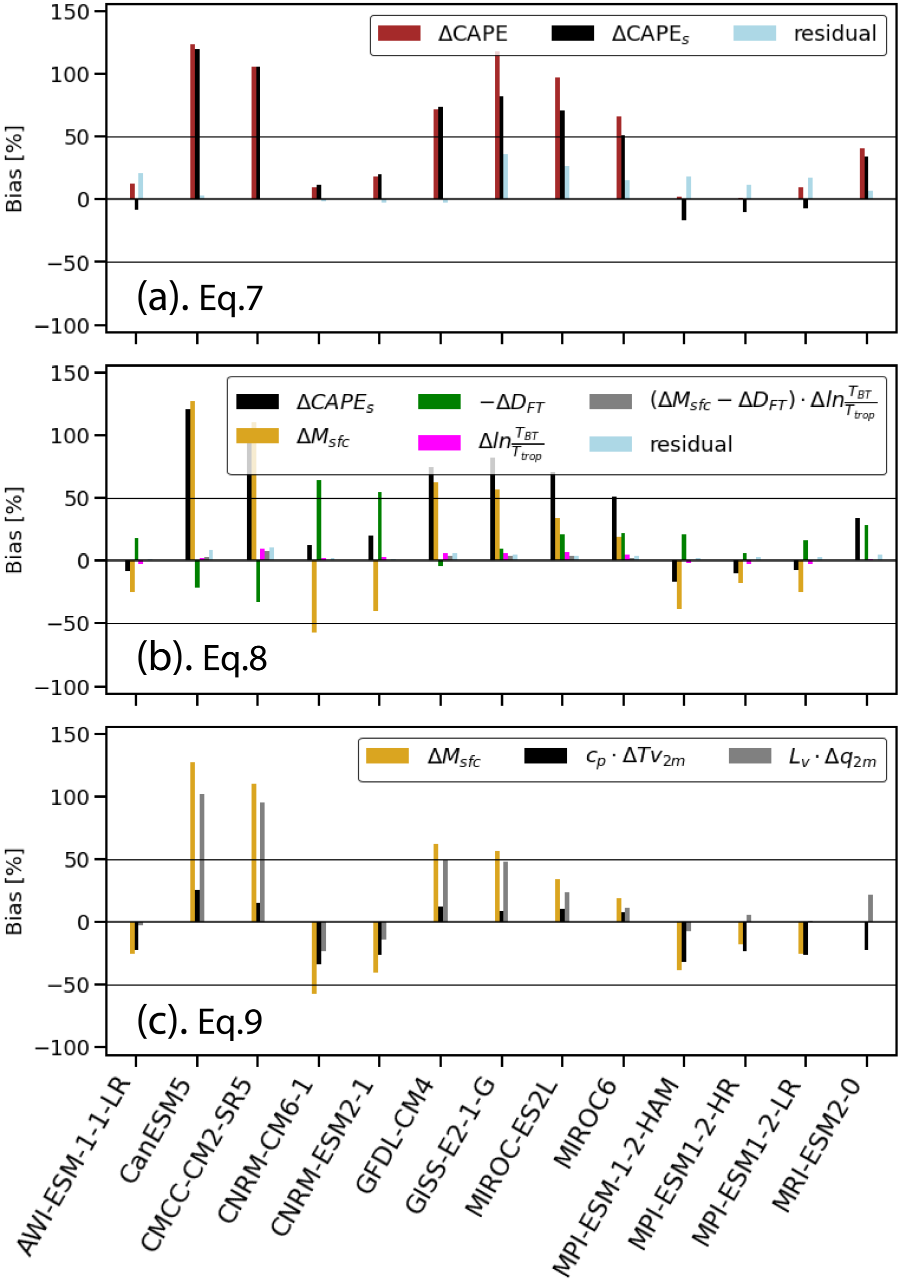

(a) CAPES06
(c) CAPE
ERA5
AWI-ESM-1-1-LR
CanESM5
CMCC-CM2-SR5
CNRM-CM6-1
CNRM-ESM2-1
GFDL-CM4
GISS-E2-1-G
MIROC-ES2L
MIROC6
MPI-ESM-1-2-HAM
MPI-ESM1-2-HR
MPI-ESM1-2-LR
MRI-ESM2-0
−80000 −60000 −40000 −20000 0 20000 40000 60000 80000
−4000 −3000 −2000 −1000 0 1000 2000 3000 4000
(b)
Bias [%]
Pattern r^2
a AWI-ESM-1-1-LR
b CanESM5
c CMCC-CM2-SR5
d CNRM-CM6-1
e CNRM-ESM2-1
f GFDL-CM4
g GISS-E2-1-G
h MIROC-ES2L
j MIROC6
k MPI-ESM-1-2-HAM
m MPI-ESM1-2-HR
n MPI-ESM1-2-LR
o MRI-ESM2-0
(d). Eq.4
Bias [%]
ΔCAPES06
ΔS06
ΔCAPE
ΔCAPE·ΔS06
residual
Δ 99th percentile CAPE
MIROC6
CanESM5
CMCC-CM2-SR5
MPI-ESM1-2-LR
AWI-ESM-1-1-LR
MPI-ESM-1-2-HAM
MPI-ESM1-2-HR
MRI-ESM2-0
MIROC-ES2L
CNRM-CM6-1
GISS-E2-1-G
GFDL-CM4
CNRM-ESM2-1

Figure 2.



