Diagnosing Near-Surface Model Errors with Candidate Physics Parameterization Schemes for the Multiphysics Rapid Refresh Forecast System (RRFS) Ensemble during Winter over the Northeastern United States and Southern Great Plains

XIAO-MING HU,[a,b] JUN PARK,[a] TIMOTHY SUPINIE,[a] NATHAN A. SNOOK,[a,b] MING XUE,[a,b] KEITH A. BREWSTER,[a,b] JERALD BROTZGE,[c] AND JACOB R. CARLEY[d]

[a] *Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma*
[b] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*
[c] *New York State Mesonet, University at Albany, State University of New York, Albany, New York*
[d] *NOAA/NWS/NCEP/Environmental Modeling Center, NOAA National Centers for Environmental Prediction, College Park, Maryland*

(Manuscript received 24 March 2022, in final form 6 September 2022)

ABSTRACT: During the winter of 2020/21 an ensemble of FV3-LAM forecasts was produced over the contiguous United States for the Winter Weather Experiment using five physics suites. These forecasts are evaluated with the goal of optimizing physics parameterizations within the future operational Rapid Refresh Forecast System (RRFS) in the Unified Forecast System (UFS) realm and for selecting suitable physics suites for a multiphysics RRFS ensemble. The five physics suites have different combinations of land surface models (LSMs), planetary boundary layer (PBL) parameterizations, and surface layer schemes, chosen from those used in current and possible future operational systems and likely to be supported in the operational UFS. Full-season evaluation reveals a persistent near-surface cold bias in the U.S. Northeast from one suite and a nighttime warm bias in the southern Great Plains in another suite, while other suites have smaller biases. A representative case is chosen to diagnose the cause for each of these biases using sensitivity simulations with different physics combinations or modified parameters and verified with additional mesonet observations. The cold bias in the Northeast is attributed to aspects of the Noah-MP LSM over snow cover, where Noah-MP simulates lower soil water content, and thus lower thermal conductivity than other LSMs, leading to less upward ground heat flux during nighttime and consequently lower surface temperature. The nighttime warm bias found in the southern Great Plains is attributed to overestimation of vertical mixing in the *K*-profile-based eddy-diffusivity mass-flux (K-EDMF) PBL scheme and insufficient land–atmospheric coupling from the GFS surface layer scheme over short vegetation. A few key parameters driving these systematic biases are identified.

KEYWORDS: Numerical weather prediction/forecasting; Short-range prediction; Ensembles; Model errors; Model evaluation/performance

1. Introduction

As part of the implementation plan for the Unified Forecast System (UFS), the U.S. National Weather Service is in the process of developing and implementing a regional convection-allowing ensemble numerical weather prediction (NWP) system to be known as the Rapid Refresh Forecast System (RRFS; Alexander and Carley 2021) using the limited area model (LAM; Black et al. 2021) version of the finite-volume cubed-sphere (FV3) dynamical core (FV3-LAM; Lin 2004; Putman and Lin 2007; Harris et al. 2021). The Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma ran a set of experimental FV3-LAM forecasts covering the contiguous United States (CONUS) at ~3-km grid spacing from November 2020 to March 2021 for use in the 11th Hydrometeorology Testbed (HMT) Winter Weather Experiment (WWE; Harnos et al. 2021). One of the goals of these forecast experiments is to examine the performance of individual physics parameterization schemes and their systematic model errors to help inform the design of the RRFS's ensemble forecast system. Five configurations with various physics schemes (Table 1) were run one or more times weekly throughout the winter. The five configurations were chosen to be similar to configurations used in existing National Oceanic and Atmospheric Administration (NOAA) operational or experimental forecasting systems and are considered suites likely to be supported in future NOAA operations.

Limited area NWP model forecast errors can come from model errors and initial and lateral boundary condition errors. Model error can be structural or parametric. Structural error refers to model equations that are inconsistent with the true laws governing the system, while parametric error refers to inaccurate parameters used within model equations. Structural errors can be converted into parametric errors when the functional equations are generalized (Aksoy et al. 2006; Hansen and Penland 2007; Tong and Xue 2008; Hu et al. 2010b; Jung et al. 2010; Liu et al. 2021; Nystrom et al. 2021). Given the inevitable uncertainties/errors associated with various model treatments, including physics parameterization schemes, NWP forecasts may systematically deviate from the real atmosphere (Hansen 2002). In such an imperfect model context, one strives to construct an ensemble of forecast models that are diverse but with the ensemble mean still close to the underlying physics that govern the real atmosphere. While data assimilation aims to minimize model initial condition errors (e.g., Banos et al.

Corresponding author: Xiao-Ming Hu, xhu@ou.edu

DOI: 10.1175/MWR-D-22-0085.1

TABLE 1. Five FV3 physics suites for Rapid Refresh Forecast System (RRFS) ensemble members. All suites use the RRTMG radiation scheme.

Experiment name	Suite characteristic	Microphysics	PBL	Surface layer	LSM
CNTL	RRFS-control-like	Thompson (Thompson and Eidhammer 2014)	MYNN (Olson et al. 2019a,b)	MYNN (Nakanishi and Niino 2009)	Noah (Chen and Zhang 2009)
LSM1	HRRR-like (Benjamin et al. 2016; Dowell et al. 2022)	Thompson	MYNN	MYNN	Rapid Update Cycle (RUC) (Smirnova et al. 2000, 2016)
LSM2	Future GFS-like	Thompson	TKE-EDMF (Han and Bretherton 2019)	GFS (Zheng et al. 2012)	Noah-MP[a] (Niu et al. 2011)
MP1	WoFS-like	NSSL (Mansell 2010; Mansell and Ziegler 2013)	MYNN	MYNN	Noah
MP2	HWRF-like (Biswas et al. 2018) HAFS-like (Hazelton et al. 2021)	Ferrier-Aligo (Aligo et al. 2018)	K-EDMF (Han et al. 2016)	GFS (Zheng et al. 2012)	Noah

[a] The surface layer drag coefficient is calculated using the Monin-Obukhov scheme (iopt_sfc=1).

2022), model errors are often addressed by improving or calibrating model physics parameterization schemes in systematic retrospective simulations using observations or high-resolution benchmarks for verification (Nystrom et al. 2021).

In this study, FV3-LAM forecasts using five different physics suites, run by CAPS during the 11th HMT WWE, are first evaluated for their accuracy in predicting near-surface variables for the entire winter of 2020/21 to identify the systematic biases over different regions of the CONUS. Then, sensitivity simulations with altered physics combinations or parameter settings (summarized in Tables 2 and 3) are conducted for a representative case to diagnose the causes of near-surface biases and attribute the systematic biases to a few key parametric inaccuracies. Results from this study can help inform design of the RRFS ensemble and improve and better calibrate physics parameterization schemes in FV3-LAM.

The rest of this paper is organized as follows: In section 2, the numerical experiment design, FV3-LAM model configurations, and evaluation data are described. In section 3, FV3-LAM forecasts are evaluated by comparing them to the operational Unrestricted Mesoscale Analysis (URMA),

TABLE 2. Model configuration for 15 RRFS members. Letters L, M and B refer to land surface model, microphysics, and PBL schemes, respectively. Numbers following the three letters are the option number for these schemes, which are marked in parentheses in columns 2, 3, and 5. These configurations are based on the five original physics suites but altering one physics parameterization at a time to isolate model bias/error sources within the five suites.

Experiment	Microphysics	PBL	Surface layer	LSM	Operational analog
M0B0L0 (CNTL)	Thompson (0)	MYNN (0)	MYNN	Noah (0)	RRFS control
M0B1L0	Thompson (0)	Shin–Hong (1)	GFS	Noah (0)	
M0B2L0	Thompson (0)	TKE-EDMF (2)	GFS	Noah (0)	
M0B2L1 (LSM2)	Thompson (0)	TKE-EDMF (2)	GFS	Noah-MP (1)	Future GFS
M0B0L1	Thompson (0)	MYNN (0)	MYNN	Noah-MP (1)	
M0B0L2 (LSM1)	Thompson (0)	MYNN (0)	MYNN	RUC (2)	HRRR
M1B0L0 (MP1)	NSSL (1)	MYNN (0)	MYNN	Noah (0)	WoFS
M1B1L0	NSSL (1)	Shin–Hong (1)	GFS	Noah (0)	
M1B2L0	NSSL (1)	TKE-EDMF (2)	GFS	Noah (0)	
M1B0L1	NSSL (1)	MYNN (0)	MYNN	Noah-MP (1)	
M2B0L0	Morrison–Gettelman (Morrison and Gettelman 2008) (2)	MYNN (0)	MYNN	Noah (0)	
M2B1L0	Morrison–Gettelman (2)	Shin–Hong (1)	GFS	Noah (0)	
M2B2L0	Morrison–Gettelman (2)	TKE-EDMF (2)	GFS	Noah (0)	
M2B0L1	Morrison–Gettelman (2)	MYNN (0)	MYNN	Noah-MP (1)	
M3B3L0 (MP2)	Ferrier–Aligo (3)	K-EDMF (3)	GFS	Noah (0)	HRWF

TABLE 3. Model configuration for sensitivity simulations modifying parameters in the Noah-MP land surface model used by the LSM2 suite and K-EDMF PBL, and GFS surface layer scheme used by the MP2 suite. Here, p is an exponent in the polynomial function determining vertical mixing strength, α is parameter used to calculate critical bulk Richardson number (Ri), z_o is surface aerodynamic roughness length over grass, and C_{zil} is a parameter to relate roughness length of heat to roughness length of momentum.

Base configuration	PBL	Surface layer	LSM	Experiment name	Changed parameters
LSM2	TKE-EDMF (Han and Bretherton 2019)	GFS (Zheng et al. 2012)	NoaH-MP (Niu et al. 2011)	LSM2	Snow emissivity = 1.0 (default)
				LSM2_SNOW_EMIS.95	Snow emissivity = 0.95
				LSM2_SNOW_EMIS.9	Snow emissivity = 0.9
MP2	K-EDMF (Han et al. 2016)	GFS (Zheng et al. 2012)	Noah (Chen and Zhang 2009)	MP2	p = 2, α = 0.16, and z_o = 0.07 (default)
				MP2_p3	p = 3
				MP2_p3_α.1	p = 3 and α = 0.1
				MP2_p3_α.1_z_0.15	p = 3; α = 0.1; z_o = 0.15
				MP2_p3_α.1_C_{zil}.1	p = 3; α = 0.1; C_{zil} = 0.1

followed by detailed diagnosis of model biases in terms of near-surface variables over the northeastern United States and the southern Great Plains. Finally, section 4 contains a summary and discussion of the main findings.

2. Numerical experiment design, model configuration, and evaluation data

a. *NWP configurations run by CAPS for the 11th HMT winter weather experiment (2021/22)*

The forecast experiments in this study use the NOAA Global Systems Laboratory (GSL) version of FV3-LAM, checked out from a Github repository maintained by GSL (https://github.com/NOAA-GSL/ufs-weather-model) on 16 October 2020. The experiments are run on the same grid (Fig. 1) as the Environmental Modeling Center (EMC) FV3-LAM run for the 11th WWE, with 1799 × 1059 horizontal grid points and a grid spacing of ~3 km. The results shown in section 3 are based on simulations that employ the Global Forecast System (GFS) 0.25° resolution initial conditions and lateral boundary conditions. An analogous set of integrations initialized with the High-Resolution Rapid Refresh (HRRR) analysis, used to assess the robustness of the GFS-initialized results, are discussed briefly at the end of that section. Five physics suites (referred to as CNTL, LSM1, LSM2, MP1, and MP2; see Table 1) are selected with different combinations of land surface models (LSMs), planetary boundary layer (PBL), surface layer, and microphysics schemes. All experiments use the same Rapid Radiative Transfer Model for GCMs (RRTMG) radiation (Iacono et al. 2008). These five physics suites bear similarity to those used by current or planned future operational or experimental forecasting systems of NOAA, including the planned control member of RRFS, the Hurricane Weather Research and Forecasting (HWRF) model and planned Hurricane Analysis and Forecasting System (HAFS), the HRRR model, the experimental Warn on Forecast System (WoFS), and the GFS. These suites are relatively well understood, and their codes have been extensively tested for robustness as part of designing their respective modeling systems, and thus are expected to be maintained in future operations. FV3-LAM forecasts using these five configurations are performed once weekly, plus other days of interest, for a total of 35 days between 26 October 2020 and 12 March 2021, which covers the period of the WWE operations plus additional significant winter weather events. Forecasts are run with an initial time of 0000 UTC on those days and are run for 84 h.

The full-season forecasts are first evaluated in terms of a few near-surface fields, then a detailed diagnosis of model errors is performed for the case of 27 December 2020, for which the five suites display biases consistent with the full-season performance (Fig. 2). To diagnose model bias in the five physics suites, we run additional forecasts with single physics parameterization alterations (for a total of 15 combinations; Table 2), for 27 December. Note that these 15 combinations include the five physics suites mentioned above, but with different names following the M#B#L# naming scheme with letters M, B, and L referring to the microphysics, PBL scheme, and land surface model, respectively, each followed by a number representing one of the individual schemes. In a separate paper, Supinie et al. (2022) provide a broad overview of the impacts of the choice of microphysics, PBL, and LSM on forecast metrics, including surface fields and precipitation. In contrast, this study focuses on diagnosing near-surface model biases that can be mainly attributed to the LSM, PBL, and surface layer schemes.

Six more sensitivity simulations are run to identify parametric errors. In these experiments, parameters are adjusted within a plausible range for the Noah-MP LSM, *K*-profile-based eddy-diffusivity mass-flux (K-EDMF) PBL, and GFS surface layer schemes (Table 3) to examine the parameter sensitivity and better determine the root cause of the model errors. The adjusted parameters include snow emissivity in Noah-MP; p, an exponent in the polynomial function determining vertical mixing strength in K-EDMF, where larger p results in smaller mixing coefficients (Troen and Mahrt 1986; Hu et al. 2010b); α, an empirical coefficient used to calculate critical bulk Richardson number (Ri) in K-EDMF; z_o, the surface aerodynamic roughness length (z_{0m}) over grass; and C_{zil}, the so-called Zilitinkevich coefficient in the GFS surface layer scheme to

