A Case Study Investigating the Low Summertime CAPE Behavior in the Global Forecast System⬙

XIA SUN 🟢,[a,b,d] DOMINIKUS HEINZELLER,[a,b,d] LIGIA BERNARDET,[b] LINLIN PAN,[a,b,d] WEIWEI LI,[c,d] DAVID TURNER,[b] AND JOHN BROWN[b]

[a] *Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado*
[b] *National Oceanic and Atmospheric Administration/Global Systems Laboratory, Boulder, Colorado*
[c] *National Center for Atmospheric Research, Boulder, Colorado*
[d] *Developmental Testbed Center, Boulder, Colorado*

ABSTRACT: Convective available potential energy (CAPE) is an important index for storm forecasting. Recent versions (v15.2 and v16) of the Global Forecast System (GFS) predict lower values of CAPE during summertime in the continental United States than analysis and observation. We conducted an evaluation of the GFS in simulating summertime CAPE using an example from the Unified Forecast System Case Study collection to investigate the factors that lead to the low CAPE bias in GFS. Specifically, we investigated the surface energy budget, soil properties, and near-surface and upper-level meteorological fields. Results show that the GFS simulates smaller surface latent heat flux and larger surface sensible heat flux than the observations. This can be attributed to the slightly drier-than-observed soil moisture in the GFS that comes from an offline global land data assimilation system. The lower simulated CAPE in GFS v16 is related to the early drop of surface net radiation with excessive boundary layer cloud after midday when compared with GFS v15.2. A moisture-budget analysis indicates that errors in the large-scale advection of water vapor does not contribute to the dry bias in the GFS at low levels. Common Community Physics Package single-column model (SCM) experiments suggest that with realistic initial vertical profiles, SCM simulations generate a larger CAPE than runs with GFS IC. SCM runs with an active LSM tend to produce smaller CAPE than that with prescribed surface fluxes. Note that the findings are only applicable to this case study. Including more warm-season cases would enhance the generalizability of our findings.

SIGNIFICANCE STATEMENT: Convective available potential energy (CAPE) is one of the key parameters for severe weather analysis. The low bias of CAPE is identified by forecasters as one of the key issues for the NOAA operational global numerical weather prediction model, Global Forecast System (GFS). Our case study shows that the lower CAPE in GFS is related to the drier atmosphere than observed within the lowest 1 km. Further investigations suggest that it is related to the drier atmosphere that already exists in the initial conditions, which are produced by the Global Data Assimilation System, in which an earlier 6-h GFS forecast is combined with current observations. It is also attributed to the slightly lower simulated soil moisture than observed. The lower CAPE in GFS v16 when compared with GFS v15.2 in the case analyzed here is related to excessive boundary layer cloud formation beginning at midday that leads to a drop of net radiation reaching the surface and thus less latent heat feeding back to the low-level atmosphere.

1. Introduction

Precipitation is a key ingredient of the tropospheric water budget and global climate. Clouds directly modulate Earth's energy budget and affect atmospheric dynamic flow (Grabowski and Moncrieff 2004). In numerical weather prediction (NWP) models, convective available potential energy (CAPE) is an important index to predict convective precipitation, as well as guiding how atmospheric buoyancy impacts the strength of convection in cumulus parameterizations via the parameterizations' closure assumption (Bechtold et al. 2014; Brooks et al. 2007). CAPE is defined as the integrated amount of work that an upward buoyant force would perform on a given mass of air if it rose vertically through the entire atmosphere, where the buoyant force is a function of the atmospheric profiles for temperature and moisture. Thus, CAPE requires realistic simulations of temperature and near-surface moisture.

Specifically, CAPE is calculated by integrating the buoyant energy present in the free convective layer (FCL) from the level of free convection (LFC) to the equilibrium level (EL). The LFC is the height at which the temperature of an air parcel exceeds the temperature of the surrounding air, resulting in the parcel being unstable relative to its environment. The

---

⬙ Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/WAF-D-22-0208.s1.

Heinzeller's current affiliation: Joint Center for Satellite Data Assimilation, UCAR Community Programs, Boulder, Colorado.

Corresponding author: Xia Sun, xia.sun@noaa.com

EL, on the other hand, is the height at which the temperature of the surrounding air exceeds the temperature of the air parcel, making it stable relative to its environment. CAPE measures the cumulative buoyant energy in the FCL from the LFC to the EL (Blanchard 1998):

$$CAPE = g \int_{z_{LFC}}^{z_{EL}} \left(\frac{T_{v_p} - T_{v_e}}{T_{v_e}} \right) dz, \qquad (1)$$

where g is the gravitational acceleration, with z_{EL} and z_{LFC} denoting the height of the EL and LFC, respectively. The T_{v_p} and T_{v_e} represent the virtual temperature of the ascending air parcel and surrounding environment, respectively.

It has been established in prior studies that high CAPE does not guarantee convection (Xie and Zhang 2000), such as with the presence of strong "convective inhibition" that can hinder convection initiation. On the other hand, low CAPE does not necessarily exclude convection, as seen in high-shear weak-CAPE events (Sherburn and Parker 2014). Nonetheless, accurate representation of CAPE in NWP models is the prerequisite to effective use of model-forecast CAPE by operational forecasters. The factors that lead to high CAPE values differ case by case, but generally, warm temperatures, high availability of moisture from the surface, and strong surface winds all contribute to the buildup of CAPE (Agard and Emanuel 2017). One example is dynamic convection, which occurs when the warm air is lifted above a surface-based cold air mass (Corfidi et al. 2008). Another example is radiative cooling in the free troposphere, which can increase instability for parcels lifted from the low levels. A third example is heating of the near-surface air due to the absorption of solar radiation or large latent heat fluxes with the presence of moisture, which can also drive an increase in CAPE magnitudes (Emanuel 1994).

Summertime convection prediction is often extremely challenging over the conterminous United States (CONUS) because of the weak synoptic-scale features. This situation results in subtle and weak patterns of meso-alpha-scale and synoptically forced vertical motion (as implied by, for example, quasigeostrophic theory) that can be resolved by the GFS (C768 resolution ∼ 12-km horizontal grid spacing). Consequently, the impact on the thermodynamic fields needed to support convection is also subtle and weak. The implication is that forecasts of convection require a high-quality initial analysis of the wind, mass and moisture fields as well as a skillful NWP model. The values of CAPE are influenced by the specific parcel being lifted (Williams and Renno 1993). There are multiple methods of determining the lifted parcel, including lifting a parcel from the surface [i.e., surface-layer-based CAPE (SBCAPE)] and lifting a parcel representative of the lowest 100-hPa layer [i.e., mixed-layer based CAPE (MLCAPE; Rochette et al. 1999)]. The model-predicted field of CAPE, regardless of flavor (surface based, mixed layer, etc.), is critical to identifying areas thermodynamically favorable for convection in the model forecast and is scrutinized by forecasters concerned with severe local storms and heavy rainfall.

Errors in forecasts made by NWP models can be attributed to several factors. These factors can be broadly categorized into three areas within the NWP forecasting process: initial conditions from the analysis, unresolved processes represented in physics parameterization schemes, and large-scale dynamics resolved by dynamic core. The initial conditions are generated through data assimilation (DA). DA integrates information from both the forecast model and available observations through an algorithmic framework (Rabier 2005). Thus, disentangling model errors from the data assimilation algorithms used in generating initial conditions is a challenging task.

The Unified Forecast System (UFS) (Jacobs 2021) represents a significant contribution to the NWP community by not only making NOAA's operational medium-range NWP model—that is, Global Forecast System (GFS)—source code publicly available, but also by providing the entire application package, allowing users to run the model on non-NOAA platforms. The current operational version of the GFS v16 differs from the previous version (GFS v15) in terms of physics, number of vertical levels (127 levels vs 64 levels), and the inclusion of four-dimensional incremental analysis update in the DA process (Tallapragada 2020). A prior evaluation of GFS v15 and v16 performance over the eastern United States from June to September in 2020 shows that GFS v16 simulates a lower CAPE than v15, which itself already underestimates CAPE when compared with sounding data (Yang 2020). The low CAPE in GFS was also identified by forecasters and raised as one of the key model issues (Sims et al. 2021). Mensch (2021) validated convection indices derived from GFS against radiosonde data from various field campaigns over midlatitude and tropical regions. Their results demonstrate the consistently underestimated CAPE in both GFS v15 and GFS v16 for events where the CAPE is larger than 3000 J kg^{-1}. However, a process-oriented comprehensive evaluation of the GFS and an investigation of the contributing factors that lead to the bias in CAPE during summertime high CAPE events has not been made and is the main motivation for the present study.

In this work, we conducted a case study focusing on CAPE bias in GFS using the summertime case selected from the UFS Case Studies Platform (Sun et al. 2021). The UFS Case Studies platform provides a set of cases that reveal major forecast challenges in GFS. The case collection covers atmospheric phenomena of winter storms, hurricanes, extreme temperature, convection, and low-level inversions. This platform serves as a repository for the research community to gather the necessary resources and instructions to conduct model runs using different UFS weather applications including global forecast applications, such as the GFS. Specifically, the platform provides initial condition datasets hosted on cloud storage, model configurations, and setups, and high-level preliminary simulation results when compared with reanalysis. In this study, we investigate a 2020 CAPE case, which occurred over the Great Plains during the daytime of 23 July 2020.

We focus on a comprehensive investigation of the GFS performance in simulating a summertime high CAPE event specifically looking at advection, boundary layer structure, and land–atmosphere interactions. Note that we will be focusing on the CONUS region, despite the fact that the GFS model is a global model. We attempt to attribute the model biases to processes and to advance the understanding of the physics

schemes' impacts on model behavior by harnessing a hierarchical testing framework (HTF). There is no single definition of HTF. A single-column model (SCM) is a simple system in model hierarchies (Ek et al. 2019). Another aspect of HTF is interaction denial, which involves denying interactions between different atmospheric processes. SCM applies interaction denial with removed physics and dynamic interactions that helps untangle impacts of model physics. In this work, we seek to address the following questions: 1) How well can the GFS predict the meteorological environment during a large-CAPE event in the summertime? 2) What are the factors that hinder the GFS from simulating the magnitudes of CAPE? It should be noted that the convection formation mechanism and statistical relationship between CAPE value and severe weather are beyond the scope of this study.