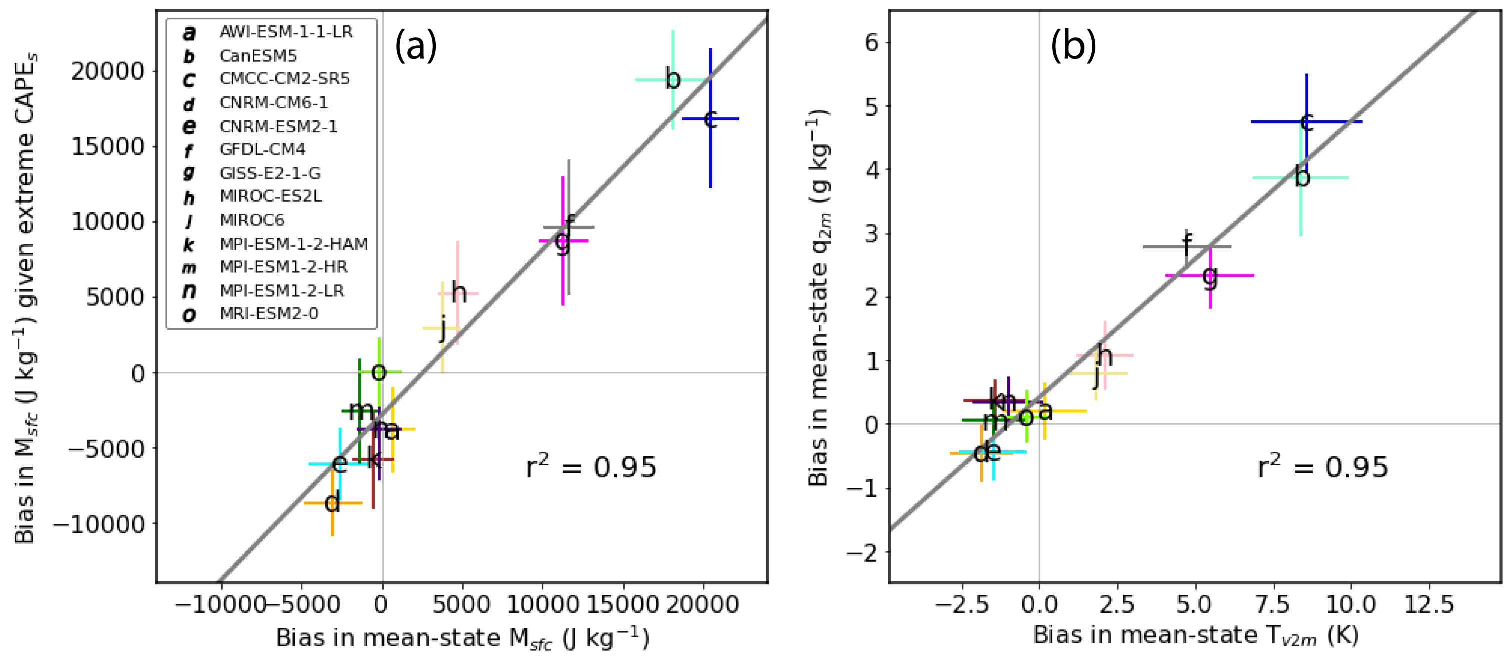
(a)
CAPE
1000
3000
5000
(b)
CAPE$_s$
2000
6000
10000
(c)
$r^2 = 0.97$
CAPE = 0.56CAPE$_s$ − 183
CAPE [J kg^{-1}]
(d)
Factor a
Pattern r^2
E ERA5
a AWI-ESM-1-1-LR
b CanESM5
c CMCC-CM2-SR5
d CNRM-CM6-1
e CNRM-ESM2-1
f GFDL-CM4
g GISS-E2-1-G
h MIROC-ES2L
j MIROC6
k MPI-ESM-1-2-HAM
m MPI-ESM1-2-HR
n MPI-ESM1-2-LR
o MRI-ESM2-0
(e)
$r^2 = 0.94$
CAPE = 0.51CAPE$_s$ + 294
CAPE [J kg^{-1}]
CAPE$_s$ [J kg^{-1}]

Figure 3.



ΔCAPE
ΔCAPE$_s$
residual
(a). Eq.7
ΔCAPE$_s$
$-\Delta D_{FT}$
$(\Delta M_{sfc} - \Delta D_{FT}) \cdot \Delta ln\frac{T_{BT}}{T_{trop}}$
ΔM_{sfc}
$\Delta ln\frac{T_{BT}}{T_{trop}}$
residual
(b). Eq.8
ΔM_{sfc}
$c_p \cdot \Delta Tv_{2m}$
$L_v \cdot \Delta q_{2m}$
(c). Eq.9
Bias [%]
150
100
50
0
−50
−100
AWI-ESM-1-1-LR
CanESM5
CMCC-CM2-SR5
CNRM-CM6-1
CNRM-ESM2-1
GFDL-CM4
GISS-E2-1-G
MIROC-ES2L
MIROC6
MPI-ESM-1-2-HAM
MPI-ESM1-2-HR
MPI-ESM1-2-LR
MRI-ESM2-0

Figure 4.



(a)
(b)
a AWI-ESM-1-1-LR
b CanESM5
c CMCC-CM2-SR5
d CNRM-CM6-1
e CNRM-ESM2-1
f GFDL-CM4
g GISS-E2-1-G
h MIROC-ES2L
j MIROC6
k MPI-ESM-1-2-HAM
m MPI-ESM1-2-HR
n MPI-ESM1-2-LR
o MRI-ESM2-0
$r^2 = 0.95$
$r^2 = 0.95$
Bias in M_{sfc} (J kg^{-1}) given extreme $CAPE_s$
Bias in mean-state M_{sfc} (J kg^{-1})
Bias in mean-state q_{2m} (g kg^{-1})
Bias in mean-state T_{v2m} (K)