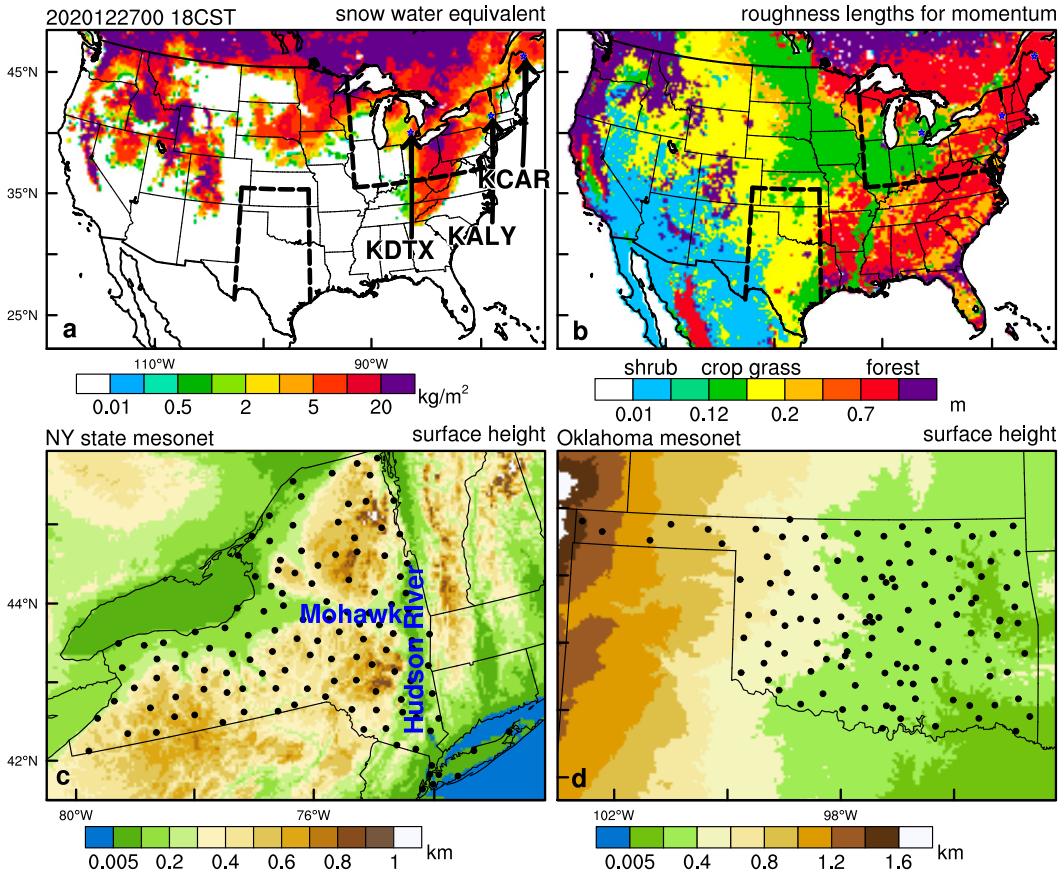

FIG. 1. (a) Snow water equivalent on 27 Dec 2020 and (b) roughness length for momentum over the model domain and site location of (c) New York State and (d) Oklahoma Mesonet. The Hudson River and Mohawk River valleys are marked in (c). Two sounding sites over significant snow, KCAR (ME) and KDTX (MI), and one sounding site with insignificant snow in the Hudson River valley, KALY (NY), are marked in (a). Dashed lines in (a) and (b) mark the U.S. Northeast and southern Great Plains regions, where evaluation against URMA is performed. The corresponding vegetation types associated with each roughness length are marked on top of the color bar in (b).

relate the roughness length of heat (z_{0h}) to that of momentum (z_{0m}) through (Zilitinkevich 1995)

$$z_{0h} = z_{0m}\exp(-kC_{\text{zil}}\sqrt{u_* z_{0m}/\nu}), \tag{1}$$

where k is the von Kármán constant, u_* is the friction velocity, and ν is the kinematic molecular viscosity of air. Note that C_{zil} typically varies between 0.01 and 1, with smaller values resulting in stronger land–atmosphere coupling (Chen et al. 1997); C_{zil} has been identified as one of the most sensitive parameters in land surface processes (Trier et al. 2011; Chaney et al. 2016).

b. Observation data sources

The forecasts produced during the 2020/21 WWE are first evaluated against the URMA gridded surface fields over the same five geographic regions as in Supinie et al. (2022): the Northeast, Southeast, Southern Great Plains, North Central, and Mountain West regions (Fig. 1 of Supinie et al. 2022). The two regions with the most severe biases found in this study, the northeastern United States (Northeast for short) and the southern Great Plains, are marked in Figs. 1a and 1b. URMA is produced using the same data assimilation system as the Real-Time Mesoscale Analysis (RTMA) (De Pondeca et al. 2011) but run 6 h afterward to capture late-arriving observations and is designed with an emphasis on fidelity to observed surface conditions (e.g., Morris et al. 2020), making it suitable for verification of near-surface forecasts.

In addition to the URMA analyses, data from two statewide mesoscale observing networks are used for evaluation. The Oklahoma Mesonet (http://mesonet.org/) is a network of 120 automated meteorological stations (Fig. 1d) with minimal influence from urban landscapes (McPherson et al. 2007; Basara et al. 2008). Each Mesonet station measures more than 20 environmental variables, including wind at 10 m and air temperature at 1.5 m AGL. Data from the Oklahoma Mesonet have been widely used in weather forecasting and atmospheric research (e.g., Hu et al. 2013b,c, 2016). The New York State Mesonet (http://nysmesonet.org) is another statewide weather observing network consisting

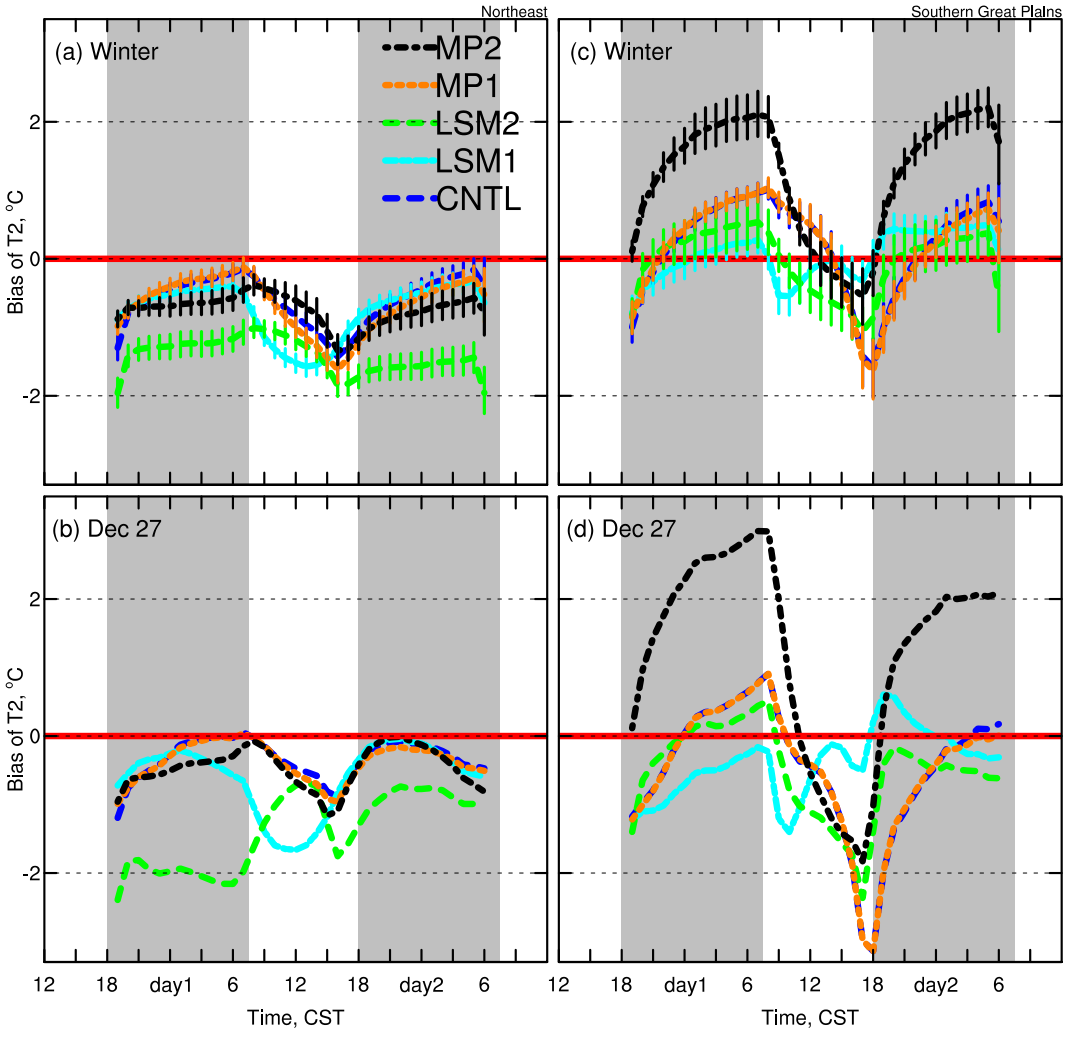

FIG. 2. (a),(c) Bias of 2-m temperature (T_2) and its 95% bootstrap confidence interval in the first 36 h of forecasts during winter of 2020/21 and (b),(d) T_2 bias during 27 Dec 2020 evaluated against the Unrestricted Mesoscale Analysis (URMA) over the (left) Northeast and (right) southern Great Plains.

of 181 state-of-the-art environmental monitoring stations (Fig. 1c). A subset network of 126 standard sites serves as the foundation of an Early Warning Severe Weather Detection network for the entire state of New York (Brotzge et al. 2020; Shrestha et al. 2021).

Radiosonde Replacement System (RRS) sounding data are used to verify the vertical extent of near-surface model biases, similar to previous boundary layer studies (e.g., Hu and Xue 2016; Hu et al. 2019b, 2020). The 1-s (vertical resolution in time per measurement) sounding data are available in BUFR format and are decoded and sampled at standard and significant levels using software provided by NOAA (available online at ftp://ftp.ncdc.noaa.gov/pub/data/ua/rrs-data/). The RRS sounding data are available at the standard observation times of 1200 UTC (0600 CST) and 0000 UTC (1800 CST). Soundings from the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) site (downloaded from https://doi.org/10.5439/1021460) are also used.

3. Results

a. *Full-season performance compared to URMA*

When evaluated against URMA surface data, trends in the performance of the HMT WWE forecasts are noted in a number of meteorological fields, particularly surface temperature and wind speed. While a detailed evaluation of full-season forecasts for the five sub-CONUS regions is reported by Supinie et al. (2022), prominent biases found in two of the regions are the primary subject of this study. These biases are summarized below.

For 2-m temperature, the LSM2 experiment performs particularly poorly over the northeast region, with a strong cold bias (from −1.5 to −2.0 K) compared to URMA during the nighttime and morning hours (Figs. 2a and 3a). Over the other four regions, MP2 is a clear outlier, exhibiting different bias behavior than the other experiments that have near-neutral bias throughout the forecast period. MP2

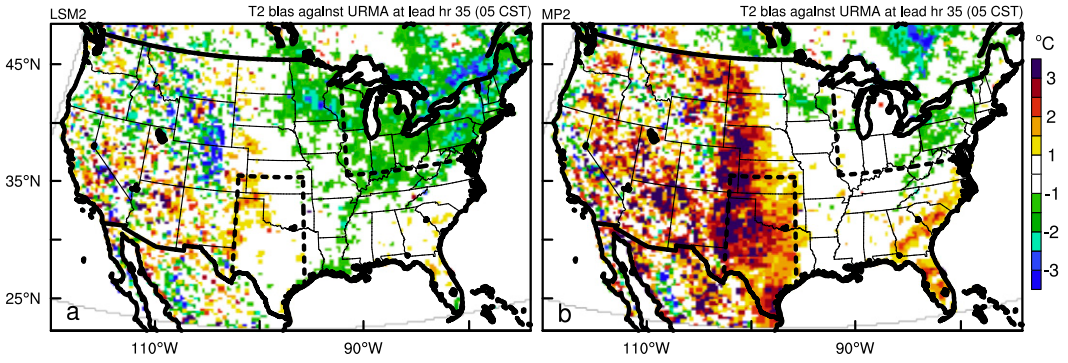

FIG. 3. Bias of 2-m temperature (T_2) of (a) LSM2, and (b) MP2 at 0500 CST (1100 UTC; lead time 35 h) evaluated against the Unrestricted Mesoscale Analysis (URMA). Two regions with significant bias, the Northeast and the southern Great Plains, are delineated in dashed black.

exhibits the highest positive bias of all experiments, particularly over the southern Great Plains (Figs. 2c and 3b) and also exhibits the largest diurnal variability in error. Furthermore, 95% confidence intervals (constructed from 1000 bootstrapped samples with replacement) reveal that LSM2 has a statistically significant cold bias in the Northeast (Fig. 2a) while MP2 has a statistically significant warm bias in the southern Great Plains compared to other experiments at all times except during the afternoon hours (Fig. 2c).

Compared to URMA, all forecasts exhibit a positive bias in 10-m wind speed (~1.0 m s^{-1}; not shown), presumably in part because the influence of obstructions and roughness on the surface winds are generally underestimated in models (Fovell and Gallagher 2020; Shen et al. 2022). Also, the RTMA/URMA analysis system is known to have a low bias in its 10-m wind speed analysis, largely owing to assimilated wind observations with nonstandard instrument siting (Morris et al. 2020). The root-mean-square error (RMSE) of wind speed grows slightly over the forecast period, from around 1.5 to around 2.0 m s^{-1}. Both the positive bias and RMSE are slightly larger in the evening, particularly in MP2, LSM2, and LSM1. The differences between this set of three members and the other two (CNTL and MP1) are substantial during the evening hours, particularly over the southern Great Plains.

There are some days with prominent examples of the season-wide trends noted here, for example, 27 December 2020 (Figs. 2b,d). On this day, the Northeast and southern Great Plains regions were south of a cold front with clear skies. The 27 December case is fairly typical in terms of winter weather and the biases observed for the season, so it is selected for detailed analysis and diagnosis in terms of the cold bias over the Northeast from LSM2 (Figs. 2a,b) and nocturnal warm bias over the Great Plains from MP2 (Figs. 2c,d).

b. *Cold bias over snow cover in the Northeast*

The near-surface bias over the Northeast is further examined by comparing predicted T_2 with New York State Mesonet observations (Fig. 4). LSM2 shows a prominent cold bias in the region (Fig. 4d). The spatial distribution of the cold bias is consistent with the spatial distribution of snow cover (Fig. 4b) except just east of Lakes Erie and Ontario, where warm air is advected onshore. In the low-elevation region of the Hudson River and Mohawk River valleys, where snow cover is not significant (Fig. 4b), the cold bias from LSM2 is not as prominent. The collocation of cold bias with snow cover suggests that elements of the LSM2 parameterization active over regions of snow cover likely lead to this error.

To isolate the root cause of the biases, additional sensitivity simulations making up a total of 15 physics combinations (Table 2) are examined. When the Noah LSM is used in place of Noah-MP (M0B2L1), the previously noted cold bias is no longer seen (Fig. 4g). By changing the configuration of CNTL to use Noah-MP, a cold bias appears over the snow-covered region with a similar spatial distribution as that from LSM2 but with a slightly smaller magnitude (Fig. 4h). This result corroborates the substantial impact of land surface models and a secondary impact from different PBL and surface layer schemes between LSM2 and M0B0L1. These sensitivity simulations suggest that it is the LSM (i.e., Noah-MP) in the LSM2 configuration that primarily leads to the cold bias over regions of snow cover.