The paper is organized as follows: section 2 details the model configurations for GFS and SCM simulations and provides an overview of the observational dataset used in this study. The results, including synoptic patterns, vertical structures, and energy budget, are analyzed in section 3. Section 4 presents the results of SCM sensitivity experiments. A summary and a future outlook are provided in section 5.

2. Model simulations and observational datasets

a. Model configuration

1) GLOBAL FORECAST SYSTEM

The Common Community Physics Package (CCPP) (Heinzeller et al. 2023) provides a framework and physics modules to the UFS applications. As of CCPP v5.0, four physics suites are supported. In this study, simulations were conducted using two physics suites: GFS v15.2 and GFS v16. The reason we examined the current operational GFS model (v16) together with an older version (v15.2) is to attempt to identify the culprits that lead to the exacerbated CAPE low bias in GFS by comparing the simulation fields from the two model versions. We leveraged the CCPP capability to output tendencies generated from the dynamical core and individual physics parameterizations to advance the understanding of model uncertainties. It is pertinent to note that the present study solely investigates the impact of modifications in the physics parameterizations between the two versions of the GFS. Other changes that were made in the operational version, such as vertical resolution and data assimilation, have not been taken into account.

The GFS v16 physics suite inherits most of the physics schemes from GFS v15.2. The radiative transfer is based on the Rapid Radiative Transfer Model for Global Climate Models (RRTMG) shortwave and longwave radiation schemes (Iacono et al. 2000). The GFS surface layer scheme (Long 1984) and Noah land surface model are used to simulate the land–atmosphere exchange (Mitchell 2005). The Geophysical Fluid Dynamics Laboratory cloud microphysics scheme applies a bulk water technique to predict five hydrometeor species (Lin et al. 1983). Scale-awareness is implemented by the assumption that subgrid variability is a function of grid spacing. For shallow convection, the two physics suites adopt the

GFS Scale-Aware Mass-Flux (sa-MF) shallow convection scheme (Han et al. 2017). For deep convection, the GFS Scale-Aware Simplified Arakawa–Schubert (sa-SAS) deep convection scheme is employed (Han et al. 2017).

The main difference between the two physics suites lies in the PBL parameterizations. The hybrid Eddy-Diffusivity Mass Flux (EDMF) PBL and free atmosphere turbulence scheme in GFS v15.2 is a first-order turbulent transport scheme that accounts for subgrid scale vertical turbulent mixing within the individual grid columns (Han et al. 2016). The hybrid approach uses the mass flux scheme in the boundary layer under strongly unstable (convective) atmospheric conditions and switches to an eddy-diffusivity countergradient scheme under slightly unstable conditions, based on a threshold of Obukhov length. The mass flux accounts for nonlocal subgrid transport related to strong updrafts. Eddy diffusivity is based on PBL height (PBLH) and Monin–Obukhov similarity parameters. The GFS v16 scale-aware turbulent kinetic energy (TKE)-based moist EDMF PBL scheme is an extended version of the hybrid EDMF scheme (Han and Bretherton 2019). One major enhancement is that the eddy diffusivity is determined by the predicted TKE. The EDMF scheme is universally applied to the boundary layer in unstable conditions. The PBL scheme in GFS v16 also adds a scale-aware capability and includes a moist-adiabatic process to simulate the formation of stratocumulus when the updraft reaches the saturation point. The turbulent mixing induced at the top of the stratocumulus is represented by a mass-flux parameterization.

The global workflow (Friedman and Kolczynski 2022) developed by the NOAA Environmental Modeling Center (EMC) is the superstructure used to run the UFS weather model applications. The UFS Case Study platform provides detailed instructions on how to set up the global workflow and contains links to initial condition datasets for the 2020 July CAPE case. Specifically, the model runs were initialized at 0000 UTC 23 July 2020, with a 48-h forecast at an operational horizontal grid resolution of C768 (approximately 13 km) and 127 vertical levels. Note that the operational GFS v16 utilizes 127 levels, whereas GFS v15.2 employs 64 levels. To eliminate any impacts of changing vertical resolutions, we have used 127 levels for both experiments in our study. For visualization, the vertical level distribution below 750 hPa and the layer thickness at different pressure levels are illustrated in Fig. 1. The lowest model level is around 20 m above ground level, and there are 22 levels within the lowest 100 hPa.

2) CCPP SCM

In our attempt to understand and attribute the model errors, we used the CCPP SCM (Firl et al. 2022) and conducted several sensitivity runs. A SCM essentially simulates a single column using the physics package from the parent model. In our case we use CCPP physics suites with the SCM version of the UFS weather model. The CCPP SCM is a simplified version of the global circulation model, yet it includes the physics parameterizations that make it an ideal tool to test physics development in a much less computationally demanding setting than running the full global model, permitting more rapid
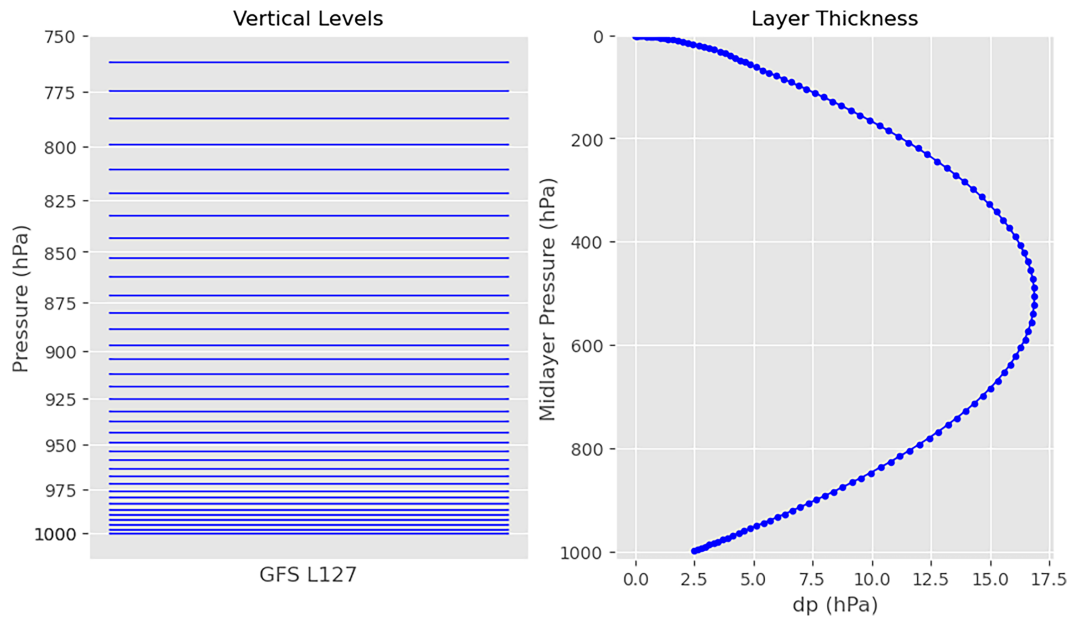
FIG. 1. Vertical levels of GFS below 750 hPa and layer thicknesses at different pressure levels.

discovery of coding errors and preliminary assessment of proposed physics improvement. The necessary input data to drive the CCPP SCM include initial vertical profiles, surface turbulent fluxes or necessary soil parameters to run a land surface model (LSM), and large-scale forcing. The advection fields are specified by a forcing dataset (Zhang et al. 2016), which can be obtained from observations or three-dimensional numerical weather models, such as GFS. The CCPP SCM provides a Python script to extract the necessary forcings from GFS global runs.

We used the suite of SCM_GFSv16 to mimic the physics of operational version of GFS v16. We employed a configuration of the SCM that allows use of specified surface fluxes (option *spec_sfc_flux*; Firl et al. 2022). The specified surface fluxes come from observations or UFS simulations as discussed in section 4. We also conducted a SCM run with active Noah LSM. The advective tendencies are estimated from UFS state variables in the initial condition and forcing generating tool provided in CCPP SCM.

b. Verification data

1) OBSERVATIONAL DATA

The Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) (Sisterson et al. 2016) observational network provides comprehensive measurement of components of the surface-energy budget, soil properties, and vertical profiles of meteorological variables. There are nine surface-energy-budget measurement sites over the SGP, with surface-energy fluxes, near-surface meteorology, and soil temperature and moisture at half-hourly frequency (Holdridge and Kyrouac 1993).

Radiosondes provide measurements of temperature, moisture, and wind speed with high vertical resolution, and are typically launched every 6 h at the SGP site. Higher temporal resolution (5 min) profiles of temperature and humidity in the boundary layer can be retrieved from the atmospheric emitted radiance interferometer (AERI) (Knuteson et al. 2004) using the tropospheric optimal estimation retrieval (TROPoe) algorithm (Turner and Löhnert 2014; Turner and Blumberg 2019). TROPoe-retrieved profiles have a maximum root-mean-square error of 1 K and 0.8 g kg^{-1} for temperature and moisture, respectively, within the lowest 2 km, when verified against radiosondes (Turner and Löhnert 2014). The TROPoe dataset has been widely used in research efforts that involve boundary layer structures (Bonin et al. 2015; Wagner et al. 2019) and data assimilation studies (e.g., (Chipilski et al. 2022; Degelia et al. 2020).

