

RESEARCH ARTICLE

Sensitivity of Fine-Resolution Urban Heat Island Simulations to Soil Moisture Parameterization

Mahdad Talebpour¹  | Elie Bou-Zeid² | Claire Welty³ | Dan Li⁴ | Benjamin Zaitchik¹

¹Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, USA | ²Department of Civil and Environmental Engineering, Princeton University, Princeton, USA | ³Department of Chemical, Biochemical and Environmental Engineering, University of Maryland Baltimore County, Baltimore, USA | ⁴Department of Earth and Environment, Boston University, Boston, USA

Correspondence: Mahdad Talebpour (mtalebp1@jhu.edu)

Received: 26 December 2023 | **Revised:** 9 August 2024 | **Accepted:** 11 October 2024

Funding: NSF CBET-1444758; JHU Glenadore and Howard L. Pim Postdoctoral Fellowship in Global Change; NSF EAR-2012340; DOE DE-SC0023218.

Keywords: heat exposure | soil moisture | urban heat island | urban hydrometeorology

ABSTRACT

Urban areas experience the impact of natural disasters, such as heatwaves and flash floods, disparately in different neighbourhoods across a city. The demand for precise urban hydrometeorological and hydroclimatological modelling to examine this disparity, and the interacting challenges posed by climate change and urbanisation, has thus surged. The Weather Research and Forecasting (WRF) model has served such operational and research purposes for decades. Recent advancements in WRF, including enhanced numerical schemes and sophisticated urban atmospheric-hydrological parameterizations, have empowered the simulation of urban geophysical processes at high resolution (~1 km), but even this resolution misses significant urban microclimate variability. This study applies the large-eddy simulations (LES) mode within WRF, coupled with single-layer urban canopy models (SLUCM), to enable even finer-scale modelling (150 m) of the Urban Heat Island (UHI) effect in the Baltimore metropolitan area. We run nine scenarios to evaluate various methods of initializing soil moisture and various spinup lead times, and to assess the impact of WRF's Mosaic approach in depicting subgrid-scale processes. We evaluate the scenarios by comparing the WRF simulated land surface temperature (LST) against Landsat LST and the WRF simulated hourly 2-m air temperatures (AT) with observations from eight weather stations across the domain. Results underscore the paramount influence of the lead spinup time on the spatiotemporal distribution of simulated soil moisture, consequently shaping WRF's efficacy in predicting the UHI. Furthermore, interpolating soil moisture-related parameters from the parent for child domain initialization yields a notable reduction in mean and root-mean-squared errors. This improvement was particularly evident in simulations with the longest spinup time, affirming the importance of carefully designing the initialization of soil moisture for improved urban temperature predictions.

1 | Introduction

Exacerbated by climate change and rapid urbanisation, the increasing frequency and intensity of heatwaves are intensifying the urban heat island (UHI) effect, posing significant threats to public health, economy, and infrastructure in cities. These threats disproportionately affect underserved communities due to the inequitable distribution of extreme heat exposure and vulnerability (Hsu et al. 2021; Venter et al. 2023; Ming et al. 2024), necessitating novel

tools to understand the complex interplay between urban microclimates and socioeconomic factors at fine scales. These tools need to be able to capture the climatic means in urban areas, the meteorological perturbations around these means, as well as the potential benefits of solutions implemented to mitigate the impact of extreme heat events on vulnerable populations and neighbourhoods.

The UHI—a key component of the aforementioned threats—is a phenomenon whereby urban areas have higher temperatures

than surrounding suburban and rural areas. The combination of anthropogenic heat emission, city texture, and hygrothermal properties of urban surfaces modulate the intensity and trends of the UHI (Arnfield 2003; Shahmohamadi et al. 2011; Oke et al. 2017; Sobstyl et al. 2018). As an example, during summer months, the daily minimum temperature in New York City can be up to $\approx 4^{\circ}\text{C}$ warmer than the surrounding areas; this urban overheating can impact public health and air quality and increase energy and water demand (Rosenzweig, Solecki, and Slosberg 2006; Yang et al. 2016). Wong, Paddon, and Jimenez (2013), among many other studies, reviewed the UHI effect worldwide and found a direct link with increased mortality.

Urban areas are characterised by distinct features such as imperviousness (streets, parking lots, roofs, compacted soil), densely populated high-rise buildings, and sparse vegetative patches arranged in a highly heterogeneous spatial distribution. This heterogeneity leads to fine-scale variation in atmospheric and hydrological processes in urban areas. Binita, Shepherd, and Gaither (2015) demonstrated how some areas of cities with higher urban densities tend to have much higher temperatures, essentially intra-urban heat islands. The difference in thermal and hydrological properties, such as higher evaporation rates and albedo and lower heat storage capacity of vegetation relative to impervious surfaces, leads to this intra-urban heat island phenomenon. Therefore, the surface temperature can vary significantly across urban areas at fine scales, exacerbating the need for atmospheric simulations at finer resolutions to examine intra-urban overheating variability.