In addition to the spatial distribution of model biases, profiles at two northeastern sounding sites with snow cover [Caribou, Maine (KCAR), and Detroit, Michigan (KDTX)] are examined to reveal the vertical distribution of model bias. There are two routine sounding sites in the state of New York: Albany, located in the Hudson River valley without substantial snow (Fig. 1), and Buffalo, which is affected by warm advection from Lake Erie. For these reasons, neither of these sites show a substantial cold bias, and thus they are not suitable for diagnosing the cold bias over snow. The surface air temperature predicted by LSM2 is 4°–5°C lower than other members (and observations) over Caribou (KCAR) and Detroit (KDTX) (Figs. 5b,e). Overly strong surface cooling from LSM2 leads to too stable of a near-surface boundary layer over KCAR and KDTX. The profile evaluation confirms that the cold bias from LSM2 is not just limited to T_2, a diagnosed variable, but extends over the entire boundary layer. Thus, the LSM2 cold bias is not just caused by the different T_2 tile-diagnosing method (Barlage et al. 2020) employed in Noah-MP, but is

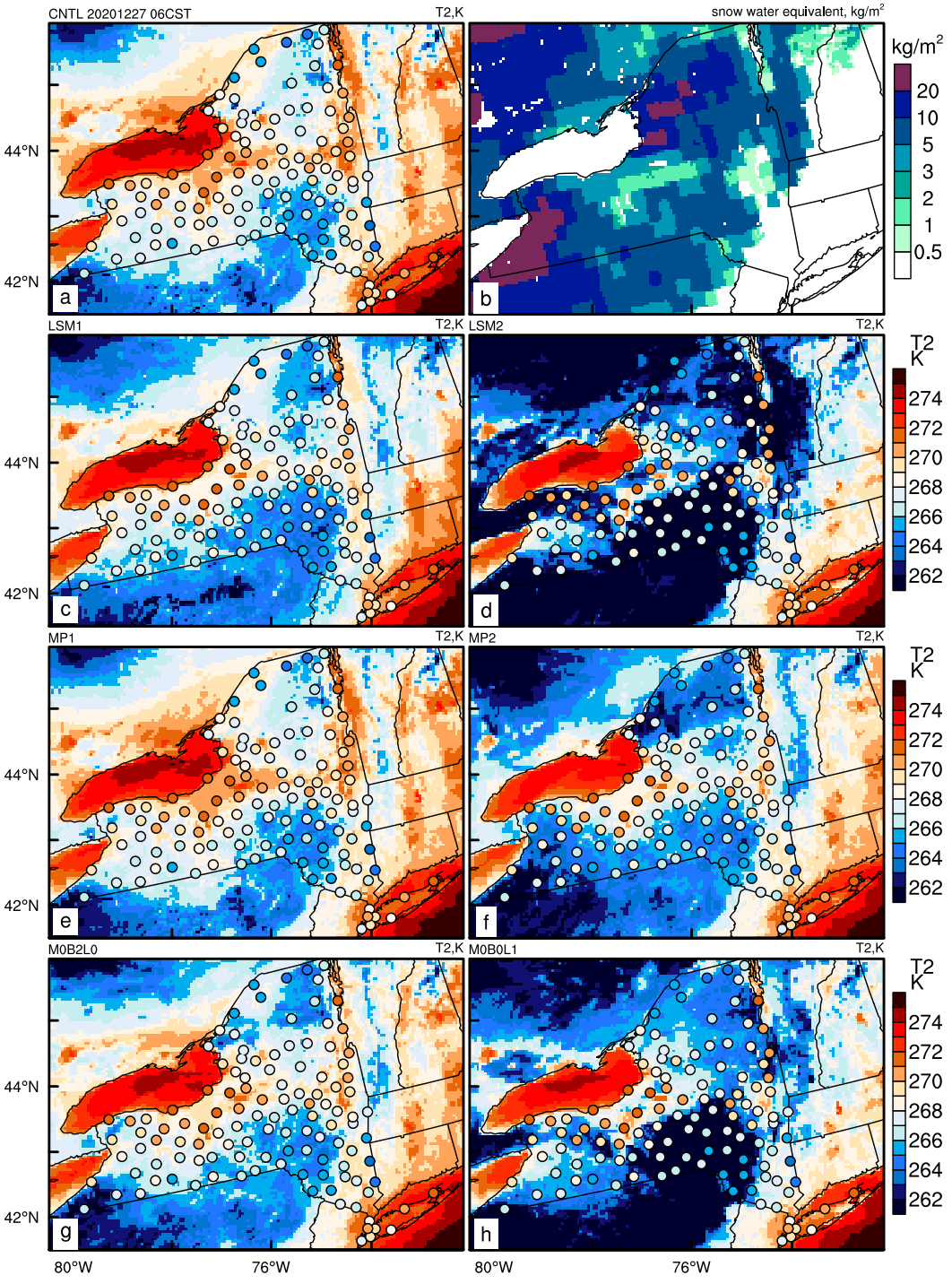

FIG. 4. 2-m temperature (T_2) at 0700 EST 27 Dec 2020 (1200 UTC; lead time 12 h) simulated by the five physics suites [(a) CNTL, (c) LSM1, (d) LSM2, (e) MP1, and (f) MP2] and two members from the 15-member ensemble [(g) M0B2L0 and (h) M0B0L1] and overlaid with New York State Mesonet observations. (b) Simulated snow water equivalent (from the GFS analysis).

a real near-surface bias caused by errors in that LSM. LSM2 also predicts negative latent heat flux over the snow cover; that is, downward water vapor transfer at the land–atmosphere interface. As a result, LSM2 predicts a near-surface dry bias and a positive vertical gradient of specific humidity ($\partial q/\partial z > 0$) in the near-surface stable boundary layer over regions of snow cover, which contrasts with the observed soundings and other simulations (Figs. 5c,f). The

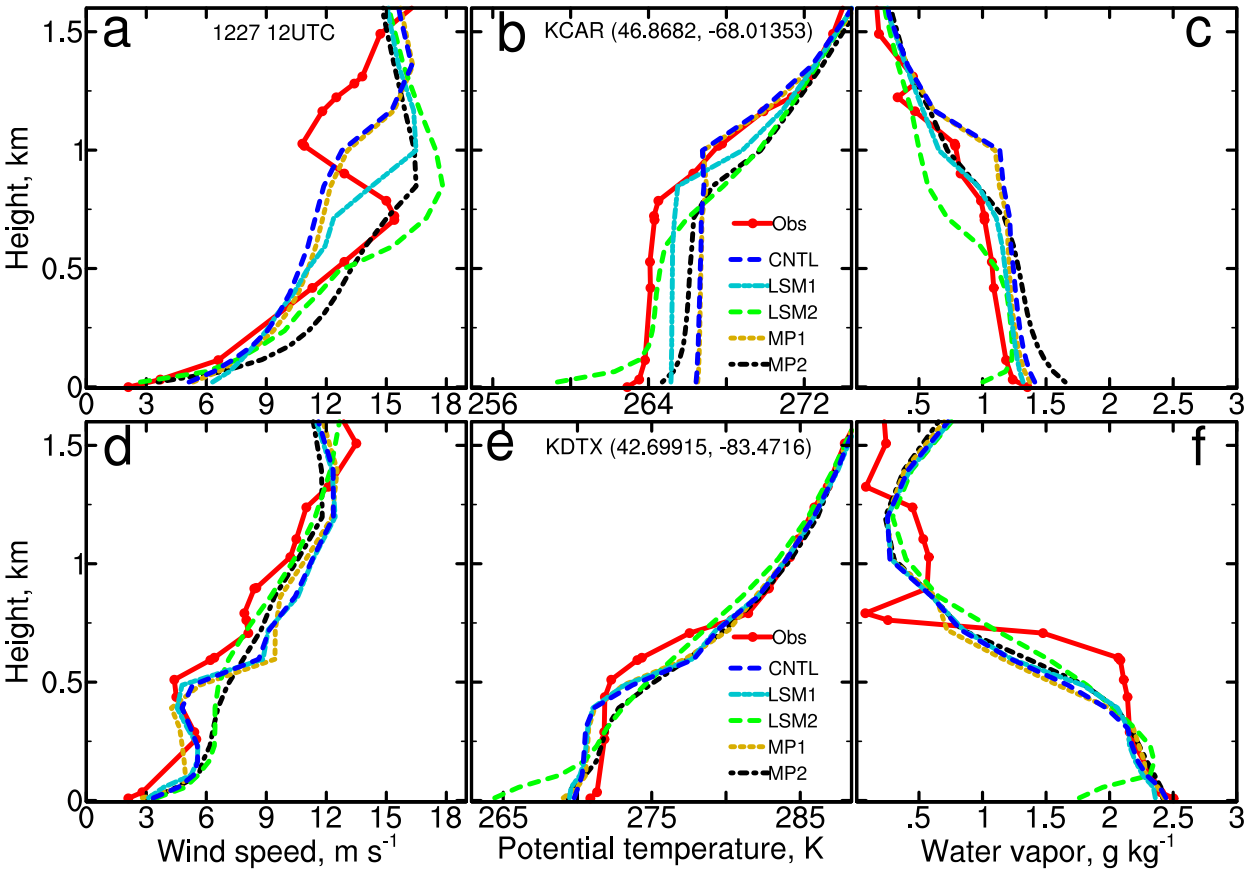

FIG. 5. Profiles of (left) wind speed, (center) potential temperature, and (right) water vapor mixing ratios over two northeastern sounding sites with snow cover, namely (a)–(c) KCAR (Caribou, Maine) and (d)–(f) KDTX (Detroit, Michigan), at 0700 EST 27 Dec 2020 (1200 UTC) from radiosonde and the 5-suite simulations (lead time 12 h).

dry bias from LSM2 also appears when evaluating the 2-m dewpoint against URMA (not shown). The opposite sign of the vertical gradient of water vapor between the different simulations is probably not due to the different PBL schemes because differences in vertical mixing within PBL schemes normally modulate the vertical gradient of variables but do not change the sign (Hu et al. 2010a; Nielsen-Gammon et al. 2010; Hu et al. 2013a, 2019a). Profiles over KDTX from the 15 experiments are further examined in Fig. 6. In those that use Noah-MP (experiment names ending with L1 including M0B0L1 in Fig. 6b, M0B2L1 and M1B0L1 in Fig. 6e, and M2B0L1 in Fig. 6h), an overly stable near-surface boundary layer and positive gradient of water vapor in the near-surface stable boundary layer are simulated, whereas the other physics suites do not exhibit this behavior. These results therefore further suggest that the Noah-MP land surface model is likely the root cause of the different behavior seen in LSM2.

Sensitivity experiments tuning the parameters within the Noah-MP land surface model are conducted to identify which aspect of the Noah-MP is leading to the cold bias from LSM2. The implementation of Noah-MP used in this study is equivalent to the Noah-MP version in the Weather Research and Forecasting (WRF) Model version 4.2.2. He et al. (2019) reported that the uncertainties in snow emissivity and model parameters in snow cover fraction formulation, and hence albedo, using this version of Noah-MP may lead to a cold bias. These factors, along with some other minor changes, were updated in Noah-MP during the WRF update from version 4.2.2 to 4.3.

Our FV3-LAM winter experiments were initialized in the early evening in the CONUS, and the cold bias from the Noah-MP in LSM2 became prominent within an hour after model initialization (not shown). Given that shortwave radiation in the eastern United States is zero at this time (1900–2000 EST) in winter and that snow cover in different simulations is nearly the same (mostly inherited from the initial conditions derived from the GFS analysis), the uncertainties in snow albedo cannot be the dominant factor. Snow emissivity is another plausible explanation for the cold bias. The snow emissivity is set to 1.0 in Noah-MP in the FV3-LAM version from autumn 2020 that is used in our 2020/21 winter experiment forecasts, while it is set to 0.95 in Noah and 0.98 in the RUC LSM. Two more FV3 sensitivity simulations with the LSM2 physics suite are conducted for the 27 December case to diagnose the impact of snow emissivity by setting it first to 0.95 and then 0.9 (Table 3). By reducing snow emissivity to 0.95, the simulated temperature over

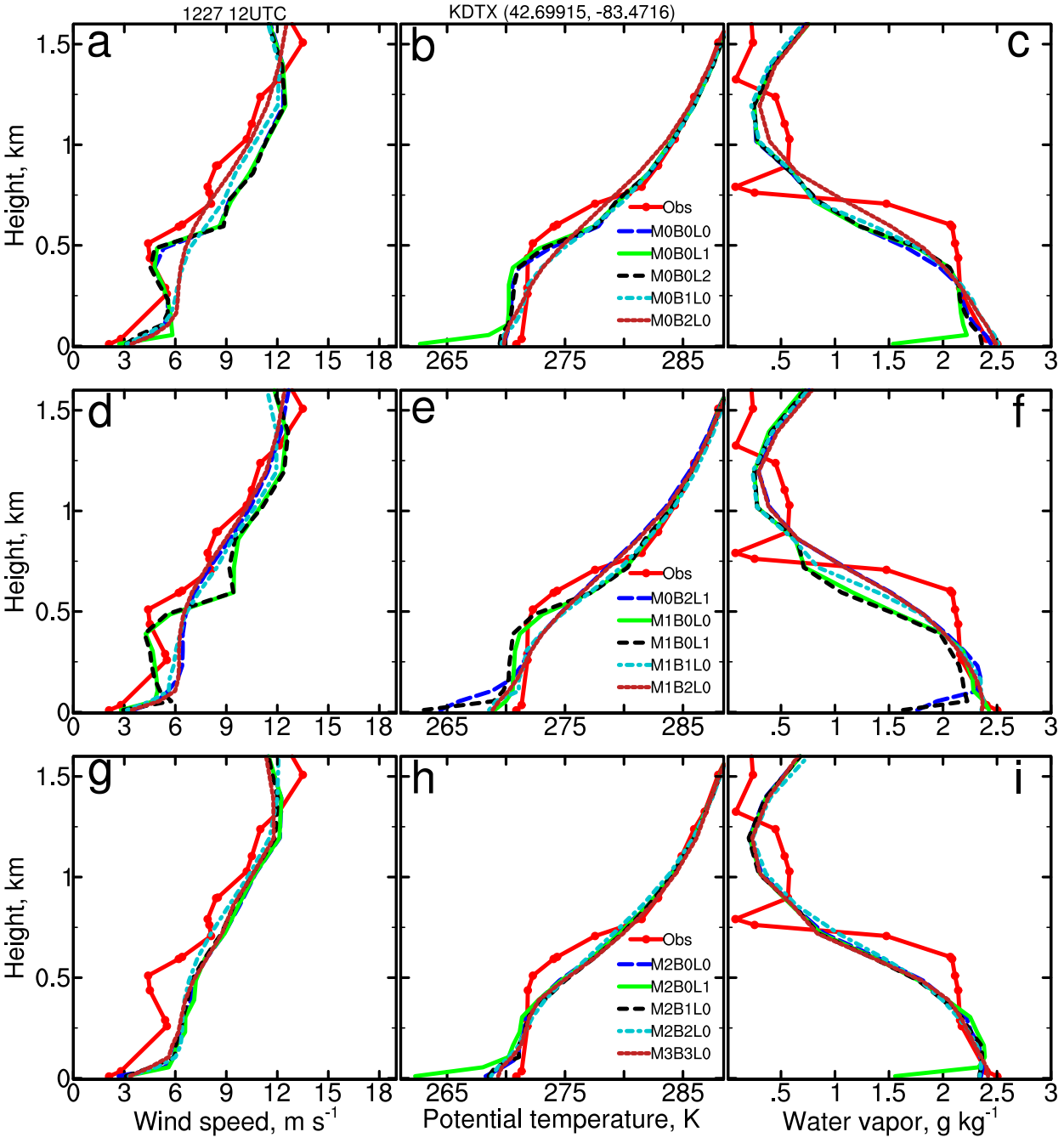

FIG. 6. As in Fig. 5, but for the 15-member simulations over KDTX (Detroit, Michigan). A separate set of five physics suites are presented in each row.

snow cover is indeed increased compared to that exhibited when using snow emissivity of 1.0, but only by 0.2°C after 12 h over the Northeast (Fig. 7a). By further reducing snow emissivity to 0.9, the simulated temperature over snow cover is further increased, but only by 0.4°C over regions of snow cover (Fig. 7b) in that same time period, an order of magnitude smaller than the overall cold bias of 4°–5°C over snow-covered regions. Even with reduced snow emissivity, the prominent cold bias still persists (Figs. 7c,d); uncertainty in snow emissivity therefore cannot entirely account for the LSM2 cold bias.