2) MODEL ANALYSIS AND REANALYSIS DATA

Here we adopt HRRR analyses (at forecast hour 0), together with radiosonde data, as the benchmark verification dataset for evaluating the spatial distribution of CAPE. The High-Resolution Rapid Refresh (HRRR) model is an hourly updated and convection-permitting model running at 3-km resolution over the CONUS (Dowell et al. 2022). HRRR is developed and maintained at NOAA's Global Systems Laboratory. The advanced data assimilation, rapid cycling, and high resolution are primary reasons why the HRRR is considered quite skillful (Fovell and Gallagher 2020). Comprehensive model verification tools (Turner et al. 2020) were used to demonstrate the accuracy of HRRR forecasts for a wide range of variables, including radiosondes, aircraft, surface meteorology, radar reflectivity, precipitation, and ceiling height (James et al. 2022). Prior studies have documented HRRR being skillful in various atmospheric processes or weather phenomena, including but not limited to PBL vertical structures
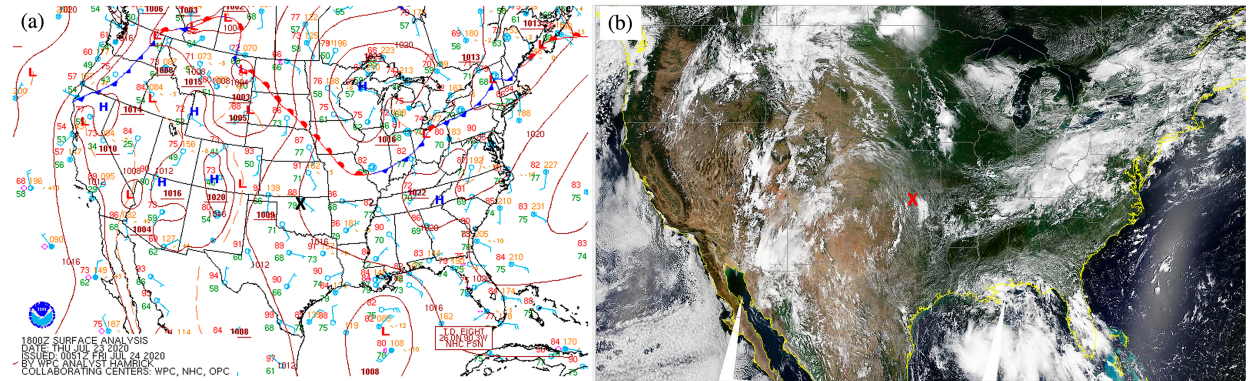
FIG. 2. (a) Weather Prediction Center surface analysis valid at 1800 UTC 23 Jul 2020, and (b) *Terra* MODIS true-color corrected reflectance on 23 Jul 2020. The approximate time when MODIS scans the CONUS is at 1900 UTC. The SGP central facility site C1 is marked by a red X.

(Fovell and Gallagher 2020), mesoscale convective systems (Pinto et al. 2015), and atmospheric rivers (English et al. 2021).

In terms of large-scale advection fields, we choose the fifth major global reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ERA5) to validate the moisture budget (details are in section 3d). ERA5 is based on the operational Integrated Forecast System (IFS) Cy41r2 (Hersbach et al. 2020) model of the ECMWF. The ERA5 dataset provides hourly data at 31-km spatial resolution and 137 vertical levels. As a reanalysis dataset, which combines observations and model ERA5 provides optimal estimates of global gridded atmospheric states. ERA5 generally performs well in comparisons with other reanalysis datasets (Guo et al. 2021; Mayer et al. 2019; Tarek et al. 2020; Trolliet et al. 2018) and several studies have used ERA5 to investigate the moisture budget (Mayer et al. 2021; Naakka et al. 2021; Ren et al. 2021).

3) LARGE-SCALE MOISTURE TENDENCY

To investigate whether the simulated large-scale dynamics contribute to the moisture tendency discrepancies, we compare the moisture budget derived from the GFS simulations with ERA5. The moisture budget is calculated as (Bellenger et al. 2015; Yanai et al. 1973)
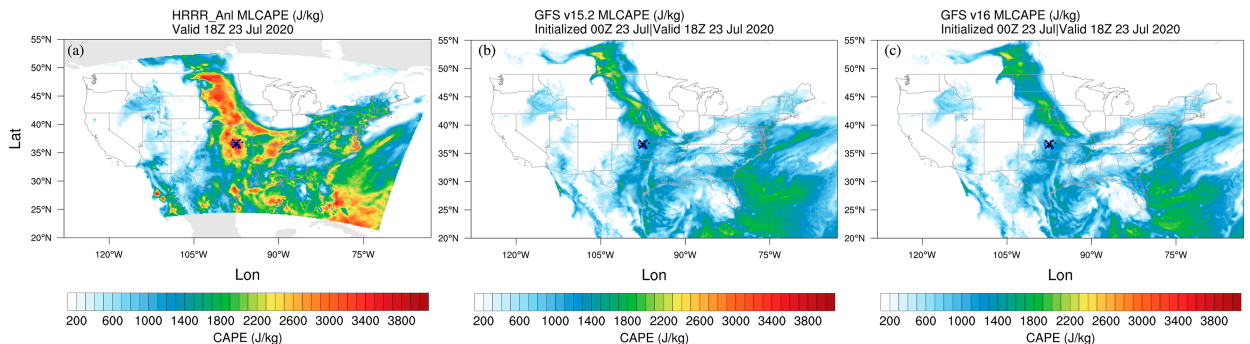
$$\frac{\partial q}{\partial t} = -\mathbf{v} \cdot \nabla q - \omega \frac{\partial q}{\partial p} - \frac{Q_2}{L}, \tag{2}$$

where $\partial q / \partial t$ is the large-scale water-vapor mixing ratio tendency; $-\mathbf{v} \cdot \nabla q$ represents the horizontal advection of moisture and $\mathbf{v}$ denotes the horizontal vector winds; $-\omega(\partial q/\partial p)$ describes the vertical advection of moisture and ω stands for the vertical velocity (Pa s^{-1}); $-(Q_2/L)$ is calculated as the residual term, in which Q_2 is called the *apparent moisture sink* and L represents the latent heat of condensation: $-(Q_2/L)$ represents condensation minus evaporation and transport by unresolved subgrid scale features. The budget items were calculated over a 120 km × 120 km region to characterize the large-scale dynamics.

3. Results

a. Case overview

The synoptic pattern and CAPE are examined at forecast hour 18 initialized at 0000 UTC 23 July 2020. The surface weather map for 1800 UTC 23 July 2020 (Fig. 2a) shows a cool air mass over the Great Lakes region, separated from a warm, humid air mass to the south and southwest by a quasi-stationary front. A large CAPE reservoir (Fig. 3a) was

FIG. 3. (a) HRRR-analyzed MLCAPE (mixed layer taken as the lowest 90 hPa of the forecast atmosphere) (shading; J kg^{-1}), and GFS (b) v15.2 and (c) v16 18-h forecast MLCAPE valid at 1800 UTC 23 Jul (shaded; J kg^{-1}). The location of the SGP central facility is indicated by an X, and the locations of the ARM surface energy balance sites at the SGP are marked with tiny blue dots.
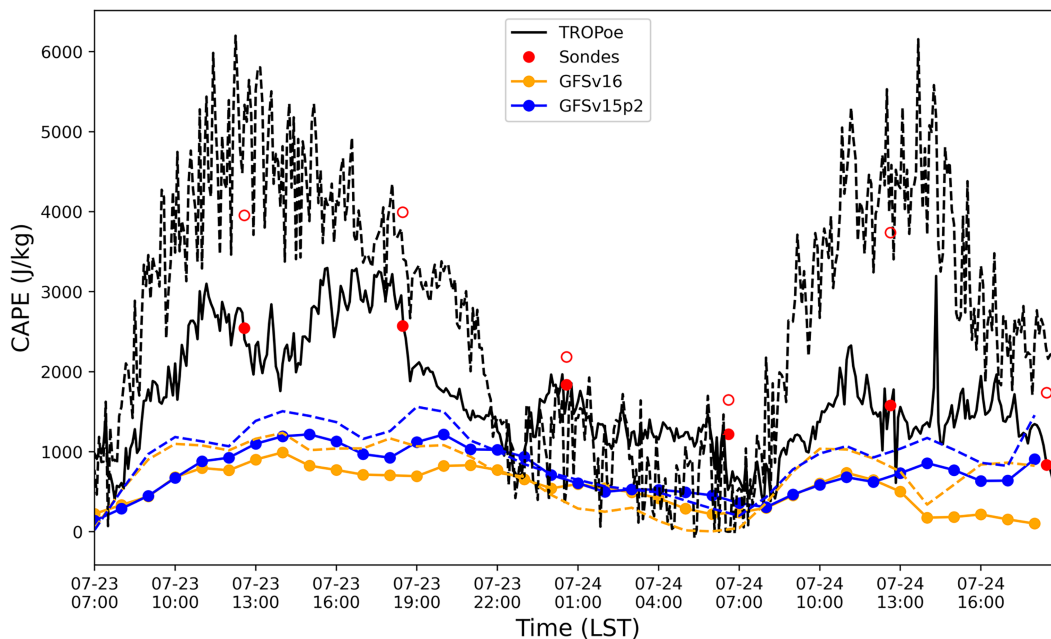
FIG. 4. Time series of SBCAPE (J kg^{-1}; dashed lines) and MLCAPE (J kg^{-1}; solid lines) derived from TROPoe (black), GFS v16 (orange), and GFS v15.2 (blue) at the SGP C1 site. The SBCAPE and MLCAPE values derived from radiosondes are given by open and filled red circles, respectively.

present over the Great Plains south of the front, north of 30°N and west of 95°W with a mean mixed-layer based CAPE (MLCAPE) value of 2090 J kg^{-1} using the routine sounding data from the University of Wyoming (http://weather.uwyo.edu/upperair/sounding.html). The spatial distributions of CAPE from GFS forecasts and HRRR analysis show that the GFS forecasts (Figs. 3b,c) captures the potential of organized convection along the boundary of the warm front over the Northern Great Plains, as well as the propagating shallow convection over the SGP. However, when compared with HRRR analysis, the MLCAPE derived from GFS forecast data at 1800 UTC exhibits a negative bias with a mean error (ME) of −1165 and −1170 J kg^{-1} for GFS v15.2 and GFS v16, respectively, over the Great Plains. Here, HRRR analysis was bilinearly interpolated to GFS grids at 1800 UTC for the region encompassing the large-CAPE reservoir over the Great Plains area, using the coordinates of the lower-left corner (31.24°N, 102.64°W) and the top-right corner (49.16°N, 96.47°W).

