

Reanalyses and a High-Resolution Model Fail to Capture the “High Tail” of CAPE Distributions

ZIWEI WANG,^{a,b} JAMES A. FRANKE,^{a,b} ZHENQI LUO,^c AND ELISABETH J. MOYER^{a,b}

^a *Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois*

^b *Center for Robust Decision-making on Climate and Energy Policy, University of Chicago, Chicago, Illinois*

^c *College of Natural Resources, Faculty of Geographical Science, Beijing Normal University, Beijing, China*

(Manuscript received 15 December 2020, in final form 9 August 2021)

ABSTRACT: Convective available potential energy (CAPE) is of strong interest in climate modeling because of its role in both severe weather and in model construction. Extreme levels of CAPE ($>2000 \text{ J kg}^{-1}$) are associated with high-impact weather events, and CAPE is widely used in convective parameterizations to help determine the strength and timing of convection. However, to date few studies have systematically evaluated CAPE biases in models in a climatological context, and none have addressed bias in the high tail of CAPE distributions. This work compares CAPE distributions in $\sim 200\,000$ summertime proximity soundings from four sources: the observational radiosonde network [Integrated Global Radiosonde Archive (IGRA)], 0.125° reanalyses (ERA-Interim and ERA5), and a 4-km convection-permitting regional WRF simulation driven by ERA-Interim. Both reanalyses and the WRF Model consistently show too-narrow distributions of CAPE, with the high tail (>90 th percentile) systematically biased low by up to 10% in surface-based CAPE and even more in alternate CAPE definitions. This “missing tail” corresponds to the most impacts-relevant conditions. CAPE bias in all datasets is driven by surface temperature and humidity: reanalyses and the WRF Model underpredict observed cases of extreme heat and moisture. These results suggest that reducing inaccuracies in land surface and boundary layer models is critical for accurately reproducing CAPE.

KEYWORDS: CAPE; Radiosonde/rawinsonde observations; Cloud resolving models; Model evaluation/performance; Reanalysis data

1. Introduction

Convective available potential energy (CAPE) is an integral quantity of buoyancy in the convective layer (Moncrieff and Miller 1976) and is considered as a key parameter in convection initiation and development. Closely linked to updraft strength and storm intensity, CAPE provides a way to understand the potential threat of some high-impact weather events such as thunderstorms, hail, and tornadoes. Brooks et al. (2003) propose a combination of CAPE and bulk wind shear as a metric for severe weather in reanalyses, with a 2000 J kg^{-1} as a threshold value for extreme events, and multiple subsequent studies confirm this relationship in models and observations. Studies relating high CAPE values to extreme precipitation or intense storms in observations include Groenemeijer and van Delden (2007), Lepore et al. (2015), Dong et al. (2019), and many others. In models, Paquin et al. (2014), for example, show that the number of extreme precipitation events in general circulation models (GCMs) grows with the covariate between CAPE and wind shear.

CAPE is also used as a key parameter in convective schemes in GCMs to determine convective mass flux (Zhang and McFarlane 1995; Yano et al. 2013; Baba 2019). In CAPE-relaxing closure (CR closure) schemes, modelers commonly

rely on CAPE to trigger convection and to determine the total vertical mass flux, so that the magnitude of vertical mass flux is directly affected by an inaccurate representation of CAPE (Lee et al. 2008; Cortés-Hernández et al. 2016). In some recently developed new schemes intended to more realistically reproduce the diurnal cycle, convective triggering is directly dependent on CAPE generation rate (dCAPE) (Xie and Zhang 2000; Wang et al. 2015). These schemes have been shown to improve model performance for precipitation diurnal peak time compared to schemes using classic CR closure (Song and Zhang 2017; Xie et al. 2019). However, it should be noted that these dynamical-based trigger functions introduce additional sensitivity to CAPE biases, and Song and Zhang (2018) find that dCAPE trigger function are highly sensitive to model resolution.

CAPE is derived from vertical profiles of temperature, pressure, and humidity, which are measured in situ only from a sparse network of specialized weather stations. Radiosondes measure atmospheric profiles from weather balloons released twice a day from ~ 1000 stations globally. Because radiosonde measurements are both spatially and temporally sparse, researchers linking measured CAPE to severe weather events have used “proximity soundings,” estimating the severity of extreme weather events based on soundings taken within a range of $\sim 200 \text{ km}$ (e.g., Brooks et al. 1994; Rasmussen and Blanchard 1998; Brooks and Craven 2002). More recent studies of CAPE and severe weather use not soundings but reanalyses that assimilate in situ and remote observations in global models to provide information at higher resolution (Brooks et al. 2003; Lepore et al. 2015; Dong et al. 2019). Global gridded reanalyses also allow ready construction of climatologies: for example,

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-20-0278.s1>.

Corresponding author: Elisabeth J. Moyer, moyer@uchicago.edu

Riemann-Campe et al. (2009) use the ERA40 reanalysis to construct a 40-yr climatology of CAPE, showing that largest values and variability are found over tropical land (mean $\sim 2000 \text{ J kg}^{-1}$), with a stronger dependence on specific humidity than temperature.

To diagnose potential changes in CAPE under future higher CO_2 conditions, studies must rely on numerical simulations. With the growth of computational resources, the horizontal resolution of models used for this purpose has increased. For example, Trapp et al. (2009) and Diffenbaugh et al. (2013) examine changes in CAPE and wind shear in GCM projections ($\sim 100 \text{ km}$) and infer a likely future increase in the number of days with severe weather events. Singh et al. (2017) use both GCMs and superparameterized GCMs (20 km) to study changes in the 95th percentile of CAPE in the tropics and subtropics during heavy precipitation and find a 6%–14% increase per kelvin regional temperature increase. [Note that CAPE values during heavy precipitation are low, e.g., Adams and Souza (2009); the 95th percentile in observations in Singh et al. (2017) is under 2000 J kg^{-1} .] Rasmussen et al. (2017) examine changes in CAPE and convective inhibition (CIN) in a 4-km dynamically downscaled simulation of North America in a pseudo-global warming scenario (driven by reanalysis or by reanalysis with an applied offset in climate variables). They find that both CAPE and CIN generally increase under warmer conditions and infer a future intensification of convective strength. Such convection-permitting models, with their improvement in convective dynamics, have been assumed to help improve the representation of CAPE.

Given the extent of the scientific use of reanalyses and model simulations, it is valuable to ask how well these products reproduce realistic CAPE values. Coarse-resolution general circulation models reproduce large-scale spatial patterns in CAPE but can produce large biases in individual locations (Chen et al. 2020). While reanalyses and high-resolution forecast models are generally assumed to be more accurate, assessments of their biases versus radiosonde observations have been limited. Studies using restricted samples of soundings near severe weather events have produced inconsistent results. For example, Thompson et al. (2003) evaluate surface-based CAPE (SBCAPE) from the Rapid Update Cycle (RUC-2) weather prediction system 0-h analysis against radiosondes sampled near supercells (149 soundings from 1999 to 2001, in the U.S. central and southern plains) and find a low bias of $\sim 16\%$ (mean bias of about -400 J kg^{-1} in mean conditions of $\sim 2500 \text{ J kg}^{-1}$). Coniglio (2012) compare SBCAPE in the RUC 0-h analysis with a different sample of soundings near supercell thunderstorms (582 soundings during the VORTEX2 campaign in 2009–10, also in the central and southern plains) and find a small high bias ($\sim 150 \text{ J kg}^{-1}$) with a large spread. Allen and Karoly (2014) compare mixed-layer CAPE (MLCAPE) in the reanalysis product ERA-Interim (ERA-I) and in the Australian Mesoscale Limited Area Prediction System (MesoLAPS) weather model with radiosonde soundings near thunderstorm events (3697 and 4988 soundings, respectively, from 2003 to 2010, from 16 stations in Australia) and find slight high biases of 6 and 74 J kg^{-1} in conditions of 234 and 255 J kg^{-1} mean nonzero MLCAPE.

To date, very few validation studies have systematically evaluated CAPE bias and errors in a climatological context, with a large enough scale to allow evaluation of the high tail of the CAPE distribution. For convection-permitting models, it is widely assumed that improved resolution also improves the representation of CAPE, but this assumption has not been explicitly tested. For reanalyses, only a few studies have compared output to large collections of soundings, and none assess distributional changes. Gensini et al. (2014) compare the North American Regional Reanalysis (NARR) to all radiosondes over 11 years from 21 stations in the eastern United States ($>100\,000$ soundings with nonzero SBCAPE from 2000 to 2011), but do not assess either mean bias or distributional differences. (They do find considerable spread in SBCAPE errors, with RMSE $\sim 1400 \text{ J kg}^{-1}$.) Taszarek et al. (2018) and Taszarek et al. (2020) use even larger sample sizes (>1 million profiles from 1979 to 2016, and >5 million profiles from 1980 to 2018, respectively, predominantly over Europe and spanning all seasons) and compare CAPE under various definitions between soundings and ERA-I and ERA5 reanalyses. They examine mean biases (in soundings with nonzero CAPE) and find them large relative to median CAPE, but median values in both studies are small (~ 68 and 100 J kg^{-1} for MLCAPE).

Even fewer studies have attempted to attribute bias in CAPE to specific model issues. While one possible cause is error in free tropospheric profiles, multiple authors have noted the potential role of incorrect temperature and humidity at the surface or boundary layer. Several studies have explicitly tested this attribution by replacing surface values in models and data products with observed ones and noting the improved match to radiosonde SBCAPE. Coniglio (2012) replaces surface values in RUC with those from the operational surface objective analysis system (SFCOA) and finds a reduction in bias in 1-h forecasts. Gartzke et al. (2017) compare 10 years of SBCAPE from a single station, the Southern Great Plains Atmospheric Radiation Measurement (ARM) site, and show that replacing surface values largely corrects CAPE values in ERA-I reanalysis and values derived from the AIRS satellite. Similarly, in a very small sample (two individual case studies), Bloch et al. (2019) find that replacing surface values of humidity and temperature corrects a low bias in SBCAPE in a satellite-derived product.

This work seeks to address both needs, for large-scale systematic assessment of CAPE distributions in reanalyses and high-resolution simulations against those in radiosondes, and for attribution of the source of any bias. To allow focusing on conditions that promote strong convection, we examine summertime data over the contiguous United States, using a total of nearly 200 000 soundings over 12 years of observations.

2. Data description

This study compares four datasets that allow calculation of CAPE over the contiguous United States from January 2001 to December 2012: radiosonde observations from the Integrated Global Radiosonde Archive (IGRA), version 2 (Durre et al. 2006, 2008); the reanalysis products ERA-I and ERA5 (Dee et al. 2011; Hersbach et al. 2020); and simulation

output from the Weather Research and Forecasting (WRF) Model at the convection-permitting resolution, forced by ERAI (Rasmussen and Liu 2017). Because our interest is in the high tail of the CAPE distribution, we focus on the summer months when convection is most active and CAPE is largest. We define summer as May–August (MJJ), following the convention of many studies (e.g., Sun et al. 2016; Rasmussen et al. 2017), though some work on extreme weather uses an earlier definition of April–July to include the late spring peak of convection (e.g., Trapp et al. 2009). With this definition, IGRA provides a total of 199 787 summertime radiosonde profiles from U.S. stations with continuous records during 2001–12. For consistency, analyses shown here involve data matched to radiosonde stations and synchronized in time, though when evaluating diurnal cycles, we also show reanalysis and model output at additional times of the day.

a. Radiosonde observations

IGRA is an archive of quality-controlled atmospheric sounding profiles from weather balloons around the world collected by a standard protocol. The archive is operated by the U.S. National Oceanic and Atmospheric Administration (NOAA) and profiles in the United States are collected by NOAA's National Weather Service. In this work we use profiles from all stations in the contiguous United States that report continuous operation through the years 2001–12, a total of 80 out of the 248 stations historically used. All stations have routine balloon launches at 0000 and 1200 UTC each day, though some soundings are missing (17.4% of all routine launches during this period). Many stations also include sporadic launches at 0600 and 1800 UTC; we include these profiles in the dataset considered here, though we generally disaggregate analyses by the time of day. Of the complete dataset of 199 787 soundings, 83 668 are from 0000 UTC, 106 455 are from 1200 UTC, and 9664 are from additional times. All four hours are used in our analysis unless otherwise stated. Of these profiles, 1496 (0.75%) are excluded by our quality control criteria (see the methods section below).