As the FV3-LAM version that was available in autumn 2020 cannot run with the latest Common Community Physics Package (CCPP; Firl et al. 2021), this FV3 version cannot run with the updated Noah-MP version used in WRFv4.3. We thus use WRF to verify impact of other updated factors in Noah-MP

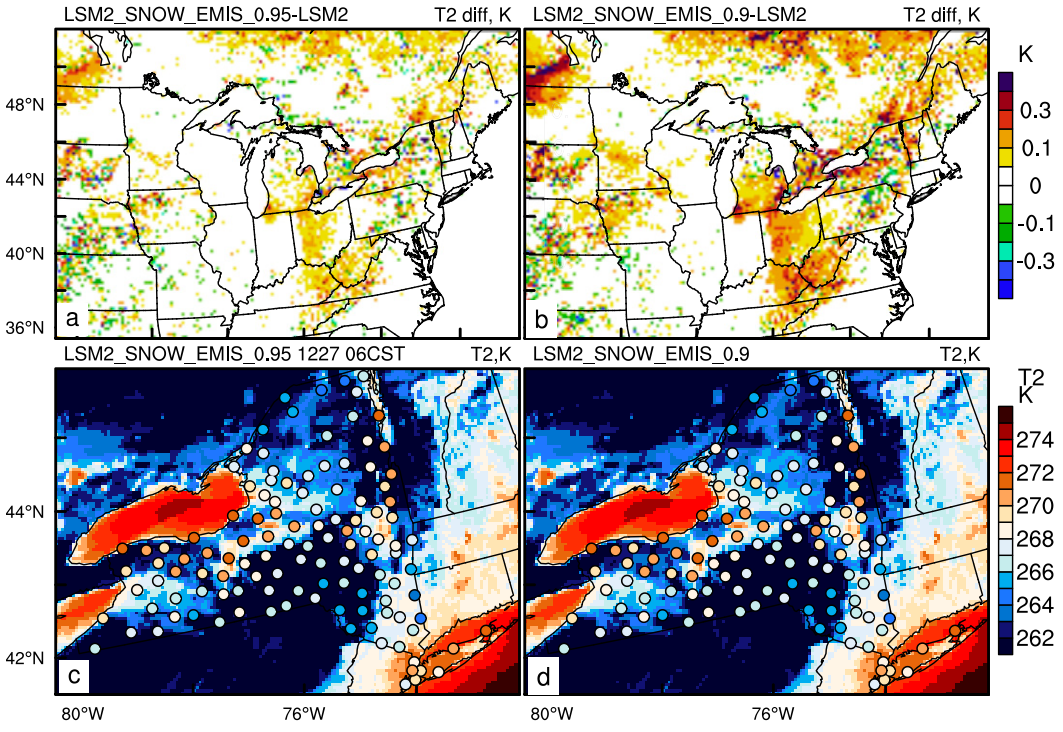

FIG. 7. Simulated (a),(b) T_2 difference from the LSM2 suite and T_2 by the (c) LSM2_SNOW_EMIS.95 and (d) LSM2_SNOW_EMIS.9 sensitivity experiments at lead time 12 h. New York State Mesonet surface temperature data (shaded circles) are overlaid in (c) and (d). See the description of the two LSM2 sensitivity simulations in Table 3.

considered in transition from WRFv4.2.2 to WRFv4.3. We ran WRFv4.2.2 and WRFv4.3 for the 27 December case with both Noah and Noah-MP. WRFv4.2.2 with Noah-MP reproduces the cold bias over regions of snow cover. Noah-MP is too cold over the majority of the snow cover (3°–5°C colder), particularly over the Northeast, similar to our FV3 simulation. WRFv4.3, using the updated version of Noah-MP, does not alleviate the cold bias to a substantial level (not shown). Thus, the updates in Noah-MP from version WRFv4.2.2 to WRFv4.3 reported by He et al. (2019) appear incapable of solving the cold bias issue over snow cover noted in our LSM2 FV3 simulation.

Other structural or parametric errors must be responsible for the cold bias from NOAH-MP over snow cover. Compared to the CNTL experiment, LSM2 with Noah-MP simulates lower soil water content under snow cover, including both soil moisture (Fig. 8d) and soil liquid water (not shown), which would lead to lower thermal conductivity because soil thermal conductivity linearly depends on water content (Ochsner et al. 2019). In the presence of nighttime surface radiative cooling, lower thermal conductivity over snow-covered regions in Noah-MP leads to heat transfer from soil layers to the land surface at a lower rate. This leads to smaller upward ground flux (Fig. 8b) and lower skin temperature, subsequently leading to lower surface temperature (Fig. 8a), but higher soil temperature (Fig. 8c). The spatial consistency among soil moisture, ground flux, soil temperature, and surface temperature shown in Fig. 8 suggests the critical role of soil thermal conductivity on the model bias of near-surface temperature associated with Noah-MP in LSM2. The difference in the fields between LSM2 with Noah-MP and M0B2L0 with Noah (which differ only by their land surface model) is very similar to that in Fig. 8 (not shown), which supports the idea that the cold bias from the LSM2 suite is dominated by model errors associated with Noah-MP. Thus, calibration of predicted soil water content or improved partitioning between different water phases, and consequently soil thermal conductivity in Noah-MP may lead to improved performance in the future.

c. *Nocturnal warm bias over short vegetation in the Great Plains*

From the full-season evaluation using both the URMA data (Fig. 2c) and Oklahoma Mesonet data (Fig. 9), the MP2 (M3B3L0) experiment shows the most prominent warm bias during nighttime (from 1900 to 0700 CST, i.e., from 0100 to 1300 UTC) over the Great Plains, with a mean nocturnal bias of 1.75°C and a root mean-square error (RMSE) of 3.18°C over Oklahoma Mesonet sites (Table 4); the warm bias increases during the overnight hours (Fig. 9). Meanwhile, MP2 also exhibits the largest positive bias for surface wind speeds (Fig. 10) with the largest nighttime mean bias of 1.0 m s^{-1} (normalized mean bias of 28.6%) and a RMSE of 2.28 m s^{-1} over Oklahoma Mesonet sites among the five suites, followed by LSM1 and

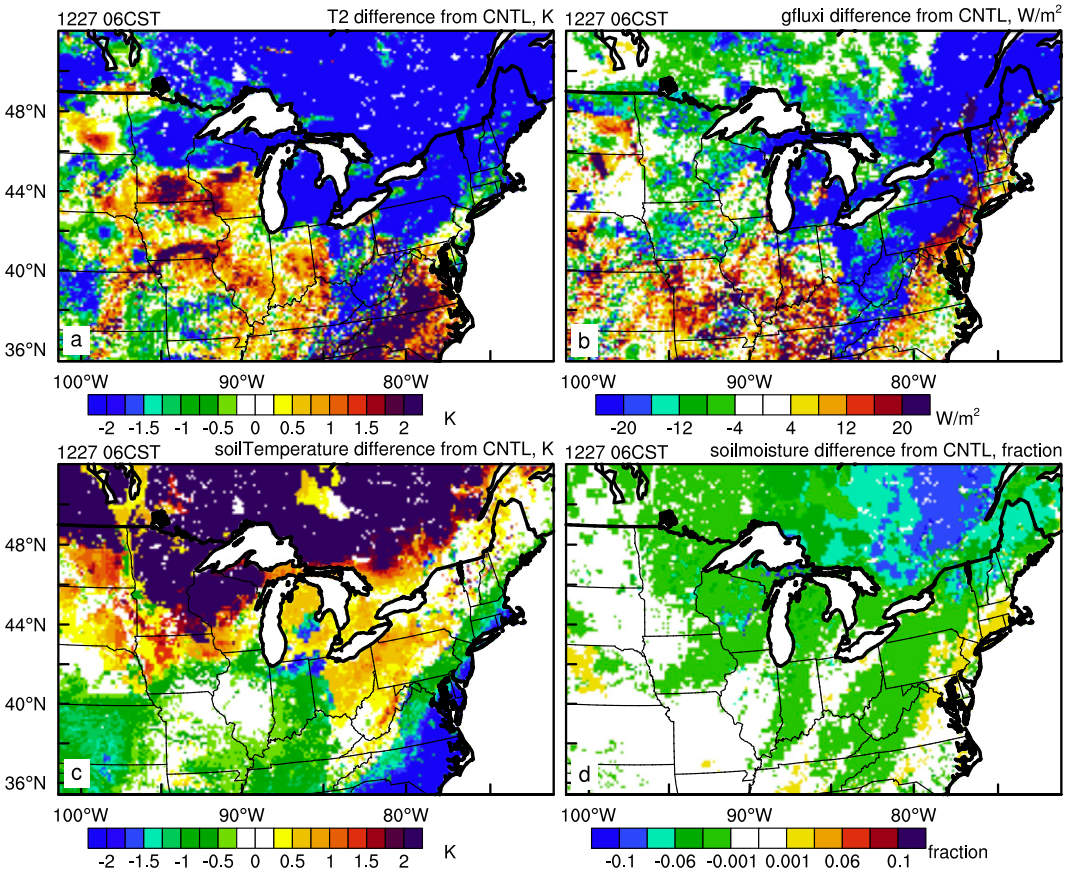

FIG. 8. Simulated difference from the CNTL simulation of (a) 2-m temperature (T_2), (b) upward ground heat flux, (c) soil temperature, and (d) soil moisture by the LSM2 suite with the Noah-MP land surface model at lead time 12 h.

LSM2 (Table 5). The case study on 27 December 2020 also reveals the same systematic bias evaluated using Oklahoma Mesonet sites (Figs. 11a and 12a,b), particularly during the two nights preceding the passage of a cold front on 28 December (Fig. 11a).

1) Overly strong nighttime vertical mixing from K-EDMF in MP2

The nighttime warm bias may be attributed to model errors in the PBL schemes (Hu et al. 2010a, 2013a). During the night, longwave radiative cooling leads to surface cooling and a positive lapse rate ($\partial\theta/\partial z > 0$) in the stable boundary layer (Figs. 13b,e). Thus, any turbulence in the stable boundary layer transports warmer air downward to the surface, creating downward sensible heat fluxes. Under such conditions, nighttime mixing that is too weak would lead to weakened downward sensible heat flux, leading to unrealistic decreases in surface temperature from radiative cooling and creating a surface layer decoupled from the layer above (Derbyshire 1999). On the other hand, nighttime mixing that is too strong (and thus strengthened downward sensible heat flux) would normally lead to a surface warm bias (Hu et al. 2010a, 2013a). Similarly, since wind speed typically increases with height near the surface, overly strong mixing would also lead to a positive surface wind speed bias (Hu et al. 2013a).

To verify the hypothesis that the warm bias and concurrent positive wind speed bias exhibited by MP2 are due to overestimated mixing, vertical mixing coefficients extracted from the five primary physics suites are examined. K-EDMF in MP2 indeed predicts a larger vertical mixing coefficient in the nighttime boundary layer than other members (Figs. 13c,f), which likely contributes to MP2's nighttime near-surface warm bias and positive wind bias. The stronger vertical mixing from K-EDMF, and thus stronger downward sensible heat flux in the stable boundary layer, provides increased heating of the surface layer. As a result, the stronger mixing overwhelms the surface radiative cooling, which results in reduced stability in the stable boundary layer (Figs. 13). Because of the reduced stability, the boundary layer simulated by K-EDMF is thicker (Figs. 12g,h). According to inertial oscillation theory, the wind vector oscillation amplitude peaks at the top of the stable boundary layer (Blackadar 1957; Shapiro and Fedorovich 2010; Klein et al. 2016). As a result, the nose of the nocturnal low-level jet simulated by K-EDMF in MP2 develops at a higher elevation, at the top of a thicker stable boundary layer as seen in soundings from both Fort Worth/Dallas (FWD), Texas, and the

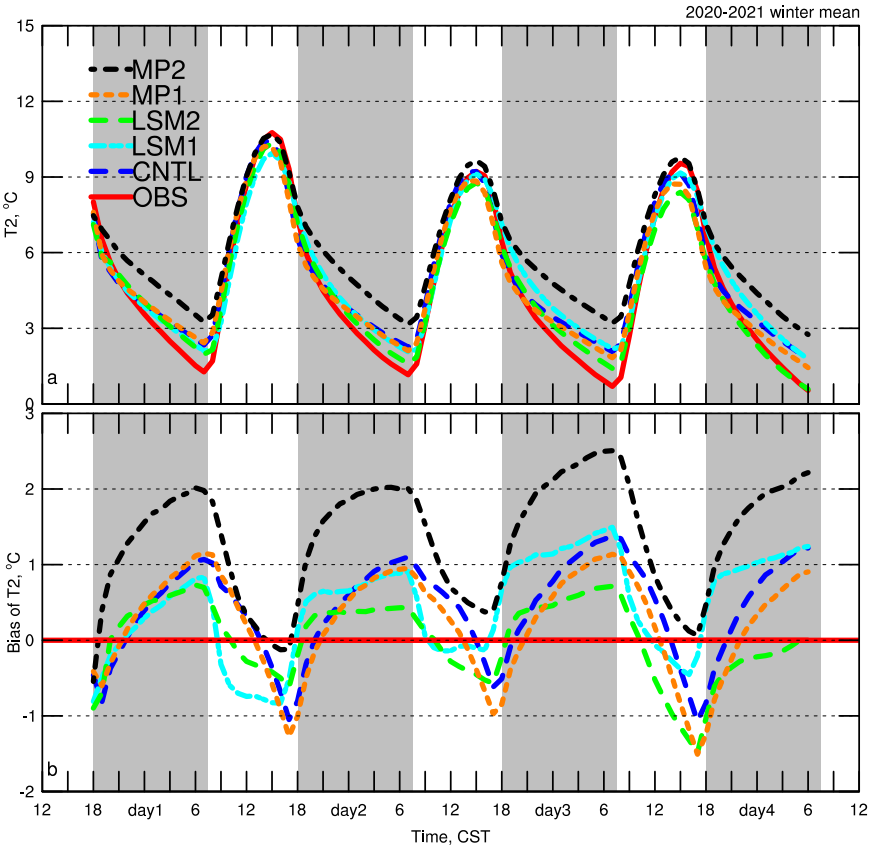

FIG. 9. (a) Simulated and observed average 2-m temperature (T_2) and (b) the corresponding model bias over Oklahoma Mesonet sites during winter of 2020/21.

ARM SGP Central Facility site in Lamont, Oklahoma (Figs. 13a,d).

Sensitivity experiments with adjusted vertical mixing coefficients in K-EDMF (Table 3) are conducted to verify the impact of vertical mixing on the surface warm bias. K-EDMF uses a first-order local mixing approach (along with mass flux nonlocal mixing) and calculates the vertical mixing coefficient as a function of height in the boundary layer, ranging from linear to cubic, as in the Asymmetric Convective Model (version 2) and Yonsei University PBL schemes (Han and Pan 2011; Han et al. 2016). The exponent p in the vertical mixing function with height determines vertical mixing strength in the boundary layer and is set to 2 by default in K-EDMF. However, it may vary between 1 and 3 depending on different flow conditions, with a larger p yielding smaller mixing coefficients (Troen and Mahrt 1986; Hu et al. 2010b). A sensitivity run with $p = 3$ (MP2_p3) is conducted; in this experiment, the vertical mixing coefficient is reduced and the warm bias over the Great Plains is reduced compared to the default MP2, but temperatures remain higher than in the CNTL simulation (not shown).

In addition to the vertical mixing coefficient, the excessively high PBL height (PBLH), dictated by other factors, predicted

TABLE 4. Evaluation statistics for simulated (sim) nighttime hourly temperature at 2 m AGL (T_2) against the Oklahoma Mesonet data during winter of 2020/21. The metrics include mean bias (MB), root mean-square error (RMSE), and normalized mean bias (NMB). Their formulas can be found in Yu et al. (2006). Simulated values at the nearest grid point to the Mesonet station is used for comparison, following our previous model evaluation practice (e.g., Hu et al. 2010a, 2013a, 2019b, 2021).