Although initiation of the deep convection that is evident in Fig. 2b in the vicinity of the front is of interest, we concentrate here on conditions farther south at the SGP site in north-central Oklahoma (approximate location indicated in Fig. 2). For this we make use of the detailed boundary layer observations at the ARM SGP site in our evaluation of GFS boundary layer forecasts for that location. Sounding and moisture budget analysis over the Northern Great Plains are included in sections S1 and S2 in the online supplemental material.

The SGP site is amid a field of small fair-weather cumulus clouds produced by diurnal heating of the surface on 23 July

2020 (Fig. 2b). Although there are evident spatial variations in this cloud field, there is no indication of active deep convection in the near vicinity. The meteorology of the SGP appears to be dominated by processes operating in the fair-weather boundary layer and not by MCSs or mobile synoptic-scale disturbances, making this case suitable for analyzing the GFS boundary layer performance.

The time series of CAPE bias for GFS was investigated using the TROPoe dataset with high temporal resolution. The TROPoe dataset enables verification of CAPE and of the diurnal cycle of boundary layer structures at finer time scales, in comparison with soundings (Blumberg et al. 2017). SBCAPE measures the instability using air parcels originating from the surface, thus errors in simulated SBCAPE imply error in forecasting meteorological fields near the surface if the forecast values above are correct. Time series of SBCAPE (dashed line) and MLCAPE (solid line) derived from the TROPoe dataset and the two model runs using the two different physics suites at SGP central facility (C1) site are shown in Fig. 4. The daytime means of SBCAPE and MLCAPE derived from observations at the SGP C1 site reached 3734 and 1904 J kg^{-1}, respectively. The CAPE values exhibit a sharp rate of increase at around 0800 LST and reached their maximum at around noon. The two GFS runs overall captured the diurnal pattern; however, GFS v15.2 underestimates SBCAPE and MLCAPE with a daytime ME of −2773 and −1247 J kg^{-1}, respectively. GFS v16 generated SBCAPE and MLCAPE about 16% and 26% lower than GFS v15.2. A notable discrepancy was observed between the difference in SBCAPE and MLCAPE from the observations versus the simulations. This is because the soundings show a notable superadiabatic

surface layer that is less evident (indicating more vertical mixing) in GFS than in the observations. The well-mixed boundary layer in GFS, as illustrated in Fig. 6, could be a contributing factor to the fewer differences between SBCAPE and MLCAPE in the simulations when compared with the observations.

b. Surface meteorology, surface fluxes, and soil properties

Following the synoptic pattern and CAPE analysis in the previous section, this section aims at describing the surface conditions at sites situated within the large CAPE region. Although vertical meteorological profiles (rawinsondes) are available at only one site (SGP C1), the ARM network at SGP surrounding C1 has 9 sites with surface measurements. We make use of surface data from all these sites. Near-surface moisture and temperature directly impact convection formation, especially the magnitude of SBCAPE. The partitioning of surface net radiation to surface turbulent fluxes redistributes heat and moisture into the lower boundary layer. The ARM network at SGP allows for verification of surface meteorological fields and surface energy budgets at several sites, enabling us to narrow down sources of error.

Figure 5 shows the observed and simulated time evolution of near-surface meteorological fields, surface energy fluxes, and soil properties. The mean daytime (0800–1800 LST) 2-m temperature (T_2m) and 2-m dewpoint temperature (Td_2m) averaged over the SGP sites were 303.60 and 296.56 K, respectively. The GFS was warm and dry with an ME of +0.44 and −3.10 K for daytime T_2m and Td_2m, respectively. GFS v16 exhibits slightly larger biases of the T_2m than GFS v15.2. Diurnal variations of 10-m wind speeds are relatively well represented in GFS until 1900 LST. Later, the observed winds decreased while the strong winds lasted longer in both GFS suites.

The mean daytime net radiation (Rn) in GFS v16 was lower than observed and in GFS v15.2 due to its sharp decrease after noon. The earlier decrease of Rn in GFS v16 (Fig. 5c) is related to liquid cloud formation due to boundary layer processes. These clouds are based just above the top of the boundary layer in GFS v16. Analogous cloud formation occurs in GFS v15.2 (Fig. S4 in the online supplemental material) but with less cloud water mixing ratio and thus a smaller liquid water path than GFS v16 (Figs. S5 and S6 in the online supplemental material).

Partitioning of the net incoming radiation flux, Rn, into turbulent surface heat flux depend on soil properties and vegetation. Magnitudes of surface latent (LHF) and sensible heat flux (SHF) drop early in GFS v16 similar to Rn. The daytime SHF and LHF at the SGP sites are shown with large biases of 20.1% and −19.8% in GFS, respectively (Fig. 5f). The daytime Bowen ratio derived from measurement was 0.42. This contrasts sharply with the Bowen ratio of 0.67 and 0.60 for v15.2 and v16, respectively. This indicates that biases in surface turbulent fluxes for this case not only stem from net radiation, but also from soil moisture (bias: −0.01; Fig. 3g) and soil temperature (bias: +0.73 K; Fig. 3h).

For this case, the GFS analysis generated from GFS v16 was utilized to drive the UFS model. In GFS v16, an offline Global Land Data Assimilation System is run for 72 h forced by the NWS Climate Prediction Center observed precipitation product and other forcings from model's own Global Data Assimilation System forecast. Soil moisture and soil temperature at snow free land points from the first guess included in the land initial conditions are replaced with the spinup values (F. Yang 2022, personal communication). However, the current study does not investigate the source of bias in soil moisture and temperature, as it is beyond the scope of this research.

c. Boundary layer structure

CAPE is directly impacted by the boundary layer structure. Biases of simulated CAPE imply discrepancies of temperature and humidity between models and observations. The temporal evolution of the vertical distribution of temperature and moisture at SGP C1 is present in Fig. 6. It reveals that GFS simulated a warmer and drier atmosphere in the low atmospheric layers when compared with the TROPoe dataset. Specifically, the average potential temperature within the lowest 1 km in GFS (305.7 K) exhibits a positive bias of 1 K when compared with the observation (304.7 K). Climatology analysis of CAPE from multidecade ERA-40 data shows a stronger relationship with near-surface humidity than near-surface temperature (Riemann-Campe et al. 2009). The mean observed water vapor mixing ratio q within 1 km AGL during the daytime was 17.4 g kg^{-1}, whereas simulated q exhibited a dry bias with a mean value of 15.4 and 14.8 g kg^{-1} in GFS v15.2 and v16, respectively.

The PBLH determined using the theta-based parcel method (Hennemuth and Lammert 2006) is illustrated in Fig. 6. The PBLH derived from the TROPoe dataset began to grow at 0700 LST and reached a plateau at 1500 LST before it collapsed at 1800 LST. GFS v15.2 PBL grows rapidly until around 1000 LST, then grows more slowly until about 1600 LST after which the PBLH decreases rapidly. In GFS v16, the simulated PBL exhibits a rather steep growth until around 1400 LST and weakens immediately after that. The maximum observed PBL depth during daytime reached 1.48 km, with a slightly higher PBL height observed on July 24 than the prior day. GFS generates a deeper mixed layer with a peak PBLH of 1.77 and 1.89 km in v15.2 and v16, respectively. The stronger PBL mixing in GFS at noon reduces water-vapor loadings at low altitudes. This correlates with our earlier findings that the magnitudes of SHF in GFS v16 were larger and dropped earlier in the afternoon, when compared with both GFS v15.2 and observations. The larger-than-observed SHF could partly explain the higher PBLH in GFS v16. One possible explanation for the slightly larger surface latent heat flux in GFS v16 than v15.2 before afternoon is the potential for enhanced drying within the PBL scheme, which could result in increased evaporation and subsequently higher surface latent heat flux.