Variables acquired from IGRA include pressure, temperature, altitude, and vapor pressure, all of which are standard reported values. We convert vapor pressure to specific humidity and dewpoint temperature for consistency across all datasets. Vertical resolution varies by station, but most stations report around 80 levels from the surface to 10-hPa pressure. (The data are available from <https://www.ncdc.noaa.gov/data-access/weather-balloon/integrated-global-radiosonde-archive>).

b. Reanalysis products

ERA-Interim and ERA5 are both reanalysis products maintained by the European Centre for Medium-Range Weather Forecasts (ECMWF). Both products assimilate observations into global models and are available from 1979 to the present. ERA-Interim has a native horizontal resolution of T255 (≈ 80 km); it has been superseded by ERA5, which has significant improvements in spatial and temporal resolution with a native horizontal resolution of TL639 (0.28° , ≈ 31 km) (Copernicus Climate Change Service 2017). Because our analysis involves matching individual

radiosonde stations, we acquire both reanalyses at a finer spatial resolution (0.125°) produced by ECMWF with bilinear interpolation for continuous fields. We use output at native model vertical levels, preserving the highest possible vertical resolution for our CAPE calculation: 60 levels for ERA-Interim (L60), and 137 for ERA5 (L137). We download profiles of temperature and specific humidity, and surface pressure; the pressure profile is then derived using surface pressure and coefficients a and b that define the hybrid-sigma coordinates of L60 and L137. Two-meter temperature and dewpoint temperature along with surface pressure are appended to the bottom level of profiles. Although ERA5 provides hourly output, we use data at 0000, 0600, 1200, and 1800 UTC to match with ERA-Interim. (Both products are available at <https://www.ecmwf.int/en/>).

Data assimilation is a key component of reanalysis products. Both ERA-Interim and ERA5 assimilate a homogenized version of IGRA radiosonde observations, the Radiosonde Observation Correction using Reanalyses (RAOBCORE) (Haimberger 2007; Haimberger et al. 2008). Reanalyses and IGRA observations are therefore not fully independent. ERA-Interim uses a bias correction for radiosonde temperature based on RAOBCORE_T_1.3, which is further adjusted and implemented to the Continuous Observation Processing Environment (COPE) framework in ERA5 (ECMWF 2016). The assimilation process of ERA-Interim uses the following exclusion criteria for radiosonde data: 1) any radiosonde observation below the model surface, and radiosonde-observed specific humidity in either 2) extreme cold conditions ($T < 193$ K for RS-90 sondes, $T < 213$ K for RS-80 sondes, $T < 233$ K otherwise), or 3) high altitude ($p < 100$ hPa for RS-80 and RS-90 sondes, $p < 300$ hPa for all other sonde types) (Dee et al. 2011).

c. High-resolution model simulation

The high-resolution model output we use is a 4-km resolution dynamically downscaled “retrospective” simulation over North America first described by Liu et al. (2017). The simulation is created as the control run of a pseudo-global warming experiment and involves forcing the WRF 3.4.1 Model with ERA-Interim reanalysis. The WRF simulation is run with 4-km grid spacing and 50 vertical levels up to 50 hPa, with parameterization schemes including: Thompson aerosol-aware microphysics (Thompson and Eidhammer 2014), the Yonsei University (YSU) planetary boundary layer (Hong et al. 2006), the Rapid Radiative Transfer Model (RRTMG) (Iacono et al. 2008), and the improved Noah-MP land surface model (Niu et al. 2011).

The model uses ERA-Interim as initial and boundary conditions, with large-scale spectral nudging applied to geopotential, temperature, and horizontal wind. Nudging is applied throughout the model domain, at all altitudes above the planetary boundary layer, and is intended to remove known large-scale issues (such as summertime high-temperature bias over the central United States; Morcrette et al. 2018) while still allowing smaller-scale processes to modify local profiles. Values are nudged at a strength corresponding to an “e-folding” time of 6 h, using a wavenumber truncation of 3 and 2 in the zonal and meridional directions, respectively. Because the experiment is intended to reproduce observed snow cover over North America, some modifications are made to the land surface model, including representing the heat transport from rainfall caused by the temperature

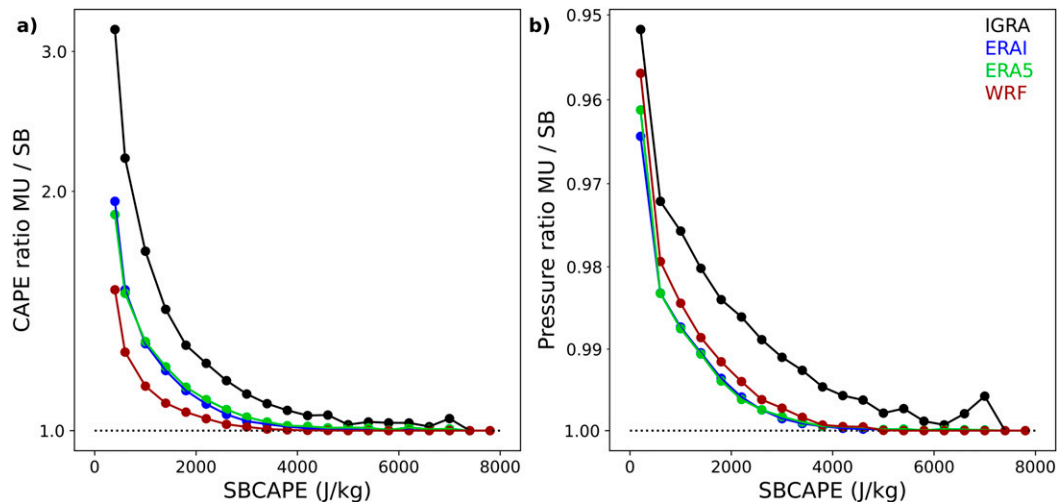


FIG. 1. Comparison of SBCAPE and MUCAPE for all datasets, using all soundings considered. Data are binned by SBCAPE value, and we exclude values under 200 J kg^{-1} . (a) Mean ratio of MUCAPE over SBCAPE and (b) mean of ratio of the most unstable pressure level over surface pressure. Note that y axes are log scale. For both CAPE and pressure level, the ratio approaches 1.0 as CAPE increases: in higher CAPE conditions, the most unstable level is closer to the surface.

difference between raindrops and land surface, and modifying the snow cover/melt curve to produce more realistic surface snow coverage and reduce wintertime low bias in temperature.

The WRF Model output is acquired from the NCAR Research Data Archive ds612.0 (Rasmussen and Liu 2017). We take the pressure, temperature, mixing ratio, height from the CTRL 3D subset, and surface topography, surface pressure, 2-m temperature, and mixing ratio from the CTRL 2D subset.

3. Methods

a. CAPE calculation

All CAPE values shown in this work are calculated with the Sounding and Hodograph Analysis and Research Program in Python (SHARPPy), version 1.4.0a4, a widely used collection of sounding and hodograph analysis routines designed to provide free and consistent analysis tools for the atmospheric sciences community (<https://github.com/sharppy/SHARPPy>; Blumberg et al. 2017). SHARPPy is an extension of SHARP, which was first released in 1991 (Hart and Korotky 1991). CAPE in the SHARPPy package is calculated following the definition of Moncrieff and Miller (1976) in which temperature is automatically corrected to virtual temperature (Doswell and Rasmussen 1994). The required variables are vertical profiles of pressure, temperature, height, and dewpoint temperature. Wind speed and direction are optional, and we do not include them. The package can produce the CAPE of parcels either at surface level (SBCAPE), at the “most unstable” level (MUCAPE), or using the averaged properties of “mixed layer” (MLCAPE). SHARPPy is the most commonly used package in the CAPE literature (e.g., Gartzke et al. 2017; King and Kennedy 2019), which provides a comprehensive list of convective indices as output.

We evaluate CAPE for all summertime profiles corresponding to radiosonde soundings other than those with the

following exclusion criteria: 1) no surface-level measurements (7 soundings or 0.004% of the total); 2) fewer than 20 vertical levels of observations (0.74% of soundings); or 3) excessive discrepancy of relative humidity between the surface and one level above, that is, $\text{RH}_{\text{sfc}} - \text{RH}_{\text{lev1}} > 65\%$ (16 soundings or 0.008%). An excessive RH gradient implies unphysical mixing; the exact threshold is somewhat arbitrary but is chosen to exclude outliers where $\text{CAPE} > 20\,000 \text{ J kg}^{-1}$. In some cases, radiosonde profiles involve missing values in the height variable, even though temperature, pressure, and humidity are reported. In these cases, we interpolate height based on pressure using the SHARPPy “INTERP” function.

b. Testing sensitivity to vertical interpolation

In the analysis here we interpolate only where data are missing in radiosonde profiles, using the SHARPPy “INTERP” function. The number of vertical levels used is therefore

TABLE 1. Fraction of observations of SBCAPE in each dataset that exceed threshold values or have zero value. Data used are the full 2001–12 MJJA dataset, inclusive of zeroes, with time/location matched to radiosonde observations. Parentheses show the ratio of incidences observed for each model or reanalysis relative to IGRA radiosondes; a number smaller than 1 means underestimation. Note the large deficits in the most extreme SBCAPE category ($>4000 \text{ J kg}^{-1}$), with the number of incidences underestimated by $\sim 40\%$ – 50% .

	IGRA	ERA1	ERA5	WRF
Zeroes	36.1%	38.1%	35.0%	39.1%
$>2000 \text{ J kg}^{-1}$	13.0%	12.8% (0.98)	13.8% (1.06)	13.2% (1.02)
$>3000 \text{ J kg}^{-1}$	5.4%	4.0% (0.74)	4.9% (0.91)	4.6% (0.85)
$>4000 \text{ J kg}^{-1}$	1.6%	0.8% (0.50)	1.0% (0.63)	1.0% (0.63)

TABLE 2. As in Table 1, but for MUCAPE. Deficits in the high tail are larger for MUCAPE than SBCAPE, as expected based on Fig. 1. Parentheses show the ratio of incidences observed for each model or reanalysis relative to IGRA radiosondes. The number of incidences of MUCAPE above the conventional severe-weather threshold (2000 J kg^{-1}) is underestimated by $\sim 25\%$ – 35% and that of extreme MUCAPE ($>4000 \text{ J kg}^{-1}$) by $\sim 65\%$ – 75% .