	Nighttime T_2 at all Oklahoma Mesonet sites during winter				
	CNTL	LSM1	LSM2	MP1	MP2
Mean obs (°C)			2.99		
No.			195 953		
Mean sim (°C)	3.59	3.79	3.24	3.44	4.75
MB (°C)	0.59	0.79	0.25	0.45	1.75
RMSE (°C)	2.70	2.64	2.74	2.62	3.18
NMB (fraction)	19.9%	26.5%	8.3%	15.0%	58.6%

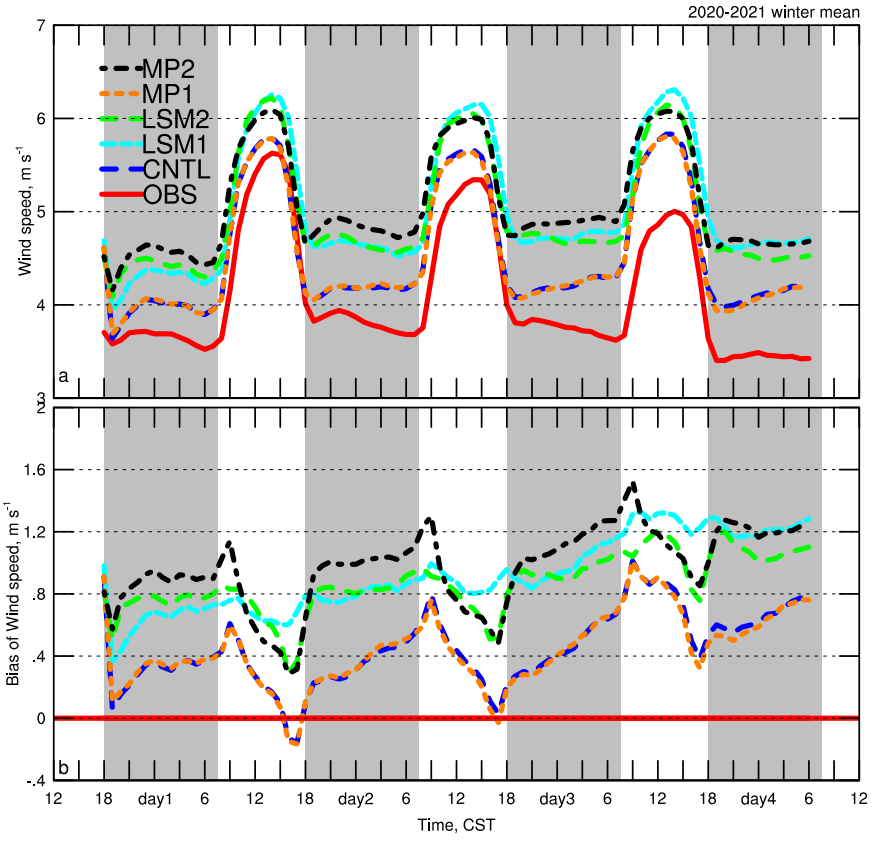

FIG. 10. As in Fig. 9, but for 10-m wind speed.

by MP2 compared to the CNTL simulation over the Great Plains (Figs. 12g,h) also contributes to higher downward heat flux. A higher PBLH can result in entrainment of warmer air from higher altitudes and lead to warmer surface temperature. On the night of 26–27 December, a prominent LLJ (Figs. 13a,d) developed over the southern Great Plains ahead of a cold front (see surface winds in Figs. 12d,e) and PBLH in the LLJ region is higher compared to that in the eastern and southeastern United States. K-EDMF in MP2 predicts a PBLH nearly twice that of the CNTL simulation, using the Mellor–Yamada–Nakanishi–Niino (MYNN)-EDMF, over the southern Great Plains. For the nighttime stable boundary layer, the critical bulk Richardson number (Ri) used by K-EDMF to determine PBLH is a function of surface wind speed and roughness length (Vickers and Mahrt 2004):

$$\mathrm{Ri} = \alpha\left(10^{-7}\frac{U_{10}}{fz_o}\right)^{-0.18}, \tag{2}$$

where z_o is surface aerodynamic roughness length, f is the Coriolis parameter, and U_{10} is the 10-m wind speed. Currently, the empirical parameter α takes a value of 0.16 based on calibration using PBLH data from two tower datasets over land and one aircraft dataset over the ocean, during which observed PBLH was mostly <150 m (Vickers and Mahrt 2004). However, the nighttime PBLH over the southern Great Plains predicted by MP2 is as high as 700 m (Fig. 12h). Thus, the empirical

TABLE 5. Evaluation statistics for simulated nighttime hourly surface wind speed against the Oklahoma Mesonet data during winter of 2020/21.

	Nighttime wind speed at all Oklahoma Mesonet sites during winter				
	CNTL	LSM1	LSM2	MP1	MP2
Mean obs (m s^{-1})			3.66		
No.			190 428		
Mean sim (m s^{-1})	4.11	4.57	4.57	4.11	4.71
MB (m s^{-1})	0.45	0.90	0.9	0.45	1.05
RMSE (m s^{-1})	2.04	2.25	2.12	2.03	2.28
NMB (fraction)	12.2%	24.7%	24.6%	12.2%	28.6%

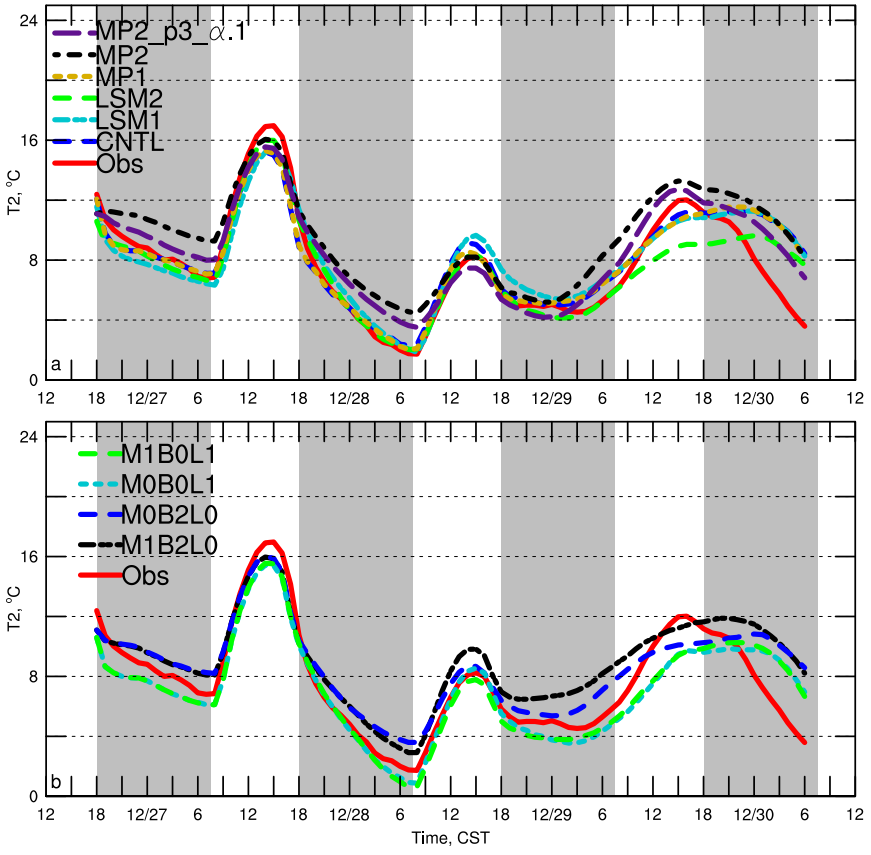

FIG. 11. Mean 2-m temperature over Oklahoma Mesonet sites observed and simulated by (a) the five physics suites and one MP2 sensitivity experiment MP2_$p3$_α.1, and (b) four members from the 15-member ensemble for the 27 Dec 2020 case including two (M1B2L0 and M0B2L0) using the GFS surface layer scheme and two (M0B0L1 and M1B0L1) using the MYNN surface layer scheme.

parameter α of 0.16 may not be appropriate when applied over the Great Plains. Since the PBLH simulated by MP2 is >60% higher than that simulated by CNTL over the southern Great Plains, we conducted another sensitivity MP2 run with $p = 3$ and $\alpha = 0.1$ (named MP2_$p3$_α.1). Reducing α further reduces the nocturnal warm bias in comparison to MP2_$p3$ (Fig. 11a), and the impact is limited to the southern Great Plains (Fig. 14c); that is, the improvement in the region does not compromise the performance in other regions. Some nocturnal warm bias remains, however, indicating there are further inaccuracies in the vertical mixing strength and/or other model errors.

2) Overly weak nighttime land–atmosphere coupling over short vegetation in the GFS surface layer scheme

In addition to PBL schemes, surface layer schemes, which determine the land–atmosphere coupling strength, also play an important role in accurately simulating near-surface variables (LeMone et al. 2008; Chen and Zhang 2009; LeMone et al. 2010; Zheng et al. 2012; Wang and Ma 2019; Gómez et al. 2020). In the MP2 forecast, the near-surface gradient of temperature ($T_2 - T_{sfc}$, where T_{sfc} is the surface/skin temperature) is significantly larger than that of the other forecast experiments over the southern Great Plains in regions of short vegetation (Figs. 12j,k). This suggests that downward sensible heat flux in the atmosphere as predicted by MP2 does not efficiently transport heat into the soil to increase T_{sfc}, but rather accumulates heat in the near-surface air (e.g., at 2 m AGL). In contrast, the near-surface gradient of temperature over the forested region just east of the southern Great Plains is much smaller in the MP2 simulation (Fig. 12k), suggesting there is reduced near-surface accumulation of heat over forests. Such a spatial contrast of $T_2 - T_{sfc}$ implies a marked spatial variability in the efficiency of heat transfer from air to soil (i.e., land–atmosphere coupling strength).

The surface exchange coefficient for heat (C_h), a quantitative measure of land–atmosphere coupling strength, drives the surface sensible heat flux (SH) through

$$SH = c_p \rho C_h U (T_{sfc} - T_2), \tag{3}$$

where ρ is the air density, c_p is the specific heat of air, and U is the surface wind speed. Over forests where the roughness

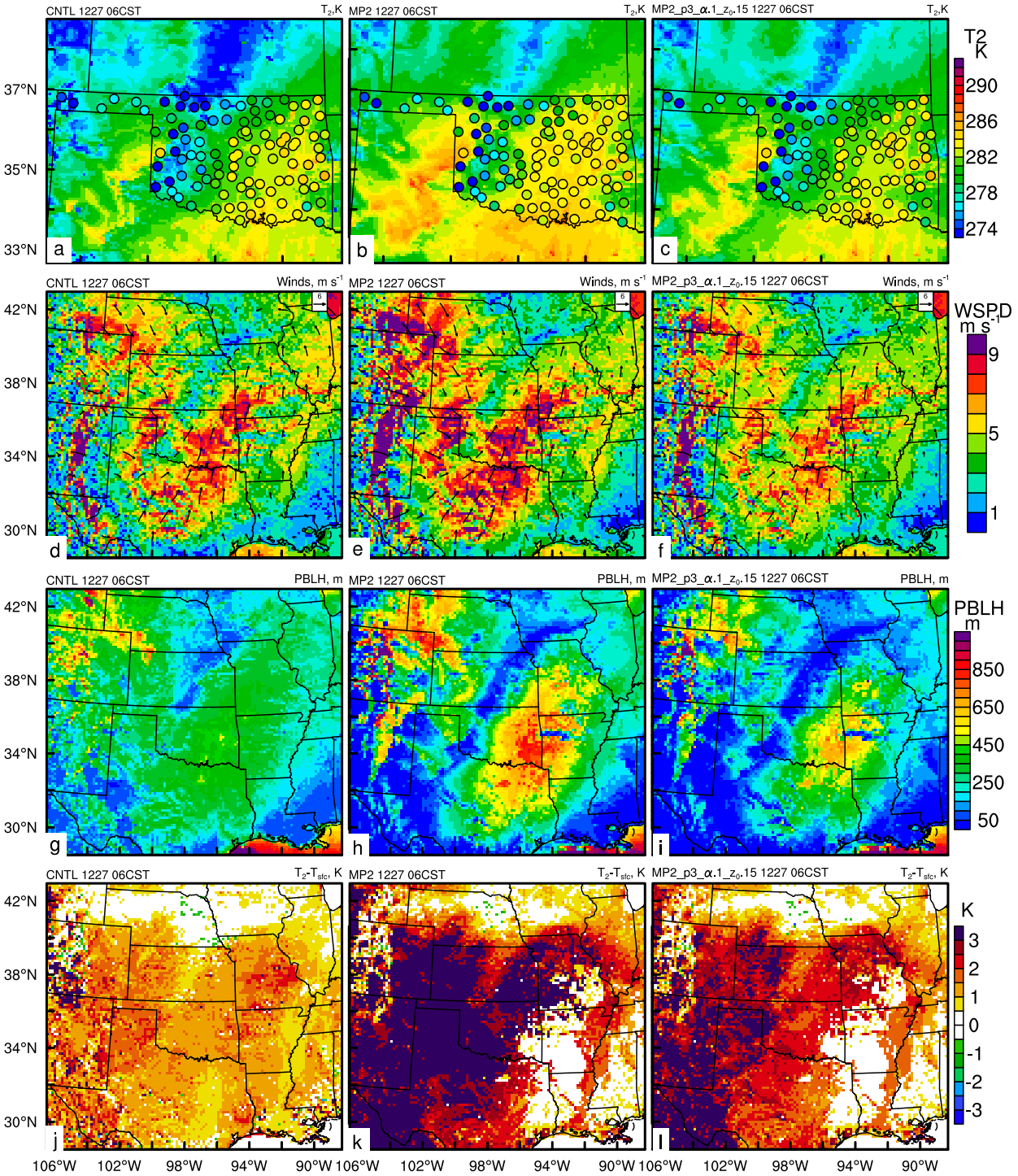

FIG. 12. Simulated (a)–(c) T_2, (d)–(f) surface winds, (g)–(i) PBL height (PBLH), and (j)–(l) near-surface temperature gradient ($T_2 - T_{sfc}$ at 0600 CST 27 Dec 2020 (1200 UTC; lead time 12 h) by the (left) CNTL, (center) MP2, and (right) MP2_$p3$_$\alpha.1$_$z_0.15$ simulations. Oklahoma Mesonet surface temperature data (shaded circles) are overlaid in (a)–(c).

length is large (Fig. 1), C_h can be 10 times as large as that over shorter vegetation, such as crops and grassland (Chen and Zhang 2009), matching the spatial distribution of $T_2 - T_{sfc}$ discussed above and confirming that the land–atmosphere coupling strength dictates the near-surface gradient of temperature. The spatial distribution of $T_2 - T_{sfc}$ in the MP2 simulation (Fig. 12k) closely resembles the T_2 difference between MP2 and the control simulation (which itself has a similar