To investigate how the individual physical processes impact the boundary layer meteorological profiles, we inspected the tendencies generated from the different physics schemes. The
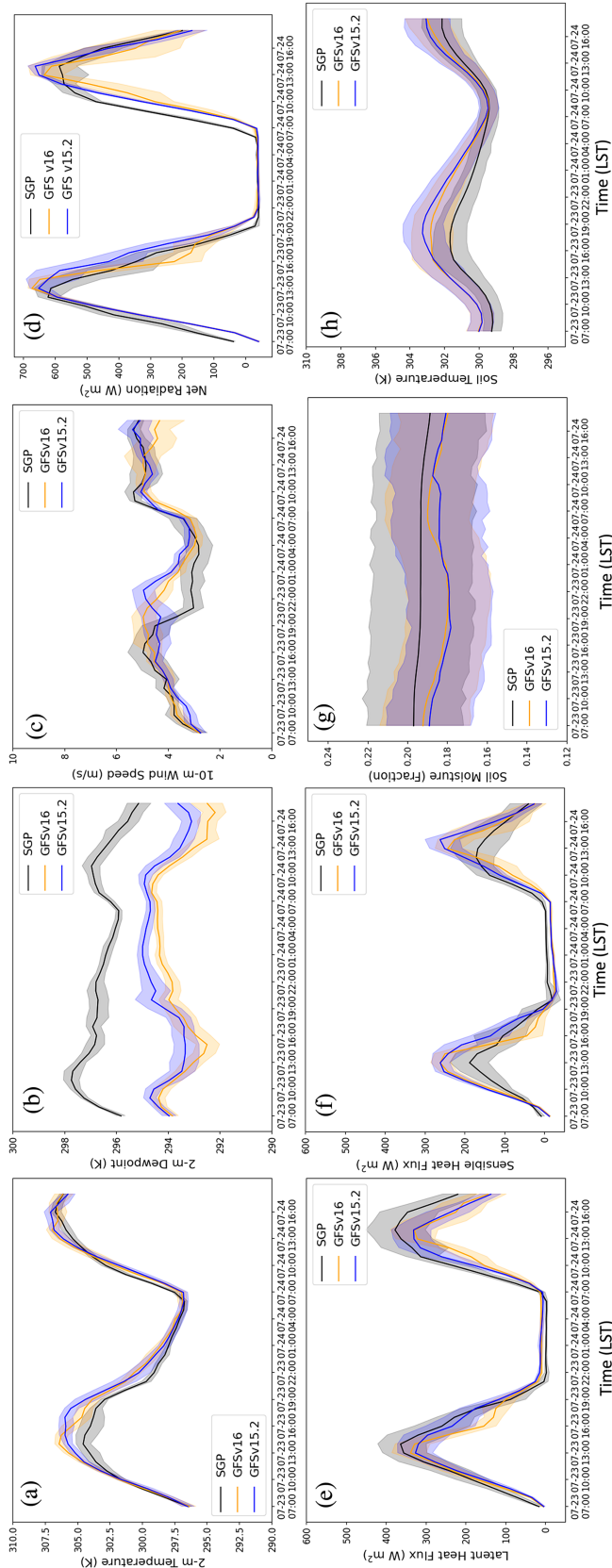
FIG. 5. Simulated and observed time series of (a) 2-m temperature (K), (b) 2-m dewpoint temperature (K), (c) 10-m wind speed (m s^{-1}), (d) net radiation (W m^{-2}), (e) surface latent heat flux (W m^{-2}), (f) surface sensible heat flux (W m^{-2}), (g) volumetric soil moisture (fraction), and (h) soil temperature (K) averaged over the ARM SGP sites with available surface observations (black), GFS v16 (orange), and GFS v15.2 (blue). The observed soil moisture was averaged over two soil depths of 5 and 10 cm to match the first soil layer (0–10 cm below ground) in GFS. Shaded areas represent the range of data values with 95% confidence with the assumption of surface homogeneity.
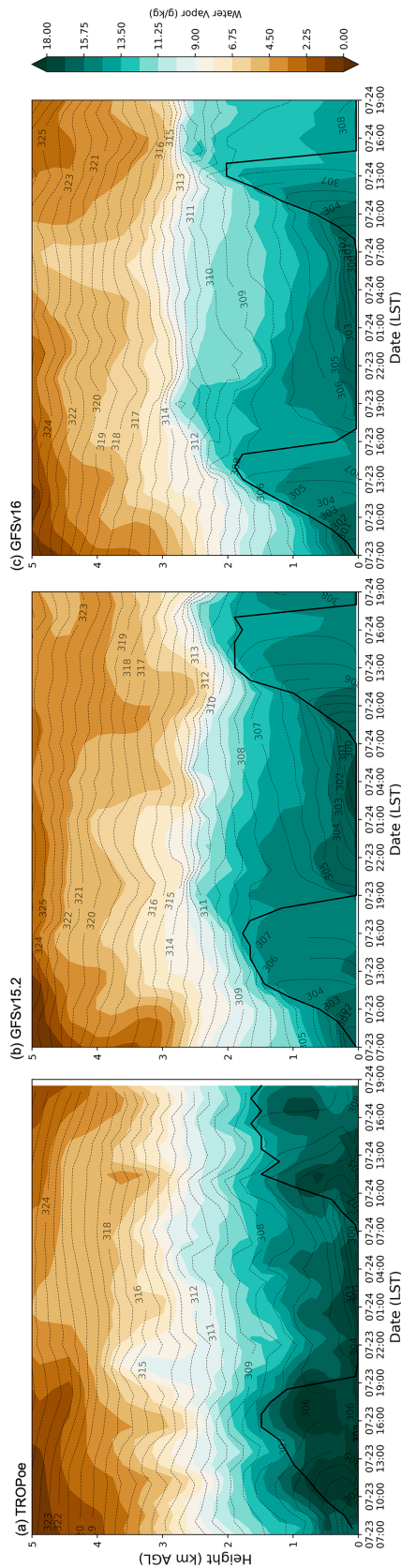
temperature tendencies and moisture tendencies generated by the PBL schemes in GFS v15.2 and GFS v16 are presented in Fig. 7. Results show that the PBL parameterization is the main driver for temperature and moisture changes (figure not shown) in the lowest 150 hPa at the SGP site. The PBL scheme warms and dries the atmosphere within the boundary layer. The buoyant elements rise within the mixed layer and continue rising and become negatively buoyant upon encountering the top of the mixed layer. The air above the mixed layer is both drier and virtually warmer than the mixed-layer air. The turbulent mixing process entrains this drier air with higher virtual potential temperature into the mixed layer, thus the mixed layer becomes virtually warmer and drier. Near the top of the PBL, the detrainment process cools and moistens the free atmosphere, leading to the thin layer of cooling and moistening above the mixed layer. The higher PBLH in GFS v16 is associated with more mixing at the low levels with more drying than in GFS v15.2, thus leading to a drier atmosphere.

d. Advection

To investigate how the simulated large-scale advection fields in GFS impact the moisture bias, we computed the moisture budget based on Eq. (2) using ERA5 and GFS forecasts within a 120 km × 120 km region centered at the SGP C1 site. The reason we use ERA5 as a benchmark is to capture large-scale dynamics without introducing small-scale heterogeneities from high resolution datasets, such as HRRR analysis. The regional average and standard deviation of terms in the moisture budget from ERA5 and GFS v16 during 1400–2200 UTC 23 July 2020 are presented in Fig. 8. Results from GFS v15.2 are not shown due to their similarity to GFS v16.

The regional mean moisture tendencies both exhibit drying tendencies (negative values) within the lowest 90 hPa with very similar values in GFS v16 and ERA5, with GFS v16 simulating a slightly lower magnitude. The residual term is the main contributor to the moisture budget within the lowest 100 hPa, exceeding the impact of advection for this case. This means the low CAPE bias in GFS does not stem from large-scale advection fields in this case. Moisture tendencies are positive above 900 hPa for both the ERA5 and GFS v16, and reduce to nearly zero at around 400 hPa. The drying impact of horizontal advection in GFS v16 is more significant than in ERA5 above 900 hPa. Meanwhile, the moistening from the residual term opposes the drying from horizontal advection. GFS v16 shows a strong peak value of mean moisture tendency at around 750 hPa, which can be attributed mainly to the horizontal advection and residual term. This large residual term could be related to moist physics, such as cloud presence and transport of moisture due to unresolved turbulence. Although the residual terms are both positive between 900 and 400 hPa in the two scenarios, the peak moisture tendency is not present in ERA5. GFS v16 exhibits less drying from vertical advection above 900 hPa with a weaker subsidence, when compared with ERA5. The standard deviations of the moisture-budget terms have similar magnitudes as the regional
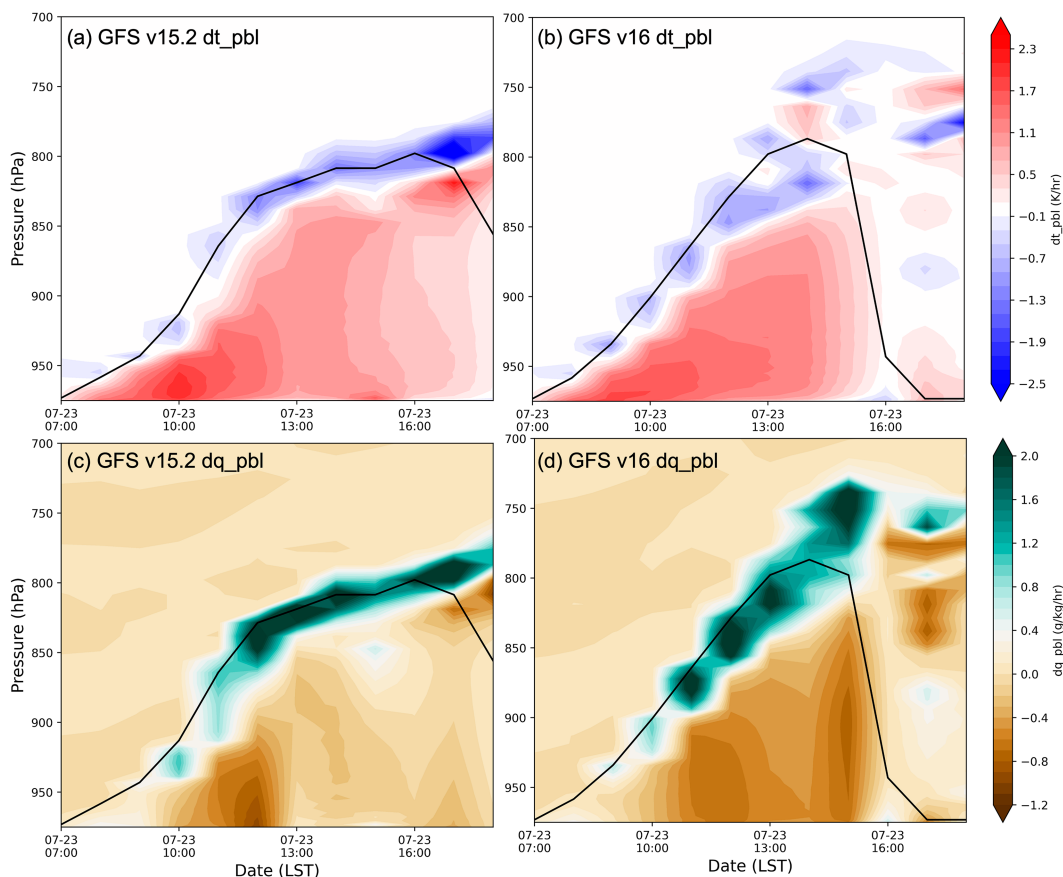

FIG. 6. Time–height plot of potential temperature (K; gray contours) and water vapor mixing ratio (g kg^{-1}; shading) from (a) TROPoe, (b) GFS v15.2, and (c) GFS v16. The PBLHs (thick black solid lines) are calculated using the theta-based parcel method.