	IGRA	ERA1	ERA5	WRF
Zeroes	22.7%	30.3%	28.2%	32.8%
$>2000 \text{ J kg}^{-1}$	22.3%	16.3% (0.73)	17.5% (0.78)	14.8% (0.66)
$>3000 \text{ J kg}^{-1}$	10.9%	5.2% (0.48)	6.5% (0.60)	5.0% (0.46)
$>4000 \text{ J kg}^{-1}$	3.9%	1.0% (0.26)	1.3% (0.33)	1.0% (0.26)

inconsistent across datasets. Other authors of CAPE comparison studies have chosen to interpolate to produce consistent vertical sampling, for example, Gartzke et al. (2017), who use 202 fixed levels (2 and 30 m, followed by 75-m spacing from 75 m to 15 km). We test the robustness of derived CAPE to this

interpolation by considering mean errors in profiles binned by number of levels, using observations from the year 2012. We find that mean absolute errors introduced by interpolation are over 25 J kg^{-1} for profiles with <10 levels but fall to $\sim 5 \text{ J kg}^{-1}$ once the number of levels exceeds 60. However, since profiles with more missing levels are more likely to have low CAPE, the mean fractional error drops even more steeply (Fig. S1): 14% for <10 levels; 7% for 10–20 levels; and 0.8% for >20 levels, justifying our choice of 20 as a cutoff. (Note that the bias introduced by interpolation across these profiles is smaller, only -0.3% . See Coniglio 2012 for similar conclusions.)

c. CAPE definitions

CAPE is the potential buoyancy of a parcel lifted to its level of free convection, but the parcel considered may be located at the surface (SBCAPE), at the most unstable vertical level (MUCAPE), or may be a hypothetical parcel initiated using the mean state of the mixed layer (MLCAPE). All are standard outputs of SHARPPy, with the lowest

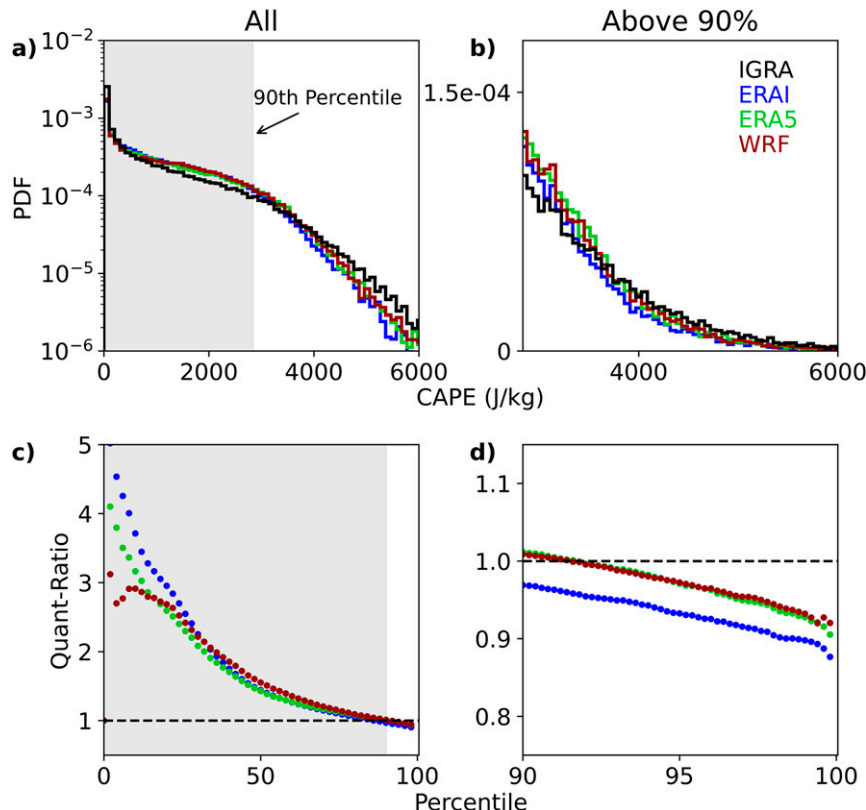


FIG. 2. (a),(b) Probability density functions and (c),(d) quantile ratio plots of CAPE from reanalyses (ERA1 and ERA5), high-resolution model output (WRF), and radiosonde observations (IGRA) for MJJA 2001–12, with times and locations matched to IGRA observations. Points with zero CAPE are excluded (36%–40% of datasets, see Table 1. (left) Full distribution and (right) the high tail (90th percentile and above). For IGRA, the 90th percentile is $\sim 2800 \text{ J kg}^{-1}$, the 95th $\sim 3200 \text{ J kg}^{-1}$, the 97.5th $\sim 4000 \text{ J kg}^{-1}$. In PDFs [in (a) and (b)], plots are cut off at 6000 J kg^{-1} on the x axis, omitting less than 0.1% of all points. In quantile ratio plots [in (c) and (d)], a slope downward to the right indicates a narrower distribution. The WRF Model and reanalyses consistently underpredict CAPE values in this high tail.

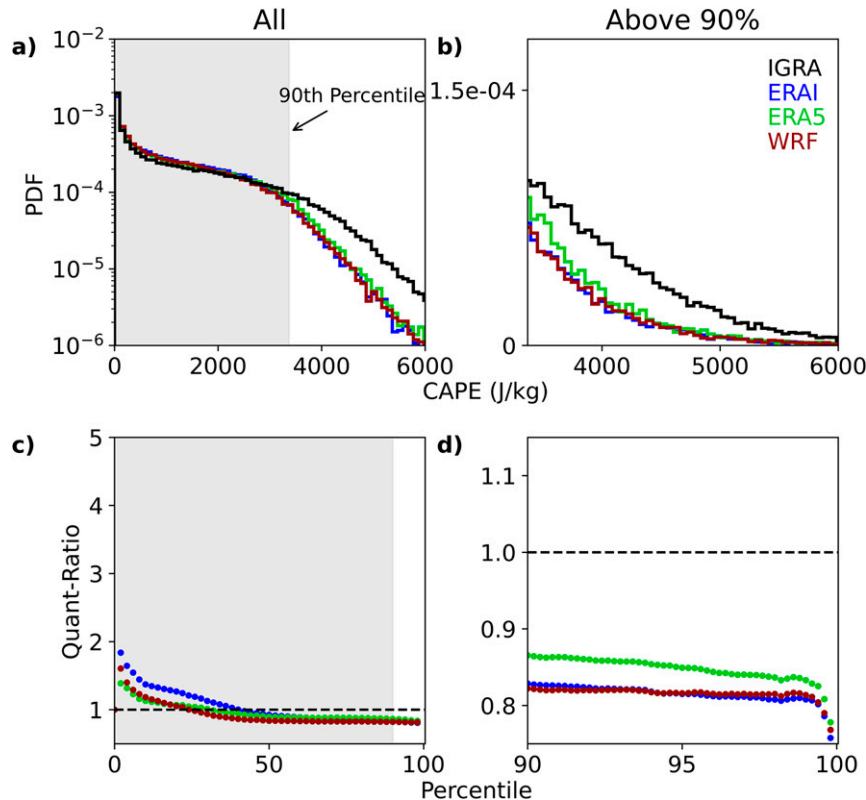


FIG. 3. As in Fig. 2, but for MUCAPE instead of SBCAPE. Points with zero CAPE are excluded from the analysis (23%–35% of the datasets, see Table 2). We match the time and locations of model output to IGRA observations. PDF x axes are cut off at 6000 J kg^{-1} , as less than 0.4% of all points lie above the limit. For IGRA, the 90th percentile is about 3370 J kg^{-1} , the 95th percentile $\sim 4010 \text{ J kg}^{-1}$, and the 97.5th percentile $\sim 4550 \text{ J kg}^{-1}$.

100 hPa used to calculate MLCAPE. [See Bunkers et al. (2002) for discussion of alternate choices.] The appropriate CAPE definition differs according to the scientific question addressed. Some authors argue that MLCAPE is most appropriate for characterizing the average properties of the parcel being initiated by convection, and MUCAPE may best capture convective extremes (Craven et al. 2002; Bunkers et al. 2002; Brooks et al. 2003), but we focus on SBCAPE in this work for several reasons. First, most prior CAPE comparison studies have used either only SBCAPE (e.g., Gensini et al. 2014; Gartzke et al. 2017), or all three definitions (Coniglio 2012; Taszarek et al. 2018). SBCAPE is the most widely used in the climate community (Riemann-Campe et al. 2009; Singh et al. 2017), and several common CR-closure convective parameterizations use SBCAPE (e.g., Zhang and McFarlane 1995; Xie and Zhang 2000; Wang et al. 2015). Finally, using SBCAPE allows the most straightforward bias attribution, since it allows us to test the effect of errors in surface properties alone.

To understand the implications of the different definitions, we compare surface-based CAPE with that of the most unstable layer, MUCAPE, the maximum possible value for each profile (Fig. 1). Because our focus is on incidences of very high CAPE, we are especially interested in whether different CAPE definitions lead to different understanding of the high

tail (defined as incidents above 90th percentile CAPE). In all datasets, the higher the CAPE value, the more similar SBCAPE and MUCAPE become (Fig. 1a). In conditions conducive to extreme weather ($>4000 \text{ J kg}^{-1}$), SBCAPE and MUCAPE are essentially identical in reanalyses and the WRF Model output. Radiosondes show a slightly larger distinction between SBCAPE and MUCAPE in all conditions. Mathematically, this means that model/radiosonde bias in MUCAPE must be more negative than those in SBCAPE. The pressure difference of the most unstable layer from the surface follows a similar pattern (Fig. 1b). The higher the CAPE value, the more the most unstable layer approaches the surface, though observations again show more distinction. In conditions with SBCAPE $\sim 1000 \text{ J kg}^{-1}$, the average most unstable parcel in radiosonde soundings lies $\sim 30 \text{ hPa}$ above the surface, but only $\sim 10 \text{ hPa}$ in reanalyses and the WRF Model. Above $>4000 \text{ J kg}^{-1}$, the most unstable layer in reanalyses and model lies at the surface.

In the last decade, some authors have argued that CAPE in any definition is not the best metric for diagnosing conditions conducive to severe weather. Grünwald and Brooks (2011) propose using instead the maximum updraft velocity calculated with parcel theory (WMAX), which arguably better represents the intensity of updrafts. Several recent papers have

followed that convention (Brooks 2013; Púčik et al. 2015; Taszarek et al. 2018). Results here can be translated to differences in WMAX distributions by using the approximation $WMAX \sim \sqrt{2CAPE}$, since the calculation of WMAX assumes no initial parcel vertical velocity.

4. Results—Biases in CAPE distributions

a. CAPE distributions across datasets

Comparison of the distribution of CAPE in the datasets considered shows immediately that reanalyses and the WRF Model output underpredict incidences of very high CAPE. Table 1 shows the breakdown of SBCAPE above or below threshold values, and Table 2 the same for MUCAPE. In all datasets, CAPE distributions are zero peaked; that is, a large fraction ($\sim 40\%$) of cases involve zero CAPE, even in the highly convective summertime. The frequency of zero CAPE is broadly similar across datasets, but in reanalyses and the WRF Model, incidences of extreme CAPE drop off sharply, with values above 4000 J kg^{-1} substantially underpredicted in both definitions. For SBCAPE, reanalyses and the WRF Model produce 40%–50% fewer incidences of values $> 4000 \text{ J kg}^{-1}$. For MUCAPE, the underprediction is even more severe, with 65%–75% of all incidences missed. (Biases in MLCAPE are intermediate between those in SBCAPE and MUCAPE.)

These biases in the high tail are related to a too-narrow distribution of CAPE in the WRF Model and reanalyses. That is, reanalyses and the WRF Model produce too few incidences of both extremely low and extremely high CAPE and too many incidences of intermediate CAPE. Figures 2 and 3 show distributions of nonzero CAPE values for SBCAPE and MUCAPE, respectively. Because valid zero values make up a large fraction of soundings, the choice whether to include them can potentially affect analysis, but in the datasets here, zero incidences are similar (Tables 1 and 2). We use two methods to show distributions: histograms (probability density functions, or PDFs) and quantile ratio plots. PDFs provide a basic sense of the CAPE distribution, and quantile ratio plots highlight distributional differences. Quantile ratio plots are constructed by taking the ratio of individual quantiles of two distributions being compared (e.g., CAPE in reanalysis and radiosondes); a value above 1 means that given quantile is overestimated. A simple multiplicative transformation produces a horizontal line whose value is the ratio of means, and a too-narrow distribution produces a slope downward to the right.