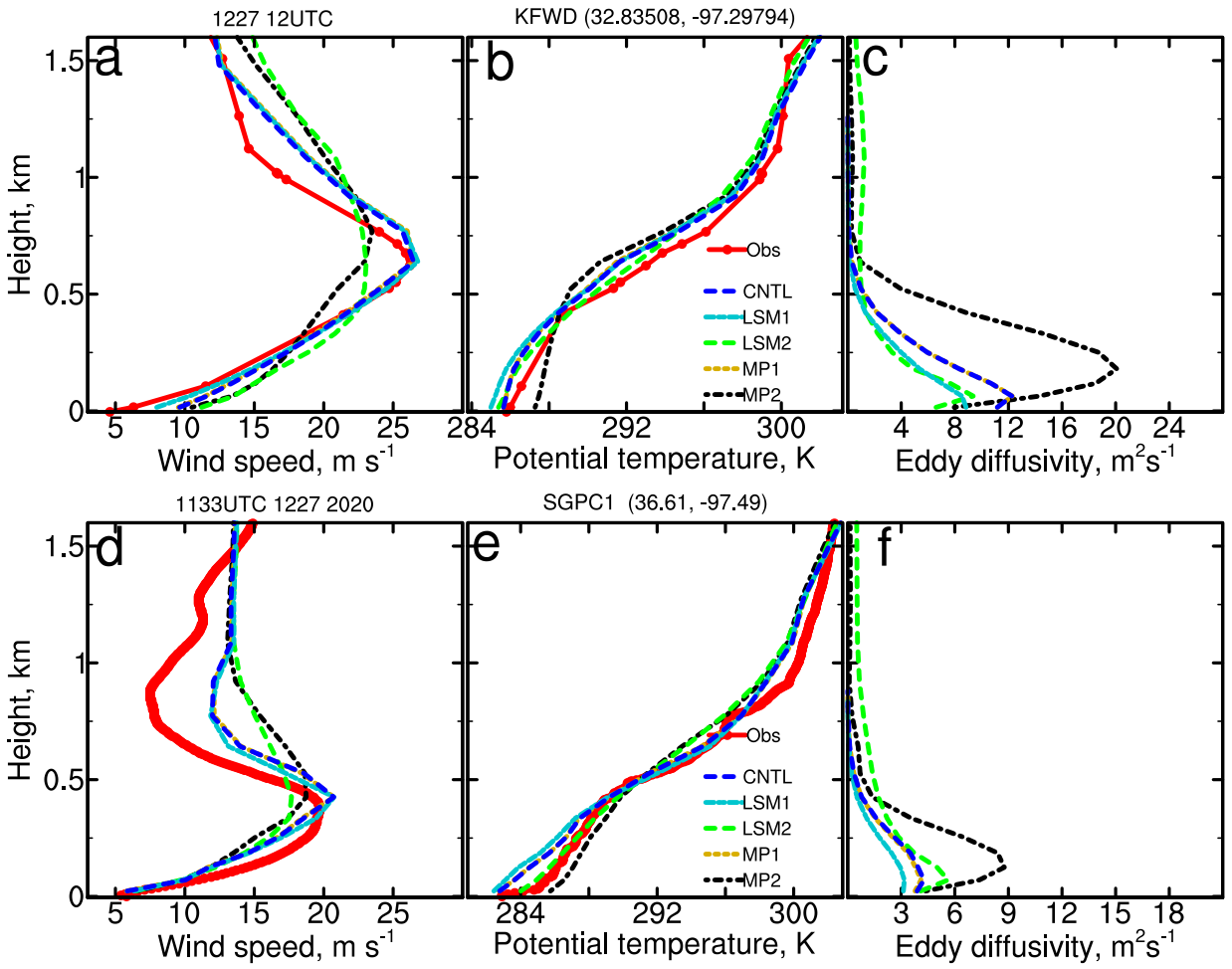

FIG. 13. Vertical profiles of (left) wind speed, (center) potential temperature, and (right) eddy diffusivity at (a)–(c) Fort Worth/Dallas, Texas (FWD), and (d)–(f) the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) site in Oklahoma.

distribution as MP2 T_2 bias against URMA) (Fig. 15). This indicates that the strength of the land–atmosphere coupling contributes to the nighttime warm bias from MP2 over short vegetation (i.e., grass and cropland) in addition to the overestimated vertical mixing by K-EDMF.

Surface fluxes and soil properties are further examined to verify the contribution of land–atmosphere coupling to the surface warm bias in MP2. K-EDMF overestimates downward sensible heat flux at least over the region with LLJs, covering both the southern Great Plains and the forest farther east, as indicated by overestimated PBLH (Figs. 12g,h). Over forested regions, where land–atmosphere coupling is strong, downward transfer of heat into soil is efficient, and thus the downward surface sensible heat flux, and consequently the soil temperature, is higher in the MP2 simulation than in the CNTL simulation (Fig. 15) under the circumstance of enhanced vertical mixing by K-EDMF. However, over the southern Great Plains, which is largely covered with grass vegetation, including short dormant winter wheat or fallow stubble fields during late December, downward surface sensible heat flux, and thus soil temperature, is lower in MP2 relative to CNTL (Figs. 15c,d), even in the presence of enhanced downward sensible heat flux in air layers in K-EDMF. This indicates the land–atmosphere coupling simulated by the GFS surface layer scheme in MP2 is underestimated at night. Other members using the GFS surface layer in the 15-member ensemble (e.g., M1B2L0 and M0B2L0), despite using a different PBL scheme and LSM from the MP2 experiment, also display a nocturnal warm bias over the Oklahoma Mesonet sites (Fig. 11b), which further corroborates the hypothesis of underestimated land–atmosphere coupling by the GFS surface layer as a major factor contributing to nighttime warm bias over short vegetation.

Sensitivity experiments adjusting the land–atmosphere coupling over grass based on the MP2 simulation (Table 3) were performed to confirm the conjecture. The land–atmosphere coupling strength is related to the surface aerodynamic roughness length (z_{0m}). In the GFS surface layer scheme, z_{0m} over grass (z_o) does not take the value from the FV3-LAM surface namelist input file (0.15 m), but rather is hard-coded as

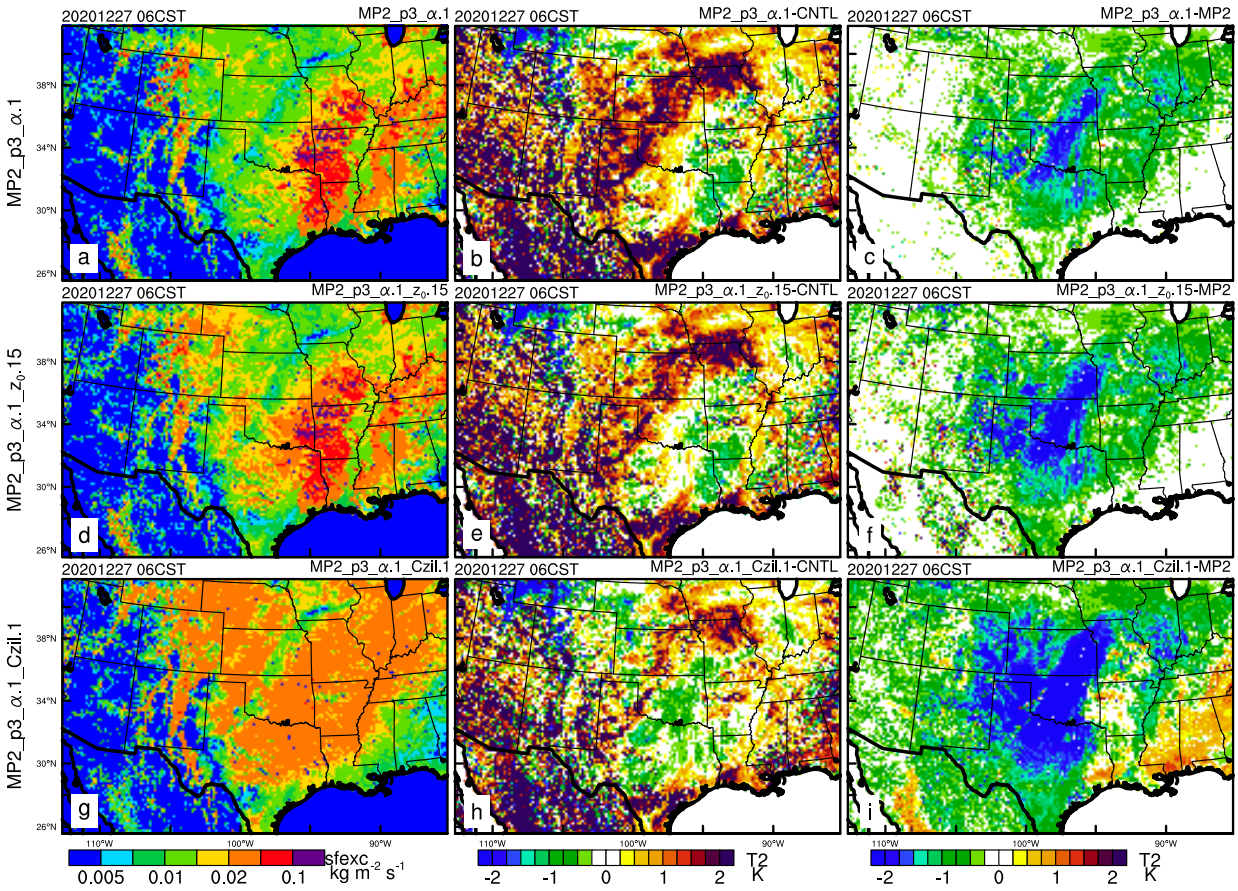

FIG. 14. Simulated (left) surface exchange coefficient ($\rho C_h U$), (center) difference of temperature at 2 m AGL (T_2) from the CNTL simulation, and (right) T_2 difference from the MP2 simulation, by three sensitivity simulations based on the MP2 configuration, referred to as (a)–(c) MP2_p3_α.1, (d)–(f) MP2_p3_α.1_z_0.15, and (g)–(i) MP2_p3_α.1_C_{zil}.1 respectively. See the description of the three MP2 sensitivity simulations in Table 3.

0.07 m following Dorman and Sellers (1989). In experiment MP2_p3_α.1_z_0.15 the z_o over grass is increased from 0.07 to 0.15 m. As a result, the surface exchange coefficient is slightly enhanced (Fig. 14d), and temperature over the Great Plains is reduced, but remains higher than in CNTL (Figs. 12a,c and 14e,f). Over western Oklahoma and Kansas where the surface exchange coefficient is still relatively small, the warm bias is still larger than over eastern Oklahoma and Texas (Fig. 14e). The remaining warm bias indicates that simply increasing z_{0m} over grass in the current framework of the GFS surface layer scheme is not enough to solve the issue of the nocturnal warm bias.

In the surface layer schemes used in this study, MYNN and GFS, the roughness length for heat (z_{0h}) is related to the roughness length for momentum (z_{0m}) through Eq. (1). In the GFS surface layer scheme (Zheng et al. 2012), the Zilitinkevich coefficient (C_{zil}) uses a vegetation height (h)-dependent function (Chen and Zhang 2009), while the MYNN surface layer scheme uses a constant C_{zil} of 0.1 (Chen et al. 1997) derived from calibration against the long-term FIFE (Sellers et al. 1992) and HAPEX (André et al. 1986) observations. A vegetation-dependent C_{zil} performs better over certain vegetation types during daytime hours. However, during other times of day (e.g., early morning and evening) when wind speeds are low compared to daytime and the land–atmospheric coupling is stronger (Sun 1999; Chen et al. 2019), a vegetation-dependent C_{zil} appears to underestimate the land–atmosphere coupling strength over short vegetation (Chen et al. 2019). A constant C_{zil} of 0 (thus stronger land–atmosphere coupling) appears to perform better over grassland during times other than midday (Chen et al. 2019). Thus, in experiment MP2_p3_α.1_C_{zil}.1, the vegetation-dependent C_{zil} in the GFS surface layer scheme is replaced with a constant C_{zil} of 0.1 as in the default MYNN surface layer scheme (see the configuration summarized in Table 3). As a result, the surface exchange coefficient over short vegetation is enhanced over the Great Plains and reduced over forests (Fig. 14g). Overall, the spatial distribution of the surface exchange coefficient simulated by MP2_p3_α.1_C_{zil}.1 is similar to that simulated by CNTL (not shown). Consequently, the warm bias is largely alleviated and the temperature over forests is slightly increased (Figs. 14h,i). These diagnoses suggest that the vegetation-dependent C_{zil}

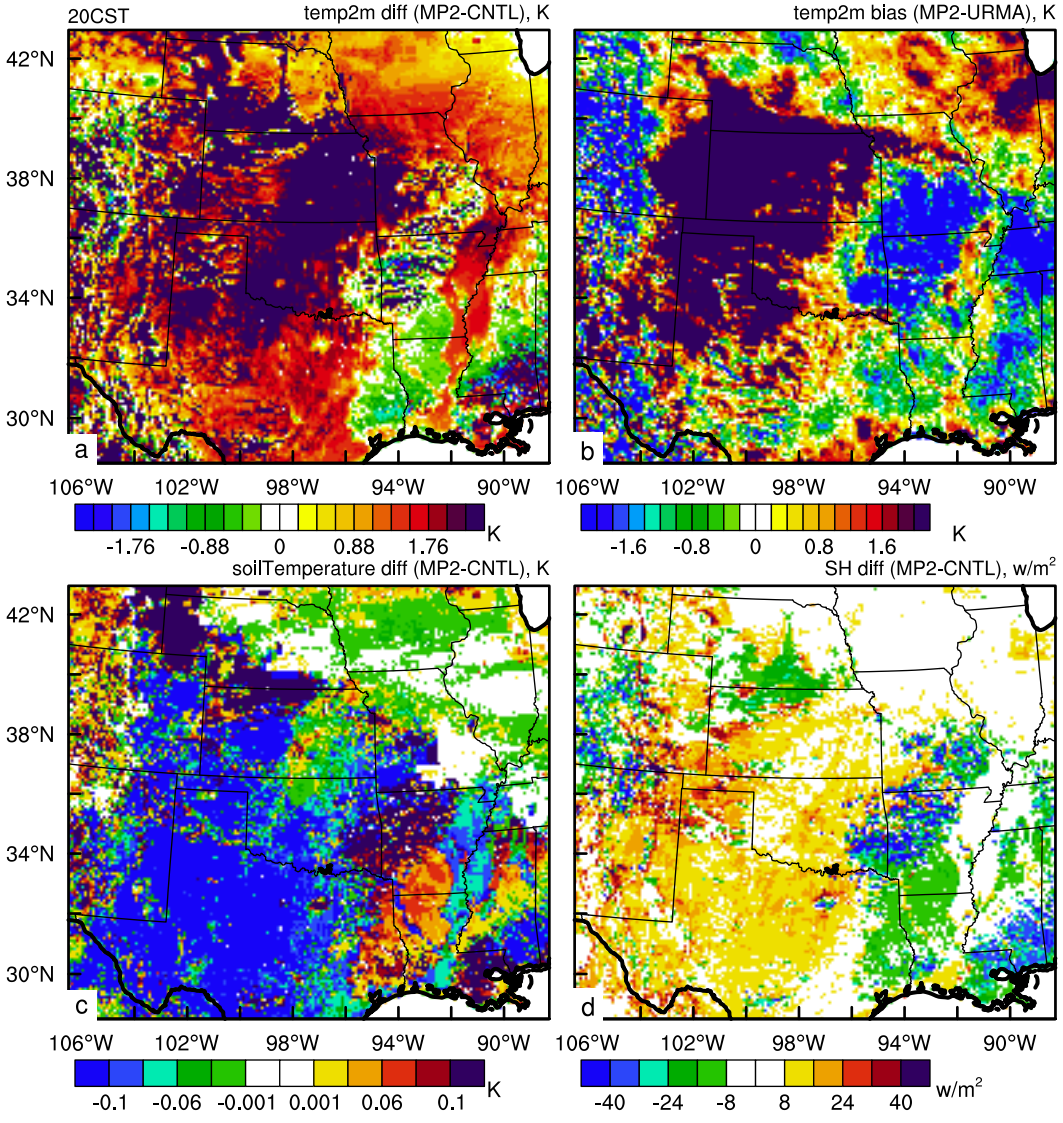

FIG. 15. Difference between MP2 and CNTL simulations of (a) temperature at 2 m AGL (T_2), (c) soil temperature, and (d) surface sensible heat flux, as well as (b) T_2 bias against URMA during the night of 26–27 Dec 2020 (lead time 2 h).

currently adopted by the GFS surface layer underestimates land surface coupling over short vegetation during the night, thus leading to a near-surface warm bias. Using a stronger coupling with a constant C_{zil} of 0.1 can help to ameliorate the nighttime warm bias.

To test the robustness of these changes, we reran all 35 cases during winter 2020/21 using MP2_p3_α.1_C_{zil}.1. In these runs, the warm bias over the southern Great Plains is substantially reduced over the whole winter in MP2_p3_α.1_C_{zil}.1 compared to the original MP2 forecasts (Fig. 16b), and the impact on other regions (e.g., the Northeast) is minor (Fig. 16a).