The Weather Research and Forecasting model (WRF; Skamarock et al. 2008) is a widely used tool for investigating the physics of the UHI effect and simulating mitigation strategies. WRF includes several Urban Canopy Model (UCM) options to represent urban-specific fluxes and hydrometeorological processes. Many researchers have used WRF with different UCMs to investigate the UHI phenomenon and its interactions with other physical and societal dynamics including climate change and heat waves at resolutions of ~ 1 km (Miao et al. 2009; Chen et al. 2011; Salamanca, Martilli, and Yagüe 2012; Li and Bou-Zeid 2014; Li, Bou-Zeid, and Oppenheimer 2014; Ramamurthy and Bou-Zeid 2014, 2017; Ramamurthy et al. 2014; Yang and Wang 2014; Kamal, Huang, and Myint 2015; Yang et al. 2015; Ryu et al. 2016; Sharma et al. 2016; Ramamurthy and Bou-Zeid 2017; Bassett et al. 2019; Li et al. 2019; Ortiz et al. 2019; Jandaghian and Berardi 2020; Liu and Morawska 2020; Wang and Hu 2021). With increased computational power and improved numerical schemes, WRF can now be readily applied at finer resolutions (Talbot, Bou-Zeid, and Smith 2012). However, the performance of WRF + UCM at higher resolutions and the value of finer-scale simulations (e.g., ~ 150 m) have not been explored except for a handful of studies (Shaffer et al. 2015; Hall et al. 2024).

In principle, a fine-scale local microclimatic and hydrological investigation can advance the understanding of intertwined interactions and processes at the interface of urban land surface, lower atmosphere, and upper subsurface. However, such simulations face two hurdles: (1) refining the resolution of the atmospheric component from 1 km to finer scales (e.g., ~ 150 m) would cause it to cross the terra-incognita—where the grid scale of the model is neither large enough to use Reynolds-averaged Navier–Stokes equations (RANS) turbulence representations adequate

for mesoscale models, nor small enough to use LES modelling for subfilter-scale turbulence (Wyngaard 2004); and (2) refining the surface model across the same range of scales results in a comparable change in physical regimes that gives rise to fine-scale variability of soil moisture content (SMC) that needs to be accounted for (Talebpour, Welty, and Bou-Zeid 2021).

For fine-scale resolutions (~ 150 m), employing the LES mode in WRF is necessary for the innermost domains, while WRF's default RANS with a planetary boundary layer scheme would still be used for the coarser parent domains at resolutions above 1 km. Moreover, the spatiotemporal distribution of SMC influences hydrometeorological processes in the upper subsurface, land surface, and lower atmosphere. The influence of SMC on hydro-meteorological and other geophysical processes is much stronger in urban areas that (1) are highly heterogeneous with various vegetated land covers such as parks, trees, and grass, and (2) in regions with shallow groundwater, where groundwater dynamics can significantly affect soil moisture variations. SMC also strongly influences the peri-urban reference temperatures that are used to calculate urban heat islands indices (Georgescu et al. 2011). Therefore, SMC initialization is critical for urban weather and climate modelling (Dy and Fung 2016; Santanello et al. 2019; Dennis and Berbery 2021; Talebpour, Welty, and Bou-Zeid 2021). For fine-scale simulation of urban hydrometeorological simulations, there is no consensus on methods for soil moisture initialization or on the spinup duration before the analysis period (i.e., the problem simulation period).

This study addresses the research gaps outlined above concerning high-resolution simulations using WRF with the single-layer urban canopy model (SLUCM; Kusaka et al. 2001), with three aims. First, we evaluate the quality and sensitivity of combining LES and SLUCM in WRF to simulate the UHI effect. Second, we examine the impact of SMC spinup on these simulations by testing three different starting points for spinning up the WRF model. Moreover, to rule out the bias introduced by SMC parameterization and initialization, several scenarios are designed to analyse the sensitivity of WRF-SLUCM to these parameters. Third, since the development and implementation of the Mosaic approach in WRF (Li et al. 2013), several studies have shown that WRF-Mosaic improves the representation of subgrid-scale heterogeneity of water and energy fluxes at resolutions of ~ 1 km by incorporating a lumped method to calculate heterogeneous fluxes (Ramamurthy and Bou-Zeid 2017; Sharma et al. 2017; Li et al. 2019; Yang and Bou-Zeid 2019). Although we use a fine-scale resolution in this study (150 m), there is still some sub-grid