Reanalyses and the WRF Model output considered here show the downward and rightward slope characteristic of too-narrow distributions: values are too large in low quantiles and too small in high quantiles. SBCAPE in the 20th–60th percentiles ($50\text{--}1000 \text{ J kg}^{-1}$) is overestimated by 84%–94%, but above the 95th percentile is underestimated by 6%–10%. These distributional errors occur even though mean SBCAPE values are similar in all datasets: within the range from +1% to +6% with zeroes included, that is, slightly *larger* in reanalyses and the WRF Model than in radiosondes. This distinction highlights the need for distributional analysis, since even

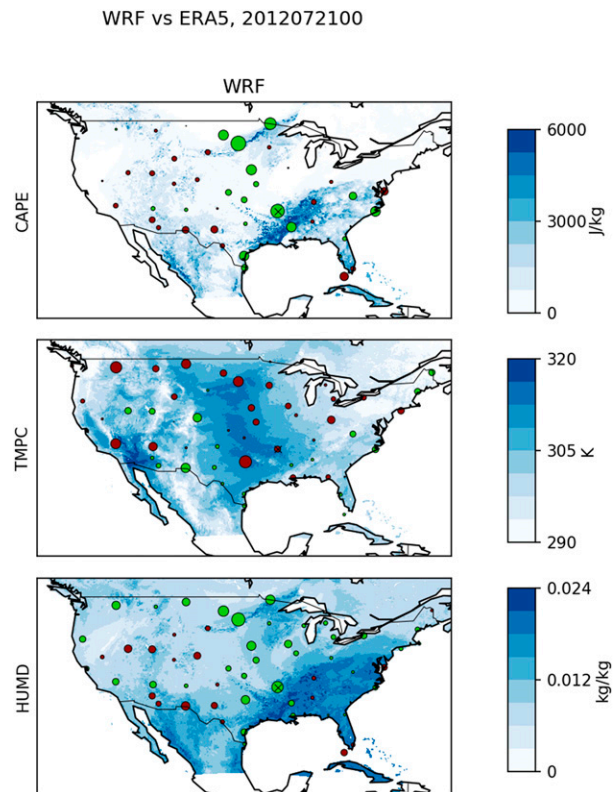


FIG. 4. Snapshot of WRF simulation output at 0000 UTC 21 Jul 2012. Panel colors show SBCAPE, 2-m temperature, and specific humidity. Ocean values are masked out. Circles show IGRA stations, with circle area showing the magnitude of bias in each variable and color indicating its sign (red = high, green = low). Note the low CAPE bias in the central United States associated with too hot and too dry model conditions. Soundings marked “X” may also be affected by errors in the location of the warm front.

severe distributional biases may not be reflected in mean values (shown in supplemental Table S1).

The too-narrow distributions in reanalyses and the WRF Model are also seen when alternate definitions of CAPE are used. (Figs. 3 and S2 show MUCAPE and MLCAPE, respectively.) However, MUCAPE and MLCAPE also show significant low mean bias, with mean values from -20% to -28% for MUCAPE and from -16% to -22% for MLCAPE. These low biases lead to even stronger deficits in the high tail, with quantiles above the 95th underestimated by $\sim 18\%$ – 20% in MUCAPE and $\sim 15\%$ – 17% in MLCAPE.

b. Spatiotemporal structure

Biases might be expected to show spatiotemporal structure, since CAPE is strongly linked to spatially complex fields of temperature and humidity. This relationship is illustrated in Fig. 4, which shows a summertime snapshot of surface values from the WRF simulation (SBCAPE, temperature, and specific humidity), coincident with the radiosonde launch time at which CAPE values are typically highest (0000 UTC, late afternoon or early evening in the contiguous United States). The

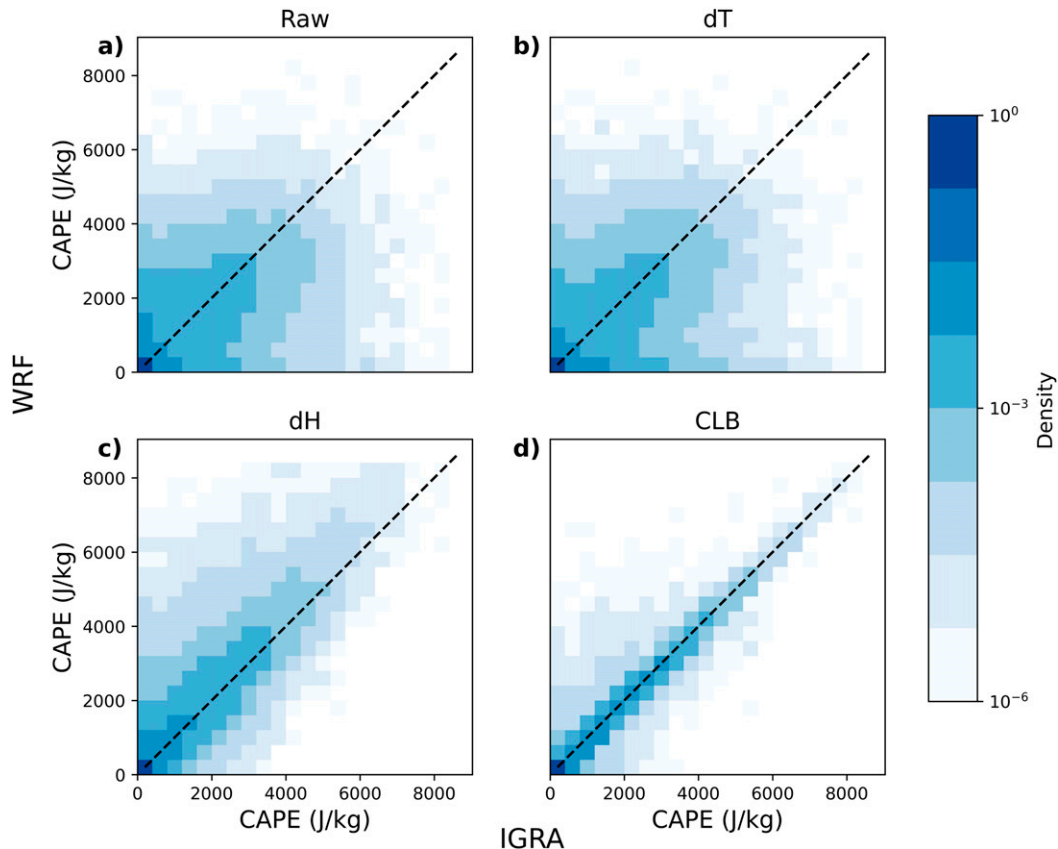


FIG. 5. Comparison of SBCAPE in WRF and radiosonde observations, for all points during summer (MJJA) 2001–12 when observations are available, inclusive of zeroes. Color bar shows log density (midpoint color is 1% of all observations), and both axes are also log scale. (a) Raw data, showing wide scatter. Recalculated WRF CAPE using (b) observed surface temperature, (c) observed surface humidity, and (d) all surface values from observations. All recalculated CAPE values also involve a pressure correction whose effects are small. For analogous figures for ERAI and ERA5, see Figs. S4–S5.

time period shown is affected by a frontal system that brings high humidity to the southeast and high temperatures to the central United States (see Fig. S3 for a weather map). CAPE reaches extreme values only where both temperature and specific humidity are high, resulting in strong spatial gradients and a narrow band of extreme CAPE extending from southeastern Texas to northern Mississippi.

Two processes appear to drive the spatially correlated CAPE errors in Fig. 4: large-scale patterns of model bias, and mismatches in the location of fronts or other weather features associated with strong gradients. The former is clearly evident in Fig. 4. The WRF Model is too warm and too dry in the central United States, coincident with and likely causing a large region of underestimated model CAPE. The warm-and-dry bias in this WRF simulation is extensively documented (Liu et al. 2017; Morcrette et al. 2018). Error in front location, on the other hand, likely produces overestimation of CAPE in stations in Tennessee and Alabama in Fig. 4. Large-scale and weather-related errors have different consequences for comparisons of CAPE in models and observations. Large-scale biases should be persistent and will affect the overall distribution

of CAPE. Fine-scale weather-related errors, on the other hand, vary rapidly on time scales of hours. While they can produce severe mismatch in individual soundings, and therefore introduce scatter in a model–observation comparison, they should have minimal effect on CAPE distributions.

c. Calibration with ground observations

Scatter in SBCAPE errors is in fact large in the WRF Model and reanalysis products considered here, with correlation coefficients against radiosonde values of only $R = 0.68$ – 0.83 . Figure 5 shows the comparison of WRF and radiosondes (Fig. 5a, $R = 0.68$; see Figs. S4 and S5 for ERAI and ERA5). Similar behavior is found in other studies, for example, Gensini et al. (2014) find correlation coefficients of 0.36 – 0.71 ; Taszarek et al. (2018) find 0.71 ; and Gartzke et al. (2017) show that reanalysis and satellite pseudosoundings cannot reproduce radiosonde observed SBCAPE at individual time steps.

Following Gartzke et al. (2017), we test to see if these inaccuracies can be corrected by simply replacing surface thermodynamics fields with those from radiosondes (Fig. 5). That is, we test whether errors in the WRF Model and reanalysis

SBCAPE are driven primarily by surface conditions rather than by the structure of atmospheric profiles. Both factors can be important because CAPE is a function of the integrated buoyancy across the convective layer, which is determined by both parcel and environmental temperature and moisture. In Fig. 5, we successively replace surface values in WRF output, first temperature and pressure (Fig. 5b), then specific humidity and pressure (Fig. 5c), then all surface fields (Fig. 5d).

Surface values do seem to govern SBCAPE bias almost entirely. For WRF, correcting the surface specific humidity raises the correlation coefficient from 0.68 to 0.91, and replacing all surface fields raises it to 0.99, removing scatter almost entirely. While correcting temperature does not raise the correlation coefficient in WRF, and instead lowers it to 0.65, for other datasets the temperature correction also contributes positively (see supplemental Table S2). We also consider an alternate measure of correspondence, the percentage of points that fall within $\pm 800 \text{ J kg}^{-1}$ of the one-to-one line (the width of two cells in Fig. 5). For raw WRF data, the percentage is 78.6% ($\text{RMSE} = 846 \text{ J kg}^{-1}$); correcting surface temperature raises the percentage slightly to 79.3% ($\text{RMSE} = 875 \text{ J kg}^{-1}$); correcting surface humidity raises it to 90.2% ($\text{RMSE} = 535 \text{ J kg}^{-1}$), and full calibration to 99.5% ($\text{RMSE} = 162 \text{ J kg}^{-1}$). Results for ERAI and ERA5 are similar. Adjustment of surface values also largely corrects the distributional problems at high CAPE, so that for quantiles above 0.9, corrected SBCAPE values in reanalyses and the WRF Model match those from radiosondes to within the range from -0.2% to $+2.0\%$. Correcting upper-tropospheric profiles has a minimal effect on CAPE values.