This study diagnoses model errors of cold start forecasts at a convection allowing scale in the presence of initial condition errors. Examining and understanding of the sensitivity of cold start short-term weather forecasting to different physics schemes/suites starting from the same initial conditions remains critical for calibration/improvement of physics schemes. We do need to make sure that the physics sensitivity is not skewed by systematic initial condition errors, such as those in the initial soil state. Typically, model error becomes more dominant in the forecast error as the forecast lead time increases. In our case, the model bias/errors remain similar in the latter stages of our 84-h forecasts (Fig. 9), suggesting that the model bias/errors identified in this study are the primary source, and are not artifacts of biases in the GFS-based initial conditions. We also repeated FV3-LAM simulations for the 27 December 2020 case starting from the operational HRRR analysis. These simulations (not shown) demonstrate that 1) LSM2 has a similar nighttime cold bias over snow regions and 2) MP2

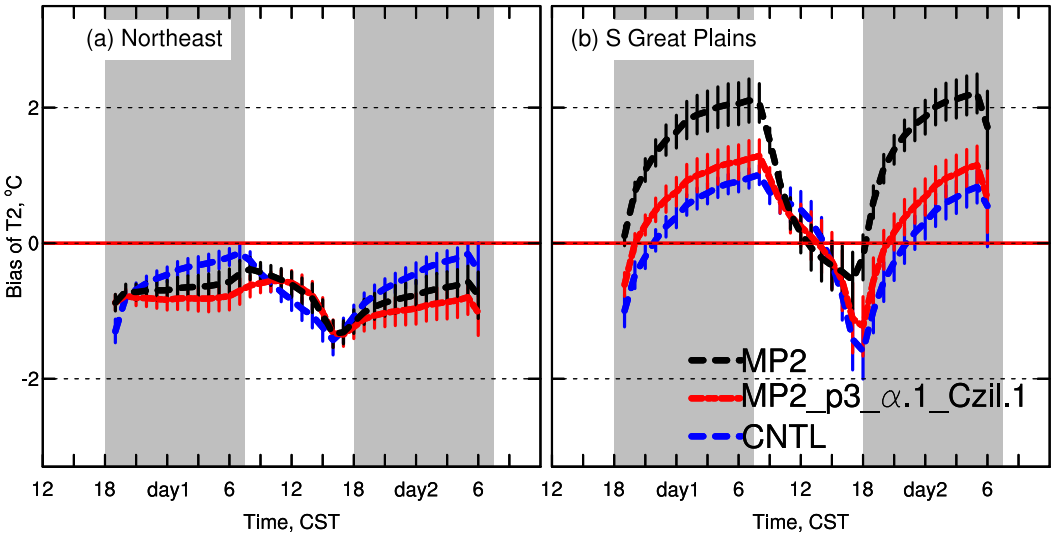

FIG. 16. Bias of 2-m temperature (T_2) and its 95% bootstrap confidence interval from simulations of CNTL, MP2, and MP2_p3_α.1_C_{zil}.1 in the first 36-h of forecasts during winter of 2020/21 evaluated against the Unrestricted Mesoscale Analysis (URMA) over the (a) Northeast and (b) southern Great Plains.

has a significant nighttime warm bias; our proposed modification (i.e., MP2_p3_α.1_C_{zil}.1) also reduces this bias. The robustness of these results across initializing analyses indicates that the proposed adjustments address systematic errors within the model itself.

4. Conclusions and discussion

A set of FV3-LAM forecasts produced by CAPS is evaluated against URMA analyses and selected mesoscale point observations to guide the selection of physics parameterizations for operational implementation of the future RRFS and guide the choice of suitable physics suites for a multiphysics RRFS ensemble. These forecasts use five common physics suites and were run over the contiguous United States at ~3-km grid spacing for 35 cases during the 11th HMT WWE, which took place from October 2020 through March 2021. Full-season evaluation reveals a systematic near-surface cold bias over the northeast United States from forecast member LSM2 using the Noah-MP land surface model and a nighttime warm bias over the southern Great Plains from the MP2 suite with the K-EDMF PBL scheme and GFS surface layer scheme.

A systematic study using a representative case (27 December 2020) is conducted to diagnose the root cause of these near-surface biases by running several additional sensitivity experiments with different combinations of physics parameterizations where experiments differ in one scheme at a time, or with modified parameter values. Additionally, mesoscale point observations from the Oklahoma and New York State Mesonets are used to help verify these near-surface biases. The treatment of snow cover in the Noah-MP LSM is identified as the primary source of the large cold biases from the LSM2 suite. Over regions of snow cover, Noah-MP simulates lower soil water content, including both soil moisture and soil liquid water, leading to lower thermal conductivity and consequently smaller upward ground flux during nighttime. As a result, Noah-MP underestimates surface temperature over snow. Overly high snow emissivity (1.0, as opposed to 0.95 adopted by other LSMs), a factor noted by He et al. (2019), is a secondary source for the cold bias, only contributing a bias of ~−0.2°C over snow (<10% of total bias) in this case.

The nighttime warm bias seen in the MP2 forecast experiment over the southern Great Plains is attributed to an overestimated vertical mixing strength from the K-EDMF PBL scheme and overly weak land–atmospheric coupling from the GFS surface layer scheme over grassland. Adjusting a few key parameters in the physics parameterization schemes can alleviate this warm bias. Such parameter changes include reducing vertical mixing by increasing the exponent p in the polynomial function of vertical diffusivity, reducing the critical bulk Richardson number in their plausible ranges in K-EDMF, and using a constant C_{zil} of 0.1 in the GFS surface layer scheme instead of a vegetation-dependent C_{zil}. This matches what is used in the MYNN surface layer scheme. A constant C_{zil} of 0.1 enhances land–atmosphere coupling over short vegetation such as grass and cropland and thus alleviates nighttime near-surface accumulation of downward sensible heat flux over the southern Great Plains in MP2.

Previous studies of the land–atmosphere coupling strength focused on the daytime convective boundary layer during the warm season (Mitchell 2004; Trier et al. 2004; LeMone et al. 2010; Zheng et al. 2015; Li et al. 2017; Chen et al. 2019; Zhang et al. 2021), which suggest that vegetation-dependent C_{zil} performs better and simulates smaller daytime sensible heat flux over short vegetation compared to C_{zil} = 0.1; while vegetation-dependent C_{zil} overestimates daytime sensible heat flux over

forests (Chen et al. 2019). During other times of day (e.g., early morning and evening), when wind speeds are low compared to daytime and the land–atmosphere coupling is stronger (Sun 1999; Chen et al. 2019), vegetation-dependent C_{zil} appears to underestimate the land–atmosphere coupling strength over short vegetation (Chen et al. 2019). An appropriate C_{zil} is therefore critical for accurately simulating near-surface meteorological variables for FV3-LAM forecasts in this study and thus warrants careful calibration in the future, particularly over short vegetation.

Optimization of the critical parameters within PBL schemes requires careful calibration using long-term observations such as soundings (e.g., Hu et al. 2019a) or large-eddy simulations (e.g., Shin and Hong 2015). Existing parameter settings may have had only limited testing. For example, the critical bulk Richardson number (Ri) used to determine the stable boundary layer in K-EDMF via Eq. (2) was calibrated in situations where observed PBLH were mostly <150 m (Vickers and Mahrt 2004). This parameter likely needs recalibration over additional climate regions such as the southern Great Plains, where nighttime PBLH reaches as high as 800 m. This cannot be derived from traditional tower data, and thus its detection requires more nighttime soundings or more advanced instruments such as lidar (Bonin et al. 2015; Klein et al. 2016; Bonin et al. 2020) or uncrewed aerial vehicles (e.g., Bonin et al. 2013).

The systematic biases in certain potential RRFS model physics schemes and the potential solutions diagnosed in this study have been communicated with scientists from National Center for Atmospheric Research (NCAR), GSL, and EMC, as a part of the development of the first operational implementation of the RRFS ensemble forecast system.

Acknowledgments. This work was supported by the NOAA Unified Forecast System Research-to-Operations (UFS-R2O) Project through Grants NA16OAR4320115 and NA21OAR4320204, and real-time forecast production and evaluation were supported by NOAA OAR Testbed Grant NA19OAR4590141. The first author is partially supported by the National Mesonet Program Grant 10558200 and DOE ASR project (DE-SC0021159). Ming Xue was also supported by NSF Grant AGS-1917701. We are grateful to Fei Chen and Cenlin He for discussion. The real-time forecasts and diagnosing sensitivity simulations were performed on supercomputers Frontera (Stanzione et al. 2020) and Stampede 2 at the Texas Advanced Computing Center (TACC), from NSF Extreme Science and Engineering Discovery Environment (XSEDE) program through allocations TG-MCA95C006, TG-ATM160014, and Frontera ATM20008. Three anonymous reviewers provided helpful comments that improved the manuscript.

Data availability statement. The URMA data were provided by Matthew Morris from NOAA/EMC. Oklahoma and New York State Mesonet data are retrieved from https://www.mesonet.org/ and http://nysmesonet.org, respectively. The Radiosonde Replacement System (RRS) sounding data were provided by Xungang Yin from NOAA. The Atmospheric Radiation Measurement (ARM) soundings were obtained from https://doi.org/10.5439/1021460. ARM is sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Climate and Environmental Sciences Division. Most of the forecast surface and precipitation products from the five real-time forecasts were sent to NOAA Hydrometeorology Testbed Winter Weather Experiment in real time for subjective evaluation. RRFS outputs and sensitivity simulation outputs produced from this study have been archived on the Ranch storage system at the Texas Advanced Computing Center (TACC) and important 2D fields are archived at CAPS and available at https://doi.org/10.15763/DBS.CAPS.001.

REFERENCES

Aksoy, A., F. Zhang, and J. W. Nielsen-Gammon, 2006: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Mon. Wea. Rev.*, **134**, 2951–2970, https://doi.org/10.1175/MWR3224.1.

Alexander, C., and J. R. Carley, 2021: Convection-allowing models: Short-range weather in operations. *UFS Quarterly*, winter 2020/21, 7–9, https://doi.org/10.25923/k3zn-xe66.

Aligo, E. A., B. Ferrier, and J. R. Carley, 2018: Modified NAM microphysics for forecasts of deep convective storms. *Mon. Wea. Rev.*, **146**, 4115–4153, https://doi.org/10.1175/MWR-D-17-0277.1.

André, J.-C., J.-P. Goutorbe, and A. Perrier, 1986: HAPEX–MOBILHY: A hydrologic atmospheric experiment the study of water budget and evaporation flux at the climatic scale. *Bull. Amer. Meteor. Soc.*, **67**, 138–144, https://doi.org/10.1175/1520-0477-67.2.138.

Banos, I. H., W. D. Mayfield, G. Ge, L. F. Sapucci, J. R. Carley, and L. Nance, 2022: Assessment of the data assimilation framework for the Rapid Refresh Forecast System v0.1 and impacts on forecasts of convective storms. *Geosci. Model Dev.*, **15**, 6891–6917, https://doi.org/10.5194/gmd-15-6891-2022.

Barlage, M., R. Yang, H. Wei, Y. Wu, W. Zheng, J. W. Bao, and V. Tallapragada, 2020: The hierarchical testing of Noah-MP as a candidate land model for the Unified Forecast System. *2020 Fall Meeting*, Online, Amer. Geophys. Union, Abstract H201-208.

Basara, J. B., P. K. Hall Jr., A. J. Schroeder, B. G. Illston, and K. L. Nemunaitis, 2008: Diurnal cycle of the Oklahoma City urban heat island. *J. Geophys. Res.*, **113**, D20109, https://doi.org/10.1029/2008JD010311.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The rapid refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Biswas, M. K., and Coauthors, 2018: Hurricane Weather Research and Forecasting (HWRF) model: 2018 scientific documentation. NCAR Tech. Note NCAR/TN-544+STR, 111 pp., https://opensky.ucar.edu/islandora/object/technotes%3A563/datastream/PDF/view.

Black, T. L., and Coauthors, 2021: A limited area modeling capability for the finite-volume cubed-sphere (FV3) dynamical core and comparison with a global two-way nest. *J. Adv. Model. Earth Syst.*, **13**, e2021MS002483, https://doi.org/10.1029/2021MS002483.

Blackadar, A. K., 1957: Boundary layer wind maxima and their significance for the growth of nocturnal inversions. *Bull. Amer. Meteor. Soc.*, **38**, 283–290, https://doi.org/10.1175/1520-0477-38.5.283.

Bonin, T., P. Chilson, B. Zielke, and E. Fedorovich, 2013: Observations of the early evening boundary-layer transition using a small unmanned aerial system. *Bound.-Layer Meteor.*, **146**, 119–132, https://doi.org/10.1007/s10546-012-9760-3.

——, 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.

——, P. M. Klein, and P. B. Chilson, 2020: Contrasting characteristics and evolution of southerly low-level jets during different boundary-layer regimes. *Bound.-Layer Meteor.*, **174**, 179–202, https://doi.org/10.1007/s10546-019-00481-0.

Brotzge, J. A., and Coauthors, 2020: A technical overview of the New York State Mesonet standard network. *J. Atmos. Oceanic Technol.*, **37**, 1827–1845, https://doi.org/10.1175/JTECH-D-19-0220.1.

Chaney, N. W., J. D. Herman, M. B. Ek, and E. F. Wood, 2016: Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning. *J. Geophys. Res. Atmos.*, **121**, 13 218–13 235, https://doi.org/10.1002/2016JD024821.

Chen, F., and Y. Zhang, 2009: On the coupling strength between the land surface and the atmosphere: From viewpoint of surface exchange coefficients. *Geophys. Res. Lett.*, **36**, L10404, https://doi.org/10.1029/2009GL037980.

——, Z. Janjic, and K. Mitchell, 1997: Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP Mesoscale Eta model. *Bound.-Layer Meteor.*, **85**, 391–421, https://doi.org/10.1023/A:1000531001463.

Chen, L., Y. Li, F. Chen, M. Barlage, Z. Zhang, and Z. Li, 2019: Using 4-km WRF CONUS simulations to assess impacts of the surface coupling strength on regional climate simulation. *Climate Dyn.*, **53**, 6397–6416, https://doi.org/10.1007/s00382-019-04932-9.

De Pondeca, M. S. F. V., and Coauthors, 2011: The real-time mesoscale analysis at NOAA's National Centers for Environmental Prediction: Current status and development. *Wea. Forecasting*, **26**, 593–612, https://doi.org/10.1175/WAF-D-10-05037.1.

Derbyshire, S. H., 1999: Boundary-layer decoupling over cold surfaces as a physical boundary instability. *Bound.-Layer Meteor.*, **90**, 297–325, https://doi.org/10.1023/A:1001710014316.

Dorman, J. L., and P. J. Sellers, 1989: A global climatology of albedo, roughness length and stomatal resistance for atmospheric general circulation models as represented by the Simple Biosphere Model (SiB). *J. Appl. Meteor. Climatol.*, **28**, 833–855, https://doi.org/10.1175/1520-0450(1989)028<0833:AGCOAR>2.0.CO;2.

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.

Firl, G., L. Carson, M. Harrold, L. Bernardet, and D. Heinzeller, 2021: Common community physics package Single Column Model (SCM) v5.0.0 user and technical guide. Developmental Testbed Center, 39 pp., https://dtcenter.org/GMTB/v5.0.0/scm-ccpp-guide-v5.0.0.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.

Gómez, I., V. Caselles, and M. J. Estrela, 2020: Improving RAMS and WRF mesoscale forecasts over two distinct vegetation covers using an appropriate thermal roughness length parameterization. *Agric. For. Meteor.*, **280**, 107791, https://doi.org/10.1016/j.agrformet.2019.107791.