FIG. 7. Time–height distribution of (a),(b) temperature tendency (K h^{-1}) and (c),(d) moisture tendency (g kg^{-1} h^{-1}) from PBL schemes in GFS (left) v15.2 and (right) v16. The PBLH from Figs. 6b and 6c, respectively, are superimposed on the tendency plots by the thick black lines.

mean at each level, implying no significant outliers within the domain of interest. Differences in moisture tendency between ERA5 and GFS would have a greater impact on MLCAPE as compared with SBCAPE, because the former is a more integrated measure that accumulates differences over a larger depth. SBCAPE will be affected by the surface value of mixing ratio, and the virtual temperature stratification above the surface. Only through the virtual temperature dependence on water vapor content will the SBCAPE be changed by tendencies above the surface.

4. SCM experiments

In global models, interactions and feedback between the multiscale atmospheric processes and compensating errors complicate the analysis of the results. In SCMs, it is convenient to change one model component at a time to pinpoint factors that impact the simulated phenomena. Initial condition vertical profiles at 0000 UTC 23 July in Fig. 9a suggest that both GFS and the observations have a dry-adiabatic mixed layer, but the GFS shows a deeper, slightly warmer and drier mixed layer. To advance our understanding of how the initial conditions impact the simulated CAPE, we drive the

CCPP SCM with realistic initial conditions using vertical profiles from sounding sites at ARM and initial conditions from GFS IC for a simulation length of 24 h, both with prescribed surface fluxes, using GFS v16 physics suites, referred to as ObsIC and UFSIC, respectively. We also perform another SCM experiment with active Noah LSM, referred to as ObsIC_LSM, to investigate how the PBL interactions with the LSM impact the simulated CAPE. The advection fields for the three SCM runs are both from the GFS v16 global runs, which show similar values in the boundary layer when compared with ERA5 (Fig. 8). Note that since the SCM outputs are sensitive to the advection fields, our focus should be placed on the intercomparisons among the three SCM runs.

Note that the GFS simulated surface heat flux bias at C1 site (figure not shown) is smaller than the ensemble data depicted in Fig. 5. We conducted a SCM run (UFSIC_ObsFlx) using GFS IC but with observed surface fluxes to examine the influence of surface flux bias on derived CAPE. It is worth noting that the GFS simulated surface heat flux bias at C1 site (figure not shown) is smaller than the ensemble data depicted in Fig. 5. The early reduction of GFS simulated surface fluxes in the early afternoon is evident at C1 site, which results in a
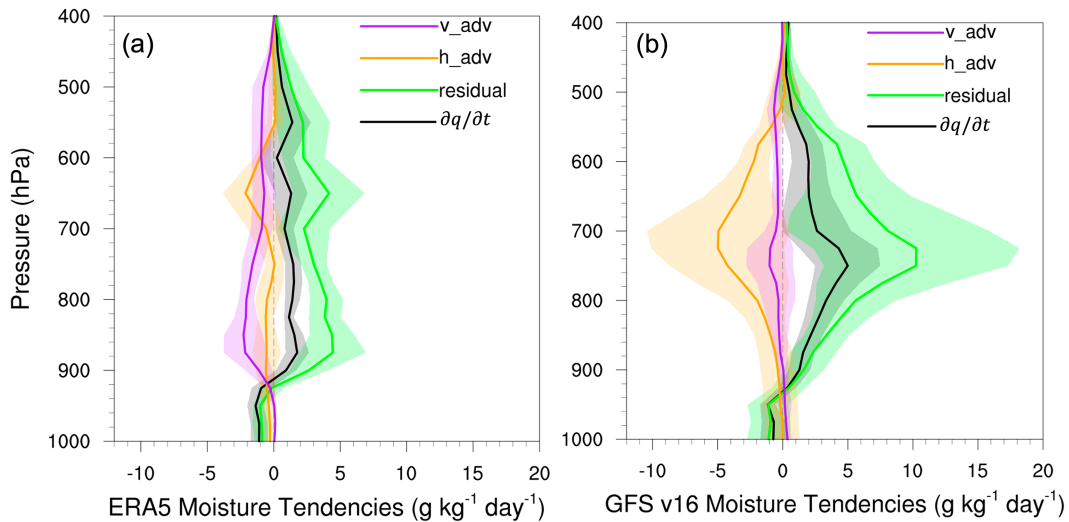
FIG. 8. Regional mean large-scale moisture budget (g kg^{-1} day^{-1}) associated with vertical advection (purple), horizontal advection (orange), residual term (green), and the local moisture change (black) from (a) ERA5 and (b) GFS v16 simulations over a 120 km × 120 km region centered at the SGP C1 site during 1400–2200 UTC 23 Jul 2020. Shaded areas represent ±1 standard deviation. The tendencies are averaged over the daytime first, and standard deviations are then calculated spatially for all grid cells within the 120 km × 120 km domain centered at the SGP C1 site.

smaller CAPE in UFSIC relative to UFSIC_ObsFlx during early afternoon (Fig. S4 in the online supplemental material).

The vertical profiles of temperature and dewpoint from SCM runs (ObsIC, ObsIC_LSM, and UFSIC) at 1500 UTC (0800 LST) 23 July are shown in Fig. 9b. We added the observed sounding profile from TROPoe for reference. At 0800 LST, the surface warms quickly, exhibiting a decoupled state from the upper levels. The SCM runs capture the surface warming with adiabatic temperature profiles. UFSIC simulations exhibit drier atmosphere than ObsIC simulations

in the lowest 100 hPa, which can be attributed to the unrealistic initial conditions shown in Fig. 9a. Air temperature in the ObsIC_LSM are colder than ObsIC in the lowest 100 hPa.

The MLCAPE derived from the four SCM runs, as well as TROPoe are shown in Fig. 10. Although the derived CAPE is not exactly the same between the SCM and the global UFS runs, the low bias of CAPE seen in the global runs is also present in the SCM runs when using the UFS forcing. Note that focus should be placed on the intercomparison among the
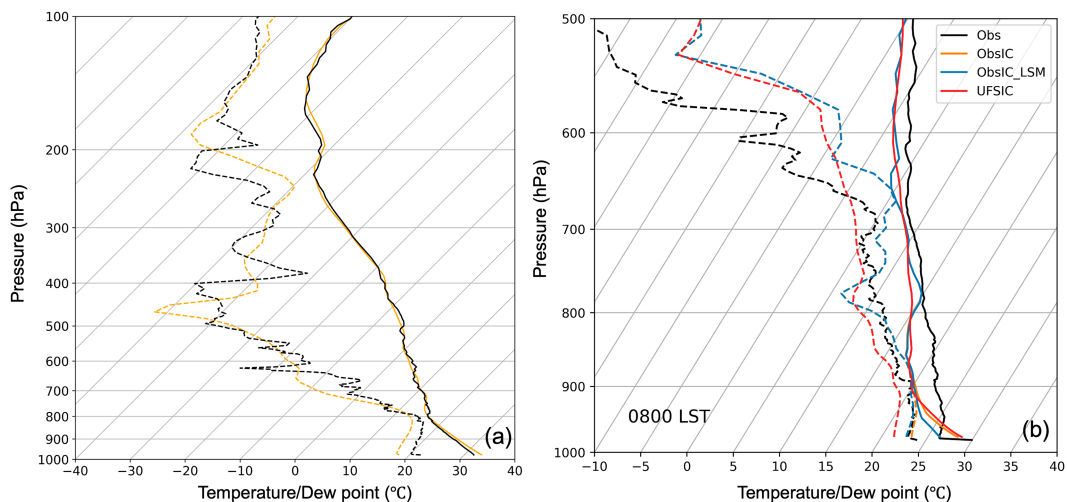

FIG. 9. (a) Initial vertical profiles of temperature (solid lines) and dewpoint (dashed lines) at the SGP central facility (C1) site from observation (black) and GFS (orange) at 0000 UTC 23 Jul. (b) Skew T–log p plots of measured soundings (Obs; black) and SCM runs from ObsIC (orange), ObsIC_LSM (blue), and UFSIC (red) at 1500 UTC 23 Jul.

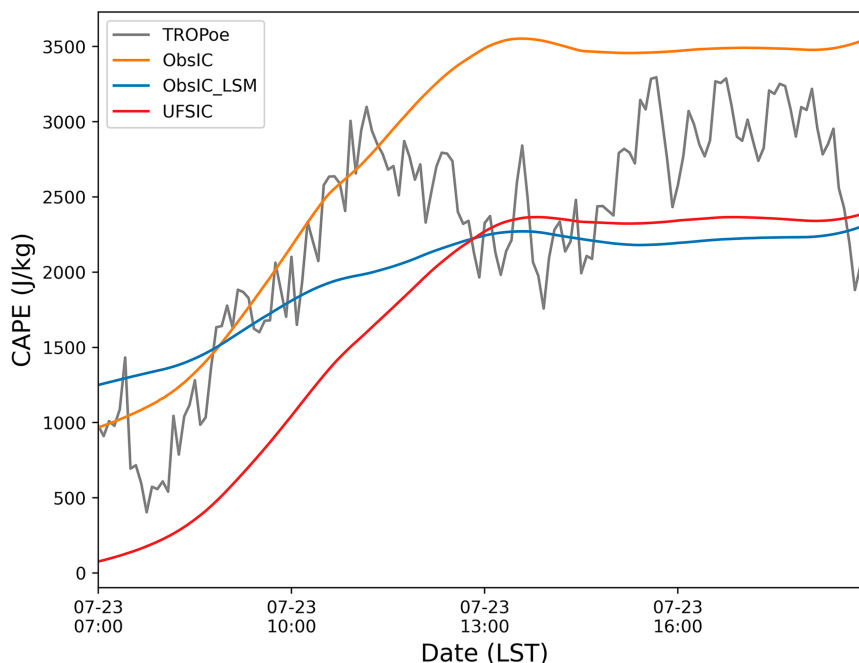Fig. 10. MLCAPE (J kg^{-1}) derived from TROPoe and SCM runs including ObsIC (orange), ObsIC_LSM (blue), and UFSIC (red).