5. Results—CAPE in temperature and humidity space

The fact that reanalyses and modeled SBCAPE can be brought into agreement with radiosondes by simply replacing surface values implies that thermodynamic fields at upper levels are not important factors in SBCAPE biases. It may then be reasonable to consider SBCAPE as a function of surface thermodynamic fields alone. We therefore examine SBCAPE in the 2D parameter space of temperature (T) and specific humidity (H) to ask the following questions: 1) Is the density distribution of SBCAPE in T – H parameter space similar in reanalyses, model, and radiosondes? 2) What surface conditions are related to the highest SBCAPE days? 3) What factors drive model and reanalysis biases in SBCAPE?

a. Dependence on surface temperature and humidity

CAPE distributions in T – H parameter space are in fact highly robust across all datasets. Figure 6 shows the heat map of mean CAPE for radiosonde measurements, with data binned in steps of 3 K and 1.35 g kg^{-1} . CAPE values show a smooth gradient from lowest values at bottom left (warm and dry conditions) to highest at top right (hot and humid). This dependence on surface T and H is similar for all datasets (supplemental Fig. S7). Contour lines at 2000 and 4000 J kg^{-1} for radiosonde observations are therefore nearly identical to those for other datasets (overlain). This similarity means that surface T and H robustly determine SBCAPE in all datasets. Of course, each T – H bin in Fig. 6 involves an underlying CAPE

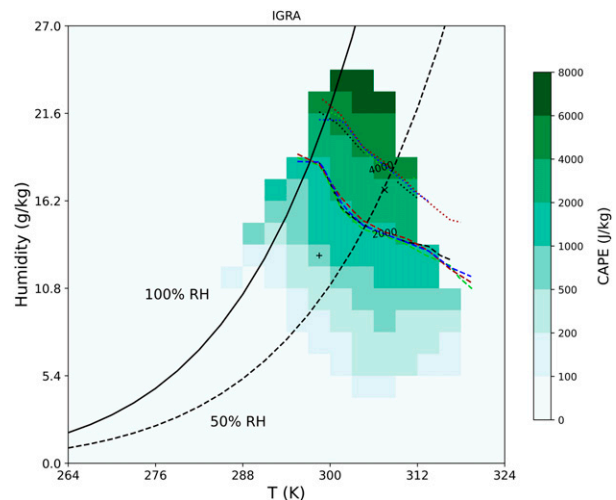


FIG. 6. Mean radiosonde observed SBCAPE in surface temperature and specific humidity parameter space, for the entire dataset: summer (MJJA) 2001–12 over the contiguous United States, inclusive of all launch times and of zero values. Only bins with at least 10 samples are colored. Colors denote mean CAPE values averaged in bins of 3 K and 1.35 g kg^{-1} . Solid and dashed lines mark contours of 100% and 50% RH at $p = 1013 \text{ hPa}$. Soundings with lower surface p will be displaced up and left from these RH contours. Symbols “+” and “x” mark two cases (“warm” and “hot”) used in Fig. 7. Contours show approximate limits for 2000 and 4000 J kg^{-1} SBCAPE for all datasets with no surface corrections applied: IGRA (black), ERAI (blue), ERA5 (green), and WRF (red). Similarity of contours means that all datasets show similar bivariate distributions. See Fig. S6 for the absolute occurrence in each bin in IGRA and Fig. S7 for analogous figures for all datasets.

distribution, but distributions are nearly identical for all datasets; see Fig. 7 for two examples. These results support the previous finding that bias in SBCAPE can be explained by bias in surface measurements alone. (See supplemental Fig. S8 for distributions of reanalyses and model errors in T , RH, and H for the profiles shown in Fig. 7 and Table S3 for summary statistics.)

Only a restricted set of conditions tend to produce the “high tail” of CAPE distributions associated with extreme, high-impact weather. We show both 2000 and 4000 J kg^{-1} contours to bracket prior definitions of extreme weather thresholds. For example, Brooks et al. (2003); Trapp et al. (2009), and Diffenbaugh et al. (2013) all use 2000 J kg^{-1} in MLCAPE, which corresponds to SBCAPE $\sim 3000 \text{ J kg}^{-1}$ in our dataset. The conditions producing mean SBCAPE above 2000 J kg^{-1} involve temperatures above 297 K for 100% relative humidity (RH), or above 304 K for 50% RH. For mean SBCAPE above 4000 J kg^{-1} , the required temperatures are 2–3 K warmer, that is, 299 K at 100% RH or 307 K at 50% RH. Significantly higher SBCAPE values are possible: in the most extreme conditions regularly sampled by radiosondes, 308 K at 65% RH, the average observed SBCAPE is over 7400 J kg^{-1} . Reanalyses and the WRF Model rarely produce SBCAPE values this high (8

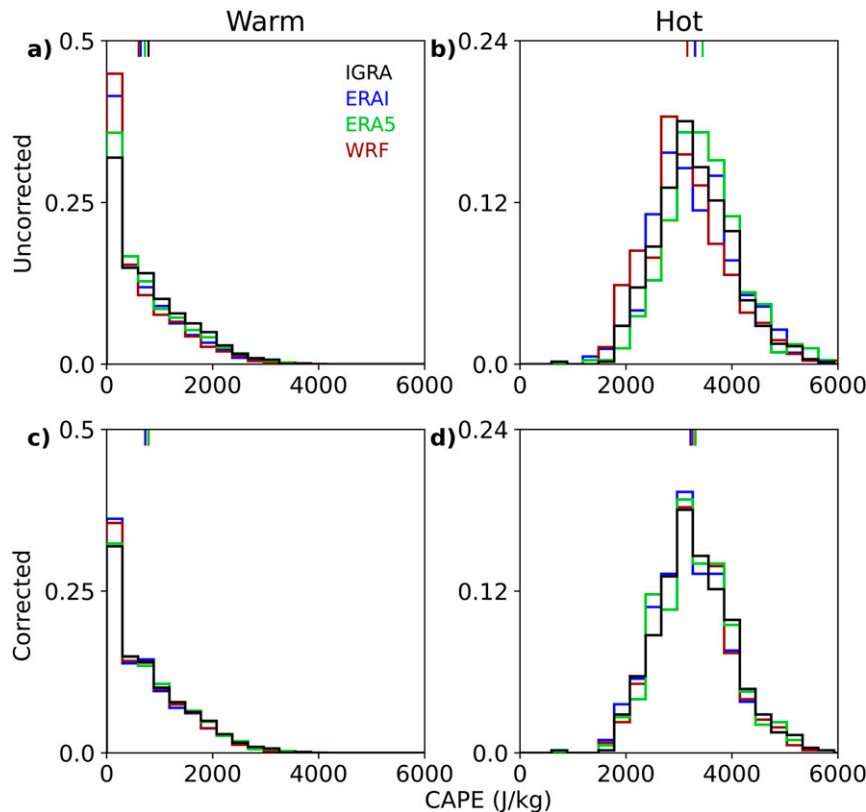


FIG. 7. Comparison of SBCAPE in all datasets for specified T - H grid cell: the “warm” example is centered at 298.5 K and 12.825 g kg^{-1} (63.4% RH) and has mean SBCAPE 791 J kg^{-1} ; the “hot” example is at 307.5 K and 16.875 g kg^{-1} (48.7% RH) with mean SBCAPE 3308 J kg^{-1} . Each bin is 3 K in width and 1.35 g kg^{-1} in height. (a),(b) Uncorrected SBCAPE from reanalyses and the WRF Model, and (c),(d) corrected with IGRA surface values. Note that since the correction involves adjusting surface T and H , the profiles sampled in top and bottom rows are different. The “warm” bin has 2438 profiles in the uncorrected data and 2063 in the corrected, while “hot” has 378 and 508, respectively. Tick marks at the top of each panel show the mean of each distribution. Distributions are very similar; correcting surface values only slightly adjusts means (from a maximum bias of -5% in uncorrected data to -2% after correction).

out of a million incidences, while observed incidences are nearly 10 times more frequent at 60 out of a million), not because they differ in fundamental atmospheric physics but because they rarely sample the appropriate surface conditions.

b. Identifying sources of CAPE bias

Because SBCAPE is strongly determined by surface temperature and humidity, biases in SBCAPE in reanalyses and the WRF Model appear driven by biases in these surface thermodynamic values. We can therefore use the T - H diagram to identify the factors that lead to underprediction of the high tail of CAPE. Figures 8 and 9 use the same T - H diagram as in Fig. 6, only now we show not the heat map of CAPE but the density of observations of each T - H grid cell and the difference in that number between datasets. Because the diurnal cycle strongly affects surface values, we show separate figures for 0000 UTC (U.S. late afternoon/evening) and 1200 UTC (U.S. early morning), omitting the limited number of samples at

other times. Reanalyses and the WRF Model all underpredict the extreme T - H values associated with extreme CAPE.

Of the two times routinely sampled by radiosondes, the cooler 1200 UTC launches—early morning in the United States—do not generally involve conditions associated with high CAPE (Fig. 8a). Conditions at this time are almost never warm enough to produce SBCAPE $> 2000 \text{ J kg}^{-1}$, even though relative humidities are high, with a tight distribution centered around $\sim 80\%$. Both reanalyses and WRF are biased dry, underpredicting incidences close to saturation, and WRF is also biased warm (Figs. 8b–d).

Most of the observed extreme CAPE values occur during the warmer late-afternoon 0000 UTC launches (Fig. 9a). Relative humidities are lower then because specific humidity does not change much during daytime warming: the modal (most probable) 0000 UTC surface conditions are between 303 and 309 K and $\sim 50\%$ RH, with mean SBCAPE of $\sim 3000 \text{ J kg}^{-1}$, similar to the “hot” example of Fig. 7. Because reanalyses and WRF are

Density at 12UTC

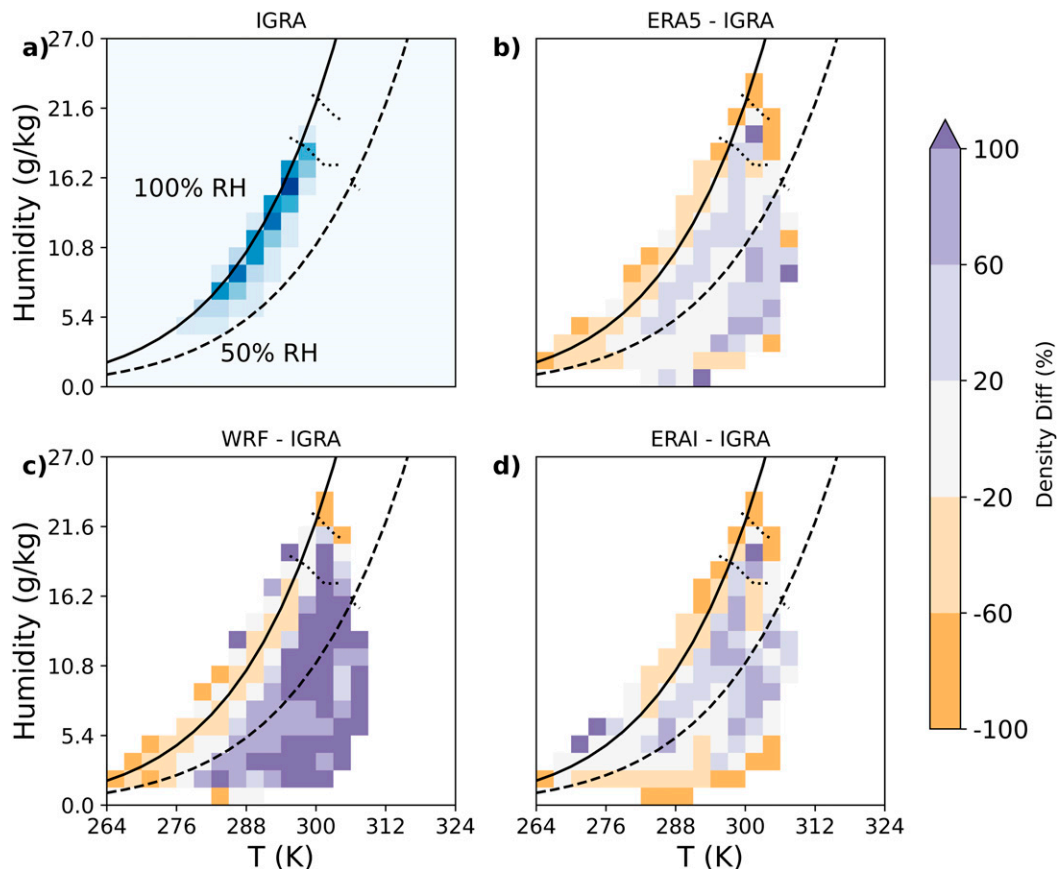


FIG. 8. (a) Density of observed surface conditions in temperature–specific humidity parameter space at 1200 UTC (early morning in the contiguous United States), again for 2001–12 MJJA radiosonde observations. Contours are repeated from Fig. 6 to mark conditions associated with 2000 and 4000 J kg^{-1} SBCAPE. Darkest blue color shown is 5.6%–6.4% of distribution; lightest is 0%–0.8%. Grids with no more than ten samples are defined as outliers and removed (only 0.03% of all model or reanalysis samples). Nighttime and early morning conditions are tightly distributed in relative humidity (RH $\sim$ 80%) and tend to be relatively cool ($T < 300$ K), with almost no conditions sampled that would tend to produce SBCAPE $> 2000 \text{ J kg}^{-1}$. (b)–(d) Heat maps of density differences between model/reanalysis and observations for ERA5, WRF, and ERAI. Color scale shows fractional difference after normalizing each bin by IGRA raw density. Orange = underpredicting and purple = overpredicting. Reanalyses and the WRF Model all underestimate relative humidities (orange near the RH=100% contour), and WRF shows a strong warm dry bias (dark purple in lower right).

dry biased, they underpredict high RH conditions in general and especially the extreme hot and humid conditions associated with the largest CAPE. On the other hand, reanalyses and WRF overpredict hot and dry conditions, WRF especially so (Figs. 9b–d). The combined warm and dry bias explains why correcting WRF surface temperatures alone does not improve the match to radiosonde CAPE measurements.