Han, J., and H.-L. Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Wea. Forecasting*, **26**, 520–533, https://doi.org/10.1175/WAF-D-10-05038.1.

——, 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.

Hansen, J. A., 2002: Accounting for model error in ensemble-based state estimation and forecasting. *Mon. Wea. Rev.*, **130**, 2373–2391, https://doi.org/10.1175/1520-0493(2002)130<2373:AFMEIE>2.0.CO;2.

——, and C. Penland, 2007: On stochastic parameter estimation using data assimilation. *Physica D*, **230**, 88–98, https://doi.org/10.1016/j.physd.2006.11.006.

Harnos, K., J. Correia, B. Albright, M. Bodner, and J. Nelson, 2021: 11th annual winter weather experiment: Findings and results. NOAA Rep., 49 pp., https://www.wpc.ncep.noaa.gov/hmt/11th_Annual_HMT_WWE_2021_Final_Report.pdf.

Harris, L., X. Chen, W. Putman, L. Zhou, and J.-H. Chen, 2021: A scientific description of the GFDL finite-volume cubed-sphere dynamical core. NOAA Tech. Memo. GFDL2021001, 109 pp., https://doi.org/10.25923/6nhs-5897.

Hazelton, A., and Coauthors, 2021: 2019 Atlantic hurricane forecasts from the global-nested hurricane analysis and forecast system: Composite statistics and key events. *Wea. Forecasting*, **36**, 519–538, https://doi.org/10.1175/WAF-D-20-0044.1.

He, C., F. Chen, M. Barlage, C. Liu, A. Newman, W. Tang, K. Ikeda, and R. Rasmussen, 2019: Can convection-permitting modeling provide decent precipitation for offline high-resolution snowpack simulations over mountains? *J. Geophys. Res. Atmos.*, **124**, 12 631–12 654, https://doi.org/10.1029/2019JD030823.

Hu, X.-M., and M. Xue, 2016: Influence of synoptic sea-breeze fronts on the urban heat island intensity in Dallas–Fort Worth, Texas. *Mon. Wea. Rev.*, **144**, 1487–1507, https://doi.org/10.1175/MWR-D-15-0201.1.

——, J. W. Nielsen-Gammon, and F. Zhang, 2010a: Evaluation of three planetary boundary layer schemes in the WRF model. *J. Appl. Meteor. Climatol.*, **49**, 1831–1844, https://doi.org/10.1175/2010JAMC2432.1.

——, F. Zhang, and J. W. Nielsen-Gammon, 2010b: Ensemble-based simultaneous state and parameter estimation for treatment of mesoscale model error: A real-data study. *Geophys. Res. Lett.*, **37**, L08802, https://doi.org/10.1029/2010GL043017.

——, P. M. Klein, and M. Xue, 2013a: Evaluation of the updated YSU planetary boundary layer scheme within WRF for wind resource and air quality assessments. *J. Geophys. Res. Atmos.*, **118**, 10 490–10 505, https://doi.org/10.1002/jgrd.50823.

——, ——, ——, A. Shapiro, and A. Nallapareddy, 2013b: Enhanced vertical mixing associated with a nocturnal cold front passage and its impact on near-surface temperature and ozone concentration. *J. Geophys. Res. Atmos.*, **118**, 2714–2728, https://doi.org/10.1002/jgrd.50309.

——, ——, ——, J. K. Lundquist, F. Zhang, and Y. Qi, 2013c: Impact of low-level jets on the nocturnal urban heat island intensity in Oklahoma City. *J. Appl. Meteor. Climatol.*, **52**, 1779–1802, https://doi.org/10.1175/JAMC-D-12-0256.1.

——, M. Xue, P. M. Klein, B. G. Illston, and S. Chen, 2016: Analysis of urban effects in Oklahoma City using a dense surface observing network. *J. Appl. Meteor. Climatol.*, **55**, 723–741, https://doi.org/10.1175/JAMC-D-15-0206.1.

——, ——, and X. Li, 2019a: The use of high-resolution sounding data to evaluate and optimize nonlocal PBL schemes for simulating the slightly stable upper convective boundary layer. *Mon. Wea. Rev.*, **147**, 3825–3841, https://doi.org/10.1175/MWR-D-19-0085.1.

——, ——, F. Kong, and H. Zhang, 2019b: Meteorological conditions during an ozone episode in Dallas–Fort Worth, Texas, and impact of their modeling uncertainties on air quality prediction. *J. Geophys. Res. Atmos.*, **124**, 1941–1961, https://doi.org/10.1029/2018JD029791.

——, and Coauthors, 2020: Dynamical downscaling of CO_2 in 2016 over the contiguous United States using WRF-VPRM, a weather-biosphere-online-coupled model. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001875, https://doi.org/10.1029/2019MS001875.

——, J. Hu, L. Gao, C. Cai, Y. Jiang, M. Xue, T. Zhao, and S. M. R. Crowell, 2021: Multisensor and multimodel monitoring and investigation of a wintertime air pollution event ahead of a cold front over eastern China. *J. Geophys. Res. Atmos.*, **126**, e2020JD033538, https://doi.org/10.1029/2020JD033538.

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.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Jung, Y., M. Xue, and G. Zhang, 2010: Simultaneous estimation of microphysical parameters and the atmospheric state using simulated polarimetric radar data and an ensemble Kalman filter in the presence of an observation operator error. *Mon. Wea. Rev.*, **138**, 539–562, https://doi.org/10.1175/2009MWR2748.1.

Klein, P. M., X.-M. Hu, A. Shapiro, and M. Xue, 2016: Linkages between boundary-layer structure and the development of nocturnal low-level jets in central Oklahoma. *Bound.-Layer Meteor.*, **158**, 383–408, https://doi.org/10.1007/s10546-015-0097-6.

LeMone, M. A., M. Tewari, F. Chen, J. G. Alfieri, and D. Niyogi, 2008: Evaluation of the Noah land surface model using data from a fair-weather IHOP_2002 day with heterogeneous surface fluxes. *Mon. Wea. Rev.*, **136**, 4915–4941, https://doi.org/10.1175/2008MWR2354.1.

——, F. Chen, M. Tewari, J. Dudhia, B. Geerts, Q. Miao, R. L. Coulter, and R. L. Grossman, 2010: Simulating the IHOP_2002 fair-weather CBL with the WRF-ARW–Noah modeling system. Part I: Surface fluxes and CBL structure and evolution along the eastern track. *Mon. Wea. Rev.*, **138**, 722–744, https://doi.org/10.1175/2009MWR3003.1.

Li, D., A. Rigden, G. Salvucci, and H. Liu, 2017: Reconciling the Reynolds number dependence of scalar roughness length and laminar resistance. *Geophys. Res. Lett.*, **44**, 3193–3200, https://doi.org/10.1002/2017GL072864.

Lin, S.-J., 2004: A "vertically Lagrangian" finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307, https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2.

Liu, Z., S. Zhang, Y. Shen, Y. Guan, and X. Deng, 2021: A study of capturing Atlantic meridional overturning circulation (AMOC) regime transition through observation-constrained model parameters. *Nonlinear Processes Geophys.*, **28**, 481–500, https://doi.org/10.5194/npg-28-481-2021.

Mansell, E. R., 2010: On sedimentation and advection in multimoment bulk microphysics. *J. Atmos. Sci.*, **67**, 3084–3094, https://doi.org/10.1175/2010JAS3341.1.

——, and C. L. Ziegler, 2013: Aerosol effects on simulated storm electrification and precipitation in a two-moment bulk microphysics model. *J. Atmos. Sci.*, **70**, 2032–2050, https://doi.org/10.1175/JAS-D-12-0264.1.

McPherson, R. A., and Coauthors, 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, **24**, 301–321, https://doi.org/10.1175/JTECH1976.1.

Mitchell, K. E., 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.*, **109**, D07S90, https://doi.org/10.1029/2003JD003823

Morris, M. T., J. R. Carley, E. Colón, A. Gibbs, M. S. F. V. De Pondeca, and S. Levine, 2020: A quality assessment of the Real-Time Mesoscale Analysis (RTMA) for aviation. *Wea. Forecasting*, **35**, 977–996, https://doi.org/10.1175/WAF-D-19-0201.1.

Morrison, H., and A. Gettelman, 2008: A new two-moment bulk stratiform cloud microphysics scheme in the Community Atmosphere Model, version 3 (CAM3). Part I: Description and numerical tests. *J. Climate*, **21**, 3642–3659, https://doi.org/10.1175/2008JCLI2105.1.

Nakanishi, M., and H. Niino, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87**, 895–912, https://doi.org/10.2151/jmsj.87.895.

Nielsen-Gammon, J. W., X.-M. Hu, F. Zhang, and J. E. Pleim, 2010: Evaluation of planetary boundary layer scheme sensitivities for the purpose of parameter estimation. *Mon. Wea. Rev.*, **138**, 3400–3417, https://doi.org/10.1175/2010MWR3292.1.

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.

Nystrom, R. G., S. J. Greybush, X. Chen, and F. Zhang, 2021: Potential for new constraints on tropical cyclone surface-exchange coefficients through simultaneous ensemble-based state and parameter estimation. *Mon. Wea. Rev.*, **149**, 2213–2230, https://doi.org/10.1175/MWR-D-20-0259.1.

Ochsner, T., E. Howerton, and B. Ellis, 2019: *Rain or Shine: An Introduction to Soil Physical Properties and Processes.* Oklahoma State University, 312 pp.

Olson, J. B., J. S. Kenyon, W. A. Angevine, J. M. Brown, M. Pagowski, and K. Sušelj, 2019a: A description of the MYNN-EDMF scheme and the coupling to other components in WRF–ARW. NOAA Tech. Memo. OAR GSD-61, 42 pp., https://doi.org/10.25923/n9wm-be49.

——, and Coauthors, 2019b: Improving wind energy forecasting through numerical weather prediction model development.

Bull. Amer. Meteor. Soc., **100**, 2201–2220, https://doi.org/10.1175/BAMS-D-18-0040.1.

Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, https://doi.org/10.1016/j.jcp.2007.07.022.

Sellers, P. J., F. G. Hall, G. Asrar, D. E. Strebel, and R. E. Murphy, 1992: An overview of the First International Satellite Land Surface Climatology Project (ISLSCP) Field Experiment (FIFE). *J. Geophys. Res.*, **97**, 18 345–18 371, https://doi.org/10.1029/92JD02111.

Shapiro, A., and E. Fedorovich, 2010: Analytical description of a nocturnal low-level jet. *Quart. J. Roy. Meteor. Soc.*, **136**, 1255–1262, https://doi.org/10.1002/qj.628.

Shen, C., and Coauthors, 2022: Spatializing the roughness length of heterogeneous urban underlying surfaces to improve the WRF simulation—Part 1: A review of morphological methods and model evaluation. *Atmos. Environ.*, **270**, 118874, https://doi.org/10.1016/j.atmosenv.2021.118874.

Shin, H. H., and S.-Y. Hong, 2015: Representation of the subgrid-scale turbulent transport in convective boundary layers at gray-zone resolutions. *Mon. Wea. Rev.*, **143**, 250–271, https://doi.org/10.1175/MWR-D-14-00116.1.

Shrestha, B., J. A. Brotzge, J. Wang, N. Bain, C. D. Thorncroft, E. Joseph, J. Freedman, and S. Perez, 2021: Overview and applications of the New York State Mesonet Profiler Network. *J. Appl. Meteor. Climatol.*, **60**, 1591–1611, https://doi.org/10.1175/JAMC-D-21-0104.1.

Smirnova, T. G., J. M. Brown, S. G. Benjamin, and D. Kim, 2000: Parameterization of cold-season processes in the MAPS land-surface scheme. *J. Geophys. Res.*, **105**, 4077–4086, https://doi.org/10.1029/1999JD901047.

——, ——, ——, and J. S. Kenyon, 2016: Modifications to the Rapid update Cycle Land Surface Model (RUC LSM) available in the Weather Research and Forecasting (WRF) model. *Mon. Wea. Rev.*, **144**, 1851–1865, https://doi.org/10.1175/MWR-D-15-0198.1.

Stanzione, D., J. West, R. T. Evans, T. Minyard, O. Ghattas, and D. K. Panda, 2020: Frontera: The evolution of leadership computing at the National Science Foundation. *Proc. PEARC'20 (Practice and Experience in Advanced Research Computing)*, Portland, OR, Association for Computing Machinery, 106–111, https://doi.org/10.1145/3311790.3396656.

Sun, J., 1999: Diurnal variations of thermal roughness height over a grassland. *Bound.-Layer Meteor.*, **92**, 407–427, https://doi.org/10.1023/A:1002071421362.

Supinie, T. A., J. Park, N. Snook, X.-M. Hu, K. Brewster, M. Xue, and J. R. Carley, 2022: Cool-season evaluation of FV3-LAM-based CONUS-scale forecasts with physics configurations of a prototype RRFS ensemble. *Mon. Wea. Rev*, **150**, 2379–2398, https://doi.org/10.1175/MWR-D-21-0331.1.

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.

Tong, M. J., and M. Xue, 2008: Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part II: Parameter estimation experiments. *Mon. Wea. Rev.*, **136**, 1649–1668, https://doi.org/10.1175/2007MWR2071.1.

Trier, S. B., F. Chen, and K. W. Manning, 2004: A study of convection initiation in a mesoscale model using high-resolution land surface initial conditions. *Mon. Wea. Rev.*, **132**, 2954–2976, https://doi.org/10.1175/MWR2839.1.

——, M. A. LeMone, F. Chen, and K. W. Manning, 2011: Effects of surface heat and moisture exchange on ARW-WRF warm-season precipitation forecasts over the central United States. *Wea. Forecasting*, **26**, 3–25, https://doi.org/10.1175/2010WAF2222426.1.

Troen, I., and L. Mahrt, 1986: A simple model of the atmospheric boundary layer; sensitivity to surface evaporation. *Bound.-Layer Meteor.*, **37**, 129–148, https://doi.org/10.1007/BF00122760.

Vickers, D., and L. Mahrt, 2004: Evaluating formulations of stable boundary layer height. *J. Appl. Meteor.*, **43**, 1736–1749, https://doi.org/10.1175/JAM2160.1.

Wang, S., and Y. Ma, 2019: On the simulation of sensible heat flux over the Tibetan Plateau using different thermal roughness length parameterization schemes. *Theor. Appl. Climatol.*, **137**, 1883–1893, https://doi.org/10.1007/s00704-018-2704-1.

Yu, S., B. Eder, R. Dennis, S. H. Chu, and S. E. Schwartz, 2006: New unbiased symmetric metrics for evaluation of air quality models. *Atmos. Sci. Lett.*, **7**, 26–34, https://doi.org/10.1002/asl.125.

Zhang, X., L. Chen, Z. Ma, and Y. Gao, 2021: Assessment of surface exchange coefficients in the Noah-MP land surface model for different land-cover types in China. *Int. J. Climatol.*, **41**, 2638–2659, https://doi.org/10.1002/joc.6981.

Zheng, W., H. Wei, Z. Wang, X. Zeng, J. Meng, M. Ek, K. Mitchell, and J. Derber, 2012: Improvement of daytime land surface skin temperature over arid regions in the NCEP GFS model and its impact on satellite data assimilation. *J. Geophys. Res.*, **117**, D06117, https://doi.org/10.1029/2011JD015901.

Zheng, Y., A. Kumar, and D. Niyogi, 2015: Impacts of land–atmosphere coupling on regional rainfall and convection. *Climate Dyn.*, **44**, 2383–2409, https://doi.org/10.1007/s00382-014-2442-8.

Zilitinkevich, S. S., 1995: Non-local turbulent transport: Pollution dispersion aspects of coherent structure of convective flows. *WIT Trans. Ecol. Environ.*, **9**, 53–60, https://doi.org/10.2495/AIR950071.