SCM runs, since the advective tendencies used to drive CCPP SCM are estimated from UFS state variables, which can surely be different from the real world. With realistic initial conditions (ObsIC), the resulting MLCAPE increases by around 1000 J kg^{-1} when compared with UFSIC. However, it is important to acknowledge that it can be challenging to distinguish between model background errors and DA algorithms in the process of generating initial conditions.

The SCM run with active Noah LSM, ObsIC_LSM simulates smaller CAPE than the SCM run with prescribed surface fluxes (ObsIC). One hypothesis is that the GFS v16 PBL scheme tends to overmix, leading to excessive drying of the atmosphere near the surface. This, in turn, causes excessive latent heat flux, drying the soil during the forecast. This overmixing and drying bias is not large, so it does not show up early in the forecast. However, it is systematic and accumulates during the forecast unless there is enough synoptic activity (specifically, quasigeostrophic forcing) forecast by the model to produce model precipitation, and so overcome the excessive drying of the soil by the excessive latent heat flux. In summer, with weak synoptic systems, is when this particular bias late in the forecast tends to show up. These results highlight the utility of SCM for model error attribution within the HTF.

5. Summary

Prediction of CAPE requires realistic simulations of vertical temperature and moisture profiles, especially at the near-surface levels. In this study, we aimed to identify the factors contributing to one of the GFS main forecast challenges,

underpredicted CAPE. To this end, we investigated a summer CAPE case (23–24 July 2020) from the case catalog on the UFS Case Studies Platform that documents forecast challenges of NOAA's global operational weather forecast model. The case is characterized by large CAPE over the Great Plains, with surface-based CAPE and mixed-layer CAPE reaching 5000 and 3000 J kg^{-1}, respectively. The large CAPE over the Northern Great Plains was associated with a warm frontal passage. No significant mesoscale convective systems were impacting the Southern Great Plains. In this study, we focused on the SGP region due to the simple synoptic situation and the detailed observations within the ARM network.

Results show that the low CAPE bias in GFS v16 at the ARM SGP central facility site can be attributed to the simulated drier air at low levels, when compared with the TROPoe dataset. The moisture-budget analysis showed that the advection fields are of the same order of magnitude as ERA5 within the PBL, hence advection is not the cause of the drier air simulated in GFS v16 in the lowest 90 hPa. The residual term, which represents subgrid physical processes, is the main contributor that impacts the low-level moisture tendencies. The sharp diurnal evolution of PBL height in GFS v16 is accompanied by overmixing within the boundary layer at noon, leading to reduced water-vapor loadings at low levels. Excessive sensible heat flux contributes to the overshooting of the PBL height (PBLH) in GFS v16. The overestimated surface heat fluxes are also due to an incorrect surface energy partitioning associated with the slightly warmer and drier soil surface in the model. The early decrease of PBLH from GFS v16 in the afternoon is associated with boundary layer cloud formation,

which leads to premature drop of surface net radiation and subsequently surface turbulent fluxes and moisture near the surface.

The single column model (SCM) is a useful tool for model error attribution within the Hierarchical Testing Framework. However, SCM results are sensitive to the advection tendencies that are used to drive the SCM. Thus, analysis is focused on the intercomparison among the SCM sensitivity runs. Three SCM experiments (ObsIC, ObsIC_LSM, and UFSIC) were conducted to investigate the impact of initial conditions and land-atmosphere interactions on the simulated CAPE. With realistic initial vertical meteorological profiles to drive the SCM runs, ObsIC run can produce a larger CAPE than the UFSIC run. This highlights the importance of optimizing the atmospheric and land initial conditions through a more accurate background state (first-guess forecast) and improved data assimilation in future GFS model development. The SCM experiment with an active LSM model simulates lower CAPE than SCM runs with prescribed surface fluxes. One hypothesis is that overmixing in the GFS v16 leads to excessive drying of the atmosphere near the surface, which in turn cause excessive evaporation and thus drying the soil. This justifies further investigation and additional development on the PBL scheme used in GFS.

It is important to note that there is continuous ongoing development of the GFS toward GFS v17. For example, efforts such as improving the surface coupling (Barlage 2022) have been made after GFS v16 became operational. However, our investigation is specific to the operational implementation of GFS v16. Interested readers are referred to the authoritative UFS Weather Model GitHub repository to retrieve the most updated developmental codes. A limitation of this study is that it only examines the CAPE evolution in a single case in the summer over the SGP site, and thus the findings are only applicable to this case study. Additional warm-season cases would help to strengthen the generalizability of our conclusions.

Data availability statement. The initial condition datasets for this case are available at the UFS Case Studies Platform (https://ufs-case-studies.readthedocs.io/en/develop/). The codes for GFS v15.2 and GFS v16 are available online (https://github.com/ufs-community/ufs-weather-model).

REFERENCES

Agard, V., and K. Emanuel, 2017: Clausius–Clapeyron scaling of peak CAPE in continental convective storm environments. *J. Atmos. Sci.*, **74**, 3043–3054, https://doi.org/10.1175/JAS-D-16-0352.1.

Barlage, M., 2022: Expanding community UFS land model development through advancing land component and land data assimilation capabilities. NOAA/Environmental Modeling Center, 20 pp., https://epic.noaa.gov/wp-content/uploads/2022/07/2.-CC1-Barlage.pdf.

Bechtold, P., N. Semane, P. Lopez, J.-P. Chaboureau, A. Beljaars, and N. Bormann, 2014: Representing equilibrium and nonequilibrium convection in large-scale models. *J. Atmos. Sci.*, **71**, 734–753, https://doi.org/10.1175/JAS-D-13-0163.1.

Bellenger, H., K. Yoneyama, M. Katsumata, T. Nishizawa, K. Yasunaga, and R. Shirooka, 2015: Observation of moisture tendencies related to shallow convection. *J. Atmos. Sci.*, **72**, 641–659, https://doi.org/10.1175/JAS-D-14-0042.1.

Blanchard, D. O., 1998: Assessing the vertical distribution of convective available potential energy. *Wea. Forecasting*, **13**, 870–877, https://doi.org/10.1175/1520-0434(1998)013<0870:ATVDOC>2.0.CO;2.

Blumberg, W. G., T. J. Wagner, D. D. Turner, and J. Correia, 2017: Quantifying the accuracy and uncertainty of diurnal thermodynamic profiles and convection indices derived from the Atmospheric Emitted Radiance Interferometer. *J. Appl. Meteor. Climatol.*, **56**, 2747–2766, https://doi.org/10.1175/JAMC-D-17-0036.1.

Bonin, T. A., W. G. Blumberg, P. M. Klein, and P. B. Chilson, 2015: Thermodynamic and turbulence characteristics of the Southern Great Plains nocturnal boundary layer under differing turbulent regimes. *Bound.-Layer Meteor.*, **157**, 401–420, https://doi.org/10.1007/s10546-015-0072-2.

Brooks, H. E., A. R. Anderson, K. Riemann, I. Ebbers, and H. Flachs, 2007: Climatological aspects of convective parameters from the NCAR/NCEP reanalysis. *Atmos. Res.*, **83**, 294–305, https://doi.org/10.1016/j.atmosres.2005.08.005.

Chipilski, H. G., X. Wang, D. B. Parsons, A. Johnson, and S. K. Degelia, 2022: The value of assimilating different ground-based profiling networks on the forecasts of bore-generating nocturnal convection. *Mon. Wea. Rev.*, **150**, 1273–1292, https://doi.org/10.1175/MWR-D-21-0193.1.

Corfidi, S. F., S. J. Corfidi, and D. M. Schultz, 2008: Elevated convection and castellanus: Ambiguities, significance, and questions. *Wea. Forecasting*, **23**, 1280–1303, https://doi.org/10.1175/2008WAF2222118.1.

Degelia, S. K., X. Wang, D. J. Stensrud, and D. D. Turner, 2020: Systematic evaluation of the impact of assimilating a network of ground-based remote sensing profilers for forecasts of nocturnal convection initiation during PECAN. *Mon. Wea. Rev.*, **148**, 4703–4728, https://doi.org/10.1175/MWR-D-20-0118.1.

Dowell, D. C., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. *Wea. Forecasting*, **37**, 1371–1395, https://doi.org/10.1175/WAF-D-21-0151.1.

Ek, M., C. DeLuca, L. Bernardet, T. Jensen, M. Vertenstein, A. Chawla, J. Kinter, and R. Rood, 2019: Hierarchical System Development for the UFS. Unified Forecast System, https://ufscommunity.org/articles/hierarchical-system-development-for-the-ufs/.

Emanuel, K. A., 1994: *Atmospheric Convection*. Oxford University Press, 592 pp.

English, J. M., D. D. Turner, T. I. Alcott, W. R. Moninger, J. L. Bytheway, R. Cifelli, and M. Marquis, 2021: Evaluating operational and experimental HRRR model forecasts of atmospheric river events in California. *Wea. Forecasting*, **36**, 1925–1944, https://doi.org/10.1175/WAF-D-21-0081.1.

Firl, G., D. Swales, L. Carson, M. Harrold, L. Bernardet, and D. Heinzeller, 2022: Common Community Physics Package Single Column Model (SCM), v6.0.0: User and technical guide. Development Testbed Center, 45 pp., https://dtcenter.org/sites/default/files/paragraph/scm-ccpp-guide-v6-0-0.pdf.

Fovell, R. G., and A. Gallagher, 2020: Boundary layer and surface verification of the High-Resolution Rapid Refresh, version 3. *Wea. Forecasting*, **35**, 2255–2278, https://doi.org/10.1175/WAF-D-20-0101.1.

Friedman, K., and W. Kolczynski, 2022: Global Superstructure/Workflow supporting the Global Forecast System (GFS). GitHub, 23 August 2022, https://github.com/NOAA-EMC/global-workflow.

Grabowski, W. W., and M. W. Moncrieff, 2004: Moisture–convection feedback in the tropics. *Quart. J. Roy. Meteor. Soc.*, **130**, 3081–3104, https://doi.org/10.1256/qj.03.135.