6. Results—Diurnal cycles of CAPE and biases

As shown in section 5, the largest CAPE biases in the WRF Model and reanalyses occur when conditions are most favorable to high CAPE, that is, in daytime. This diurnal difference could result from inherent nonlinearity but could also reflect a

bias in some aspect of the diurnal cycle of surface thermodynamic fields. We therefore examine the diurnal cycle of surface temperature and specific humidity in reanalyses, model, and radiosondes. As an illustration, we show in Fig. 10 a 5-day episode exhibiting large CAPE error, which is broadly representative of problematic reanalyses and model pseudosoundings, and in Fig. 11 we compare this episode to summertime climatological mean diurnal cycles for all, low-CAPE, and high-CAPE conditions (10th/90th SBCAPE percentiles). The sequence in Fig. 10 runs from 24 to 28 July 2012 at a station in Wilmington, North Carolina. On three occasions, radiosonde profiles show extreme CAPE of nearly 5000 J kg^{-1} , but model and reanalyses grossly underpredict these excursions, producing CAPE values $\sim 2500 \text{ J kg}^{-1}$ too low. Since the temporal pattern of

Density at 00UTC

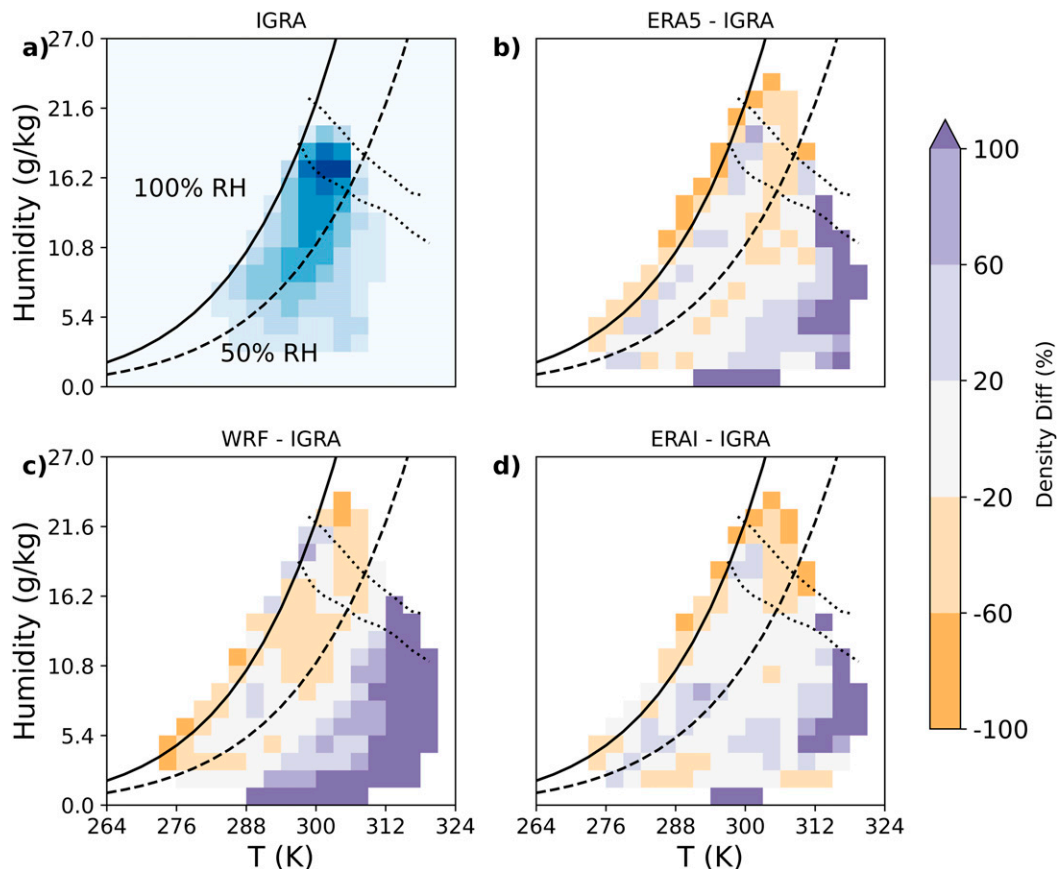


FIG. 9. As in Fig. 8, but for 0000 UTC (late afternoon/early evening in the contiguous United States) (a) Density of observed surface conditions in T - H diagram. At this time period the density distribution peaks in conditions associated with 2000–4000 J kg^{-1} CAPE. Darkest blue color shown is 2.1%–2.4% of distribution; lightest is 0%–0.3%. (b)–(d) Density differences between reanalyses / model and radiosondes. ERA5 and ERAI underpredict both the highest relative humidities and the highest temperatures (orange near the $\text{RH} = 100\%$ contour and on the right side), while WRF shows a warm dry bias (purple in lower right). Reanalyses and model all severely underpredict the conditions associated with extreme CAPE (orange in upper right).

temperature evolution appears synchronous in all datasets, these biases appear unrelated to any mismatch of frontal systems. (See Fig. S9 for the WRF Model output and bias structure and Fig. S10 for the 26 July weather map.)

In the example episode of Fig. 10, strong CAPE discrepancies result when models and reanalyses fail to capture short-term increases in specific humidity associated with extreme CAPE. Biases are driven by humidity, since throughout the 5-day period the WRF Model and reanalyses are slightly too warm, with a fairly accurate diurnal temperature cycle (~ 5 – 6 K, with the WRF Model exhibiting the largest amplitude). Comparison with climatological means on the T - H diagram of Fig. 11 suggests that the daytime humidity rise in the example episode is extreme even for high-CAPE conditions, but the climatological biases are otherwise broadly similar. Reanalyses and WRF have overall dry biases that are exacerbated in WRF during the day (in the climatological case, by an actual daytime

drop in specific humidity). Diurnal cycles of temperature cycles are similar, though daytime warming is slightly too weak in reanalyses, and WRF has an overall high temperature bias of ~ 1.3 K. In all cases, too-low surface-level humidity appears to be the driving factor that strongly suppresses incidences of extreme CAPE.

CAPE biases in the example episode of Fig. 10 differ in ERA reanalyses and the WRF Model, but both produce deficits in specific humidity. In all datasets, temperatures match reasonably well in early morning (1200 UTC), but daytime temperature rise is slightly too small in ERA reanalyses and considerably too large in WRF. ERA RH is reasonably accurate throughout, so its too-low temperatures are associated with a small specific humidity deficit. In WRF, specific humidity actually falls during the day, something not seen in reanalyses or radiosondes, contributing to erroneously low relative humidities. During the two “missed-high-CAPE”

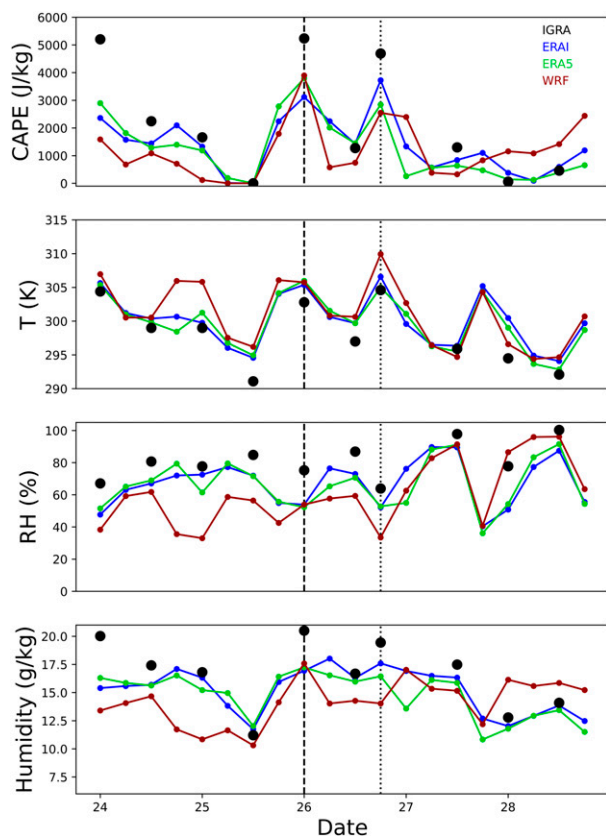


FIG. 10. An example episode of high CAPE and substantial CAPE error: 5 days from 24 to 28 Jul 2012 over Wilmington, North Carolina, color coded as before. Reanalyses and the WRF Model are shown every 6 hours; IGRA soundings are generally every 12 hours (although note the irregular timing for radiosonde launching on the 26th, when daytime sampling occurred 6 hours earlier than usual, at 1800 UTC). Vertical lines mark the two examples discussed in text. Over this entire period, reanalyses and WRF Model show hot and dry bias; when the bias in absolute humidity is large, the dry bias produces too-low CAPE even despite too-high temperatures.

episodes, WRF RH is ~ 25 percentage points below that in radiosondes.

To demonstrate that biases during this single-station episode are typical for warm conditions conducive of extreme CAPE formation, we compare its diurnal cycle to climatological means across all stations. Figure 11 uses the temperature–humidity (T – H) diagram to show mean diurnal cycles of the 5-day episode at station Wilmington (Fig. 11a) and across the whole dataset (Fig. 11b), showing both the overall summertime average and also subsets of days involving the highest and lowest radiosonde SBCAPE values (90th/10th percentiles). The biases in the July 2012 example episode are similar to those generally experienced in high-CAPE conditions. On average, the ERA reanalyses experience slightly too-weak daytime warming (i.e., a too-small diurnal cycle in temperature) and are slightly too dry. WRF has an overall high bias in temperature of ~ 1.3 K and is substantially too dry. The warm

dry bias in WRF is exacerbated at midday by the fact that specific humidity erroneously drops during the day. This daytime humidity loss strongly suppresses incidences of extreme CAPE.

7. Conclusions and discussion

Despite the importance of CAPE to both model construction and meteorology, few prior studies have evaluated CAPE biases against radiosondes on a large enough scale to evaluate climatological distributions. This study of nearly 200 000 proximity soundings in two reanalyses and a convection-permitting model confirms consistent patterns of distributional bias. CAPE distributions are too narrow in all cases, with underprediction of the most extreme values that are associated with severe weather events. Values in the high tail (95th percentile and above) are 6%–10% too low in surface-based CAPE and even more severely underestimated under alternative definitions, at 18%–20% too low in MUCAPE and 15%–17% too low in MLCAPE.

In this study, both distributional biases and “mismatch error” in CAPE appear driven by conditions at the surface and/or boundary layer. SBCAPE shows a tight and similar dependence on surface temperature and humidity in all datasets; the dependence is so strong that CAPE distributions as a function of surface T , H are almost identical, even though individual profiles may have inaccurate surface values. The underprediction of the high tail of CAPE occurs simply because reanalyses and WRF runs underpredict the hot and humid conditions associated with extreme CAPE values.

These results emphasize the importance of land and boundary layer treatment in the performance of high-resolution models. Discussion of improving models has tended to focus on increased resolution and its effect on the atmospheric profile (e.g., Fossler et al. 2015). However, the similarity of biases in model and reanalysis output with resolved and parameterized convection suggests that surface biases are unrelated to the treatment of convection. Many authors have noted that SBCAPE is strongly dependent on surface conditions (e.g., Maddox and Doswell 1982; Zhang 2002; Donner and Phillips 2003; Guichard et al. 2004), but land surface feedbacks may be crudely treated even in state-of-the-art high-resolution models (Prein et al. 2015).

Dry biases such as those seen here could be produced by misrepresentation of land surface evaporation, by excess vertical mixing of the boundary layer, or, for the central United States, by too-weak advection of moisture from the Gulf of Mexico (Feng et al. 2021). In this study, the greater bias in MUCAPE than SBCAPE across all datasets points to boundary layer processes as the problematic elements. It is well established that treatment of mixing in boundary layer (PBL) schemes can modify the diurnal cycle of temperature and humidity (Kalthoff et al. 2009; García-Díez et al. 2013; Xu et al. 2019). Several recent studies have evaluated the impacts of PBL treatment on CAPE, and consistently find that “local” schemes tend to undermix and moisten the boundary layer while “nonlocal” ones, like those used in the datasets shown here, overmix and dry it (Coniglio et al. 2013; Cohen et al. 2017; Evans et al. 2018). The YSU scheme used in our WRF runs is especially prone to producing a dry bias (Coniglio et al. 2013; Liu et al. 2017).

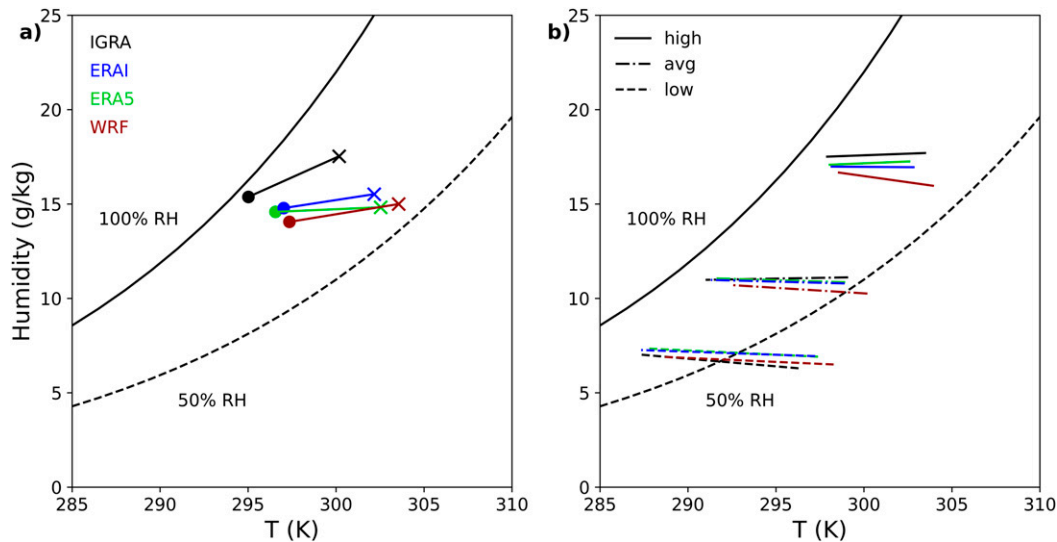


FIG. 11. The diurnal cycle in T - H space in all datasets. Thick lines connect 0000 UTC (right end, marked by “x”) and 1200 UTC (left end, marked by “o”) values. Color code follows the convention throughout this work. (a) The average over the 5-day episode in Fig. 10. (b) Mean summertime diurnal cycles over the entire domain, for all profiles (dot-dashed), and for high-CAPE (solid) and low-CAPE (dashed) subsets, defined as 0000 UTC SBCAPE values above 90th/below 10th percentile in each dataset, and values 12 hours later. In all cases, WRF and ERA are biased dry, and WRF is biased warm. In the example in (a), daytime observed specific humidity increases more than in WRF or ERA. In the climatological mean in (b), specific humidity is roughly constant in observations and ERA but erroneously decreases in daytime in the WRF Model, exacerbating midday bias.

While prior studies evaluating the effect of boundary layer treatment on CAPE have generally evaluated only mean values and have found only small biases, results here suggest that the high tails can be much more strongly affected. This finding is consistent with Evans et al. (2018), who note in a small sample of soundings that dry biases produced by nonlocal PBL schemes appear larger when observed CAPE is larger. The dependence of biases on underlying conditions means that even models and data products whose mean CAPE is well validated may be inaccurate in capturing the strong convective events that lead to large socioeconomic losses. This problem cannot be assessed with studies that match soundings to severe weather events, since model displacement of weather features means that “mismatch” error is large and proximity soundings will not necessarily capture the same meteorological context. On the flip side, the accuracy of relatively low CAPE is also critical for convective parameterization schemes, since convective initiation thresholds are commonly set at only 65 J kg^{-1} . Subtle distributional biases can therefore affect convective triggering and total mass flux, and indirectly affect precipitation diurnal timing and amplitude. Given the importance of CAPE as a key meteorological parameter linking the large-scale environment to weather-scale events, and its sensitivity to details of boundary layer treatment, its evaluation warrants careful distributional analysis.

Acknowledgments. The authors thank Zhihong Tan for valuable discussion and insight, and three anonymous reviewers for helpful comments and suggestions. This work is supported by the Center for Robust Decision-making on Climate and Energy

Policy (RDCEP), which is funded by the NSF Decision Making Under Uncertainty program, Award SES-1463644. This work was completed in part with resources provided by the University of Chicago Research Computing Center. The authors thank the National Center for Atmospheric Research (NCAR) for providing the WRF dataset that made this article possible.

Data availability statement. The 4-km high-resolution WRF simulations of the current and future climate of North America are available from NCAR Research Data (<https://doi.org/10.5065/D6V40SXP>). The radiosonde observations are obtained from NOAA Integrated Global Radiosonde Archive (IGRA; <https://www1.ncdc.noaa.gov/pub/data/igra/>). The reanalysis output is available from ECMWF (<https://www.ecmwf.int/en/forecasts/datasets/browse-reanalysis-datasets>).

REFERENCES

- Adams, D. K., and E. P. Souza, 2009: CAPE and convective events in the southwest during the North American monsoon. *Mon. Wea. Rev.*, **137**, 83–98, <https://doi.org/10.1175/2008MWR2502.1>.
- Allen, J. T., and D. J. Karoly, 2014: A climatology of Australian severe thunderstorm environments 1979–2011: Inter-annual variability and ENSO influence. *Int. J. Climatol.*, **34**, 81–97, <https://doi.org/10.1002/joc.3667>.
- Baba, Y., 2019: Spectral cumulus parameterization based on cloud-resolving model. *Climate Dyn.*, **52**, 309–334, <https://doi.org/10.1007/s00382-018-4137-z>.
- Bloch, C., R. O. Knuteson, A. Gambacorta, N. R. Nalli, J. Gartzke, and L. Zhou, 2019: Near-real-time surface-based CAPE from merged hyperspectral IR satellite

- sounder and surface meteorological station data. *J. Appl. Meteor. Climatol.*, **58**, 1613–1632, <https://doi.org/10.1175/JAMC-D-18-0155.1>.
- Blumberg, W. G., K. T. Halbert, T. A. Supinie, P. T. Marsh, R. L. Thompson, and J. A. Hart, 2017: SHARPPy: An open-source sounding analysis toolkit for the atmospheric sciences. *Bull. Amer. Meteor. Soc.*, **98**, 1625–1636, <https://doi.org/10.1175/BAMS-D-15-00309.1>.
- Brooks, H. E., 2013: Severe thunderstorms and climate change. *Atmos. Res.*, **123**, 129–138, <https://doi.org/10.1016/j.atmosres.2012.04.002>.
- , and J. P. Craven, 2002: A database of proximity soundings for significant severe thunderstorms, 1957–1993. *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 16.2, <https://ams.confex.com/ams/pdfpapers/46680.pdf>.
- , C. A. Doswell, and J. Cooper, 1994: On the environments of tornadic and nontornadic mesocyclones. *Wea. Forecasting*, **9**, 606–618, [https://doi.org/10.1175/1520-0434\(1994\)009<0606:OTEOTA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0606:OTEOTA>2.0.CO;2).
- , J. W. Lee, and J. P. Craven, 2003: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res.*, **67–68**, 73–94, [https://doi.org/10.1016/S0169-8095\(03\)00045-0](https://doi.org/10.1016/S0169-8095(03)00045-0).
- Bunkers, M. J., B. A. Klimowski, and J. W. Zeitler, 2002: The importance of parcel choice and the measure of vertical wind shear in evaluating the convective environment. *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., P8.2, <https://ams.confex.com/ams/pdfpapers/47319.pdf>.
- Chen, J., A. Dai, Y. Zhang, and K. L. Rasmussen, 2020: Changes in convective available potential energy and convective inhibition under global warming. *J. Climate*, **33**, 2025–2050, <https://doi.org/10.1175/JCLI-D-19-0461.1>.
- Cohen, A. E., S. M. Cavallo, M. C. Coniglio, H. E. Brooks, and I. L. Jirak, 2017: Evaluation of multiple planetary boundary layer parameterization schemes in southeast U.S. cold season severe thunderstorm environments. *Wea. Forecasting*, **32**, 1857–1884, <https://doi.org/10.1175/WAF-D-16-0193.1>.
- Coniglio, M. C., 2012: Verification of RUC 0–1-h forecasts and SPC mesoscale analyses using VORTEX2 soundings. *Wea. Forecasting*, **27**, 667–683, <https://doi.org/10.1175/WAF-D-11-00096.1>.
- , J. Correia, P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF Model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, **28**, 842–862, <https://doi.org/10.1175/WAF-D-12-00103.1>.
- Copernicus Climate Change Service, 2017: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Climate Data Store, accessed 9 September 2019, <https://cds.climate.copernicus.eu/cdsapp#!/home>.
- Cortés-Hernández, V. E., F. Zheng, J. Evans, M. Lambert, A. Sharma, and S. Westra, 2016: Evaluating regional climate models for simulating sub-daily rainfall extremes. *Climate Dyn.*, **47**, 1613–1628, <https://doi.org/10.1007/s00382-015-2923-4>.
- Craven, J. P., R. E. Jewell, and H. E. Brooks, 2002: Comparison between observed convective cloud-base heights and lifting condensation level for two different lifted parcels. *Wea. Forecasting*, **17**, 885–890, [https://doi.org/10.1175/1520-0434\(2002\)017<0885:CBOCCB>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0885:CBOCCB>2.0.CO;2).
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Diffenbaugh, N. S., M. Scherer, and R. J. Trapp, 2013: Robust increases in severe thunderstorm environments in response to greenhouse forcing. *Proc. Natl. Acad. Sci. USA*, **110**, 16 361–16 366, <https://doi.org/10.1073/pnas.1307758110>.
- Dong, W., Y. Lin, J. S. Wright, Y. Xie, X. Yin, and J. Guo, 2019: Precipitable water and CAPE dependence of rainfall intensities in China. *Climate Dyn.*, **52**, 3357–3368, <https://doi.org/10.1007/s00382-018-4327-8>.
- Donner, L. J., and V. T. Phillips, 2003: Boundary layer control on convective available potential energy: Implications for cumulus parameterization. *J. Geophys. Res.*, **108**, 4701, <https://doi.org/10.1029/2003JD003773>.
- Doswell, C. A., and E. N. Rasmussen, 1994: The effect of neglecting the virtual temperature correction on CAPE calculations. *Wea. Forecasting*, **9**, 625–629, [https://doi.org/10.1175/1520-0434\(1994\)009<0625:TEONTV>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0625:TEONTV>2.0.CO;2).
- Durre, I., R. S. Vose, and D. B. Wuertz, 2006: Overview of the integrated global radiosonde archive. *J. Climate*, **19**, 53–68, <https://doi.org/10.1175/JCLI3594.1>.
- , —, and —, 2008: Robust automated quality assurance of radiosonde temperatures. *J. Appl. Meteor. Climatol.*, **47**, 2081–2095, <https://doi.org/10.1175/2008JAMC1809.1>.
- ECMWF, 2016: Part I: Observations. IFS Documentation CY41R2, 72 pp., <https://www.ecmwf.int/node/16646>.
- Evans, C., S. J. Weiss, I. L. Jirak, A. R. Dean, and D. S. Nevius, 2018: An evaluation of paired regional/convection-allowing forecast vertical thermodynamic profiles in warm-season, thunderstorm-supporting environments. *Wea. Forecasting*, **33**, 1547–1566, <https://doi.org/10.1175/WAF-D-18-0124.1>.
- Feng, Z., F. Song, K. Sakaguchi, and L. R. Leung, 2021: Evaluation of mesoscale convective systems in climate simulations: Methodological development and results from MPAS-CAM over the United States. *J. Climate*, **34**, 2611–2633, <https://doi.org/10.1175/JCLI-D-20-0136.1>.
- Fosser, G., S. Khodayar, and P. Berg, 2015: Benefit of convection permitting climate model simulations in the representation of convective precipitation. *Climate Dyn.*, **44**, 45–60, <https://doi.org/10.1007/s00382-014-2242-1>.
- García-Díez, M., J. Fernández, L. Fita, and C. Yagüe, 2013: Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe. *Quart. J. Roy. Meteor. Soc.*, **139**, 501–514, <https://doi.org/10.1002/qj.1976>.
- Gartzke, J., R. Knuteson, G. Przybyl, S. Ackerman, and H. Revercomb, 2017: Comparison of satellite-, model-, and radiosonde-derived convective available potential energy in the southern Great Plains region. *J. Appl. Meteor. Climatol.*, **56**, 1499–1513, <https://doi.org/10.1175/JAMC-D-16-0267.1>.
- Gensini, V. A., T. L. Mote, and H. E. Brooks, 2014: Severe-thunderstorm reanalysis environments and collocated radiosonde observations. *J. Appl. Meteor. Climatol.*, **53**, 742–751, <https://doi.org/10.1175/JAMC-D-13-0263.1>.
- Groenemeijer, P. H., and A. van Delden, 2007: Sounding-derived parameters associated with large hail and tornadoes in the Netherlands. *Atmos. Res.*, **83**, 473–487, <https://doi.org/10.1016/j.atmosres.2005.08.006>.
- Grünwald, S., and H. E. Brooks, 2011: Relationship between sounding derived parameters and the strength of tornadoes in Europe and the USA from reanalysis data. *Atmos. Res.*, **100**, 479–488, <https://doi.org/10.1016/j.atmosres.2010.11.011>.
- Guichard, F., and Coauthors, 2004: Modelling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single-column models. *Quart. J. Roy.*