Guo, J., and Coauthors, 2021: Investigation of near-global daytime boundary layer height using high-resolution radiosondes: First results and comparison with ERA5, MERRA-2, JRA-55, and NCEP-2 reanalyses. *Atmos. Chem. Phys.*, **21**, 17 079–17 097, https://doi.org/10.5194/acp-21-17079-2021.

Han, J., and C. S. Bretherton, 2019: TKE-based moist eddy-diffusivity mass-flux (EDMF) parameterization for vertical turbulent mixing. *Wea. Forecasting*, **34**, 869–886, https://doi.org/10.1175/WAF-D-18-0146.1.

——, M. L. Witek, J. Teixeira, R. Sun, H.-L. Pan, J. K. Fletcher, and C. S. Bretherton, 2016: Implementation in the NCEP GFS of a hybrid eddy-diffusivity mass-flux (EDMF) boundary layer parameterization with dissipative heating and modified stable boundary layer mixing. *Wea. Forecasting*, **31**, 341–352, https://doi.org/10.1175/WAF-D-15-0053.1.

——, W. Wang, Y. C. Kwon, S.-Y. Hong, V. Tallapragada, and F. Yang, 2017: Updates in the NCEP GFS cumulus convection schemes with scale and aerosol awareness. *Wea. Forecasting*, **32**, 2005–2017, https://doi.org/10.1175/WAF-D-17-0046.1.

Heinzeller, D., L. Bernardet, G. Firl, M. Zhang, X. Sun, and M. Ek, 2023: The Common Community Physics Package (CCPP) Framework v6. *Geosci. Model Dev.*, **16**, 2235–2259, https://doi.org/10.5194/gmd-16-2235-2023.

Hennemuth, B., and A. Lammert, 2006: Determination of the atmospheric boundary layer height from radiosonde and lidar backscatter. *Bound.-Layer Meteor.*, **120**, 181–200, https://doi.org/10.1007/s10546-005-9035-3.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Holdridge, D., and J. Kyrouac, 1993: ARM: ARM-standard meteorological instrumentation at surface. U.S. Department of Energy, 16 September 2022, https://doi.org/10.5439/1025220.

Iacono, M. J., E. J. Mlawer, S. A. Clough, and J.-J. Morcrette, 2000: Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR Community Climate Model, CCM3. *J. Geophys. Res.*, **105**, 14 873–14 890, https://doi.org/10.1029/2000JD900091.

Jacobs, N. A., 2021: Open innovation and the case for community model development. *Bull. Amer. Meteor. Soc.*, **102**, E2002–E2011, https://doi.org/10.1175/BAMS-D-21-0030.1.

James, E. P., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part II: Forecast performance. *Wea. Forecasting*, **37**, 1397–1417, https://doi.org/10.1175/WAF-D-21-0130.1.

Knuteson, R. O., and Coauthors, 2004: Atmospheric Emitted Radiance Interferometer. Part I: Instrument design. *J. Atmos. Oceanic Technol.*, **21**, 1763–1776, https://doi.org/10.1175/JTECH-1662.1.

Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092, https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2.

Long, P., 1984: GUST: A General Unified Similarity Theory for the calculation of turbulent fluxes in the numerical weather prediction models for unstable conditions. NCEP Office Note 302, NOAA, 30 pp., https://repository.library.noaa.gov/view/noaa/11507.

Mayer, J., M. Mayer, and L. Haimberger, 2019: Improved estimates of surface energy fluxes derived from CERES [radiation] and the ERA5 energy budget. *Geophysical Research Abstracts*, Vol. 21, Abstract 10505, https://meetingorganizer.copernicus.org/EGU2019/EGU2019-10505.pdf.

——, ——, and ——, 2021: Consistency and homogeneity of atmospheric energy, moisture, and mass budgets in ERA5. *J. Climate*, **34**, 3955–3974, https://doi.org/10.1175/JCLI-D-20-0676.1.

Mensch, C. A., 2021: Impacts of vertical grid resolution on the representation of a convective environment. M.S. thesis, The University of North Dakota, 141 pp.

Mitchell, K., and Coauthors, 2005: The Community Noah Land-Surface Model (LSM). User's guide public release version 2.7.1. NOAA, 26 pp., https://ral.ucar.edu/document-or-file/noah-lsm-users-guide.

Naakka, T., T. Nygård, and T. Vihma, 2021: Air moisture climatology and related physical processes in the Antarctic on the basis of ERA5 reanalysis. *J. Climate*, **34**, 4463–4480, https://doi.org/10.1175/JCLI-D-20-0798.1.

Pinto, J. O., J. A. Grim, and M. Steiner, 2015: Assessment of the High-Resolution Rapid Refresh Model's ability to predict mesoscale convective systems using object-based evaluation. *Wea. Forecasting*, **30**, 892–913, https://doi.org/10.1175/WAF-D-14-00118.1.

Rabier, F., 2005: Overview of global data assimilation developments in numerical weather-prediction centres. *Quart. J. Roy. Meteor. Soc.*, **131**, 3215–3233, https://doi.org/10.1256/qj.05.129.

Ren, P., D. Kim, M.-S. Ahn, D. Kang, and H.-L. Ren, 2021: Intercomparison of MJO column moist static energy and water vapor budget among six modern reanalysis products. *J. Climate*, **34**, 2977–3001, https://doi.org/10.1175/JCLI-D-20-0653.1.

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.

Rochette, S. M., J. T. Moore, and P. S. Market, 1999: The importance of parcel choice in elevated CAPE computations. *Natl. Wea. Dig.*, **23**, 20–32.

Sherburn, K. D., and M. D. Parker, 2014: Climatology and ingredients of significant severe convection in high-shear, low-CAPE environments. *Wea. Forecasting*, **29**, 854–877, https://doi.org/10.1175/WAF-D-13-00041.1.

Sims, J., D. Koch, H. Tolman, and D. Achuthavarier, 2021: UFS forecasters workshops 2020 & 2021. Bulletin of the UFS Community, April–June 2021, 13–14 (https://ufscommunity.org/wp-content/uploads/2021/07/Bulletin_UFS_Community_April-June_2021_Full.pdf), https://www.weather.gov/media/sti/Final%20Consolidated%20Forecasters%20requests%20April%202021.pdf.

Sisterson, D. L., R. A. Peppler, T. S. Cress, P. J. Lamb, and D. D. Turner, 2016: The ARM Southern Great Plains (SGP) site. *The Atmospheric Radiation Measurement (ARM) Program: The First 20 Years*, Meteor. Monogr., No. 57, Amer. Meteor. Soc., https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0004.1.

Sun, X., D. Heinzeller, L. Bernardet, L. Pan, and J. M. Brown, 2021: Case studies that exemplify known biases of the Unified Forecast System (UFS) weather model. *First UFS User's Workshop*, online, UFS, 11 pp., https://dtcenter.org/sites/default/files/events/2020/3-xia-sun.pdf.

Tallapragada, V., 2020: Unified Forecast System development and operational implementation plans at NCEP/EMC. *10th Conf. on Transition of Research to Operations*, Boston, MA, Amer. Meteor. Soc., 5A.2, https://ams.confex.com/ams/2020Annual/webprogram/Paper362787.html.

Tarek, M., F. P. Brissette, and R. Arsenault, 2020: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrol. Earth Syst. Sci.*, **24**, 2527–2544, https://doi.org/10.5194/hess-24-2527-2020.

Trolliet, M., J. P. Walawender, B. Bourlès, A. Boilley, J. Trentmann, P. Blanc, M. Lefèvre, and L. Wald, 2018: Downwelling surface solar irradiance in the tropical Atlantic Ocean: A comparison of re-analyses and satellite-derived data sets to PIRATA measurements. *Ocean Sci.*, **14**, 1021–1056, https://doi.org/10.5194/os-14-1021-2018.

Turner, D. D., and U. Löhnert, 2014: Information content and uncertainties in thermodynamic profiles and liquid cloud properties retrieved from the ground-based Atmospheric Emitted Radiance Interferometer (AERI). *J. Appl. Meteor. Climatol.*, **53**, 752–771, https://doi.org/10.1175/JAMC-D-13-0126.1.

——, and W. G. Blumberg, 2019: Improvements to the AERIoe Thermodynamic Profile Retrieval Algorithm. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **12**, 1339–1354, https://doi.org/10.1109/JSTARS.2018.2874968.

——, and Coauthors, 2020: A verification approach used in developing the Rapid Refresh and other numerical weather prediction models. *J. Oper. Meteor.*, **8**, 39–53, https://doi.org/10.15191/nwajom.2020.0803.

Wagner, T. J., P. M. Klein, and D. D. Turner, 2019: A new generation of ground-based mobile platforms for active and passive profiling of the boundary layer. *Bull. Amer. Meteor. Soc.*, **100**, 137–153, https://doi.org/10.1175/BAMS-D-17-0165.1.

Williams, E., and N. Renno, 1993: An analysis of the conditional instability of the tropical atmosphere. *Mon. Wea. Rev.*, **121**, 21–36, https://doi.org/10.1175/1520-0493(1993)121<0021:AAOTCI>2.0.CO;2.

Xie, S., and M. Zhang, 2000: Impact of the convection triggering function on single-column model simulations. *J. Geophys. Res.*, **105**, 14 983–14 996, https://doi.org/10.1029/2000JD900170.

Yanai, M., S. Esbensen, and J.-H. Chu, 1973: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *J. Atmos. Sci.*, **30**, 611–627, https://doi.org/10.1175/1520-0469(1973)030<0611:DOBPOT>2.0.CO;2.

Yang, F., 2020: Development and evaluation of NCEP's Global Forecast System version 16 (GFSv16). Unified Forecast System Webinar Series, 61 pp., https://ufscommunity.org/wp-content/uploads/2020/10/UFS_Webnair_GFSv16_20201022_Fanglin Yang.pdf.

Zhang, M., R. C. J. Somerville, and S. Xie, 2016: The SCM concept and creation of ARM forcing datasets. *The Atmospheric Radiation Measurement (ARM) Program: The First 20 Years*, Meteor. Monogr., No. 57, Amer. Meteor. Soc., https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0040.1.