- Meteor. Soc.*, **130**, 3139–3172, <https://doi.org/10.1256/qj.03.145>.
- Haimberger, L., 2007: Homogenization of radiosonde temperature time series using innovation statistics. *J. Climate*, **20**, 1377–1403, <https://doi.org/10.1175/JCLI4050.1>.
- , C. Tavalato, and S. Sperka, 2008: Toward elimination of the warm bias in historic radiosonde temperature records—Some new results from a comprehensive intercomparison of upper-air data. *J. Climate*, **21**, 4587–4606, <https://doi.org/10.1175/2008JCLI1929.1>.
- Hart, J., and W. Korotky, 1991: The SHARP workstation v1.50 users guide. NOAA National Weather Service Final Tech. Rep., 30 pp. [Available from NWS Eastern Region Headquarters, 630 Johnson Ave., Bohemia, NY 11716.]
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, <https://doi.org/10.1175/MWR3199.1>.
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res. Atmos.*, **113**, D13103, <https://doi.org/10.1029/2008JD009944>.
- Kalthoff, N., and Coauthors, 2009: The impact of convergence zones on the initiation of deep convection: A case study from COPS. *Atmos. Res.*, **93**, 680–694, <https://doi.org/10.1016/j.atmosres.2009.02.010>.
- King, A. T., and A. D. Kennedy, 2019: North American supercell environments in atmospheric reanalyses and RUC-2. *J. Appl. Meteor. Climatol.*, **58**, 71–92, <https://doi.org/10.1175/JAMC-D-18-0015.1>.
- Lee, M.-I., S. D. Schubert, M. J. Suarez, J.-K. E. Schemm, H.-L. Pan, J. Han, and S.-H. Yoo, 2008: Role of convection triggers in the simulation of the diurnal cycle of precipitation over the United States Great Plains in a general circulation model. *J. Geophys. Res.*, **113**, D02111, <https://doi.org/10.1029/2007JD008984>.
- Lepore, C., D. Veneziano, and A. Molini, 2015: Temperature and CAPE dependence of rainfall extremes in the eastern United States. *Geophys. Res. Lett.*, **42**, 74–83, <https://doi.org/10.1002/2014GL062247>.
- Liu, C., and Coauthors, 2017: Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dyn.*, **49**, 71–95, <https://doi.org/10.1007/s00382-016-3327-9>.
- Maddox, R. A., and C. A. Doswell, 1982: An examination of jet stream configurations, 500 mb vorticity advection and low-level thermal advection patterns during extended periods of intense convection. *Mon. Wea. Rev.*, **110**, 184–197, [https://doi.org/10.1175/1520-0493\(1982\)110<0184:AEJSC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<0184:AEJSC>2.0.CO;2).
- Moncrieff, M. W., and M. J. Miller, 1976: The dynamics and simulation of tropical cumulonimbus and squall lines. *Quart. J. Roy. Meteor. Soc.*, **102**, 373–394, <https://doi.org/10.1002/qj.49710243208>.
- Morcrette, C. J., and Coauthors, 2018: Introduction to CAUSES: Description of weather and climate models and their near-surface temperature errors in 5 day hindcasts near the southern Great Plains. *J. Geophys. Res. Atmos.*, **123**, 2655–2683, <https://doi.org/10.1002/2017JD027199>.
- Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, <https://doi.org/10.1029/2010JD015139>.
- Paquin, D., R. de Elía, and A. Frigon, 2014: Change in North American atmospheric conditions associated with deep convection and severe weather using CRCM4 climate projections. *Atmos.–Ocean*, **52**, 175–190, <https://doi.org/10.1080/07055900.2013.877868>.
- Prein, A. F., and Coauthors, 2015: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Rev. Geophys.*, **53**, 323–361, <https://doi.org/10.1002/2014RG000475>.
- Púčik, T., P. Groenemeijer, D. Rýva, and M. Kolář, 2015: Proximity soundings of severe and nonsevere thunderstorms in central Europe. *Mon. Wea. Rev.*, **143**, 4805–4821, <https://doi.org/10.1175/MWR-D-15-0104.1>.
- Rasmussen, E. N., and D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148–1164, [https://doi.org/10.1175/1520-0434\(1998\)013<1148:ABCOSD>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<1148:ABCOSD>2.0.CO;2).
- Rasmussen, K. L., A. F. Prein, R. M. Rasmussen, K. Ikeda, and C. Liu, 2017: Changes in the convective population and thermodynamic environments in convection-permitting regional climate simulations over the United States. *Climate Dyn.*, **55**, 383–408, <https://doi.org/10.1007/s00382-017-4000-7>.
- Rasmussen, R., and C. Liu, 2017: High resolution WRF simulations of the current and future climate of North America. National Center for Atmospheric Research Computational and Information Systems Laboratory Research Data Archive, accessed 30 October 2019, <https://doi.org/10.5065/D6V40SXP>.
- Riemann-Campe, K., K. Fraedrich, and F. Lunkeit, 2009: Global climatology of convective available potential energy (CAPE) and convective inhibition (CIN) in ERA-40 reanalysis. *Atmos. Res.*, **93**, 534–545, <https://doi.org/10.1016/j.atmosres.2008.09.037>.
- Singh, M. S., Z. Kuang, E. D. Maloney, W. M. Hannah, and B. O. Wolding, 2017: Increasing potential for intense tropical and subtropical thunderstorms under global warming. *Proc. Natl. Acad. Sci. USA*, **114**, 11 657–11 662, <https://doi.org/10.1073/pnas.1707603114>.
- Song, F., and G. J. Zhang, 2017: Improving trigger functions for convective parameterization schemes using GOAmazon observations. *J. Climate*, **30**, 8711–8726, <https://doi.org/10.1175/JCLI-D-17-0042.1>.
- , and —, 2018: Understanding and improving the scale dependence of trigger functions for convective parameterization using cloud-resolving model data. *J. Climate*, **31**, 7385–7399, <https://doi.org/10.1175/JCLI-D-17-0660.1>.
- Sun, X., M. Xue, J. Brotzge, R. A. McPherson, X.-M. Hu, and X.-Q. Yang, 2016: An evaluation of dynamical downscaling of central plains summer precipitation using a WRF-based regional climate model at a convection-permitting 4 km resolution. *J. Geophys. Res. Atmos.*, **121**, 13 801–13 825, <https://doi.org/10.1002/2016JD024796>.
- Taszarek, M., H. E. Brooks, B. Czernecki, P. Szuster, and K. Fortuniak, 2018: Climatological aspects of convective parameters over Europe: A comparison of ERA-Interim and sounding data. *J. Climate*, **31**, 4281–4308, <https://doi.org/10.1175/JCLI-D-17-0596.1>.
- , N. Pilgus, J. T. Allen, V. Gensini, H. E. Brooks, and P. Szuster, 2020: Comparison of convective parameters derived from ERA5 and MERRA2 with rawinsonde data over Europe and North America. *J. Climate*, **34**, 3211–3237, <https://doi.org/10.1175/JCLI-D-20-0484.1>.
- Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large

- winter cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, <https://doi.org/10.1175/JAS-D-13-0305.1>.
- Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the rapid update cycle. *Wea. Forecasting*, **18**, 1243–1261, [https://doi.org/10.1175/1520-0434\(2003\)018<1243:CPSWSE>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2).
- Trapp, R. J., N. S. Diffenbaugh, and A. Gluhovsky, 2009: Transient response of severe thunderstorm forcing to elevated greenhouse gas concentrations. *Geophys. Res. Lett.*, **36**, L01703, <https://doi.org/10.1029/2008GL036203>.
- Wang, Y.-C., H.-L. Pan, and H.-H. Hsu, 2015: Impacts of the triggering function of cumulus parameterization on warm-season diurnal rainfall cycles at the atmospheric radiation measurement Southern Great Plains site. *J. Geophys. Res. Atmos.*, **120**, 10 681–10 702, <https://doi.org/10.1002/2015JD023337>.
- Xie, S., and M. Zhang, 2000: Impact of the convection triggering function on single-column model simulations. *J. Geophys. Res. Atmos.*, **105**, 14 983–14 996, <https://doi.org/10.1029/2000JD900170>.
- , and Coauthors, 2019: Improved diurnal cycle of precipitation in E3SM with a revised convective triggering function. *J. Adv. Model. Earth Syst.*, **11**, 2290–2310, <https://doi.org/10.1029/2019MS001702>.
- Xu, L., H. Liu, Q. Du, and X. Xu, 2019: The assessment of the planetary boundary layer schemes in WRF over the central Tibetan Plateau. *Atmos. Res.*, **230**, 104644, <https://doi.org/10.1016/j.atmosres.2019.104644>.
- Yano, J.-I., M. Bister, Ž. Fuchs, L. Gerard, V. T. J. Phillips, S. Barkidija, and J.-M. Piriou, 2013: Phenomenology of convection-parameterization closure. *Atmos. Chem. Phys.*, **13**, 4111–4131, <https://doi.org/10.5194/acp-13-4111-2013>.
- Zhang, G. J., 2002: Convective quasi-equilibrium in midlatitude continental environment and its effect on convective parameterization. *J. Geophys. Res.*, **107**, 4220, <https://doi.org/10.1029/2001JD001005>.
- , and N. A. McFarlane, 1995: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmos.–Ocean*, **33**, 407–446, <https://doi.org/10.1080/07055900.1995.9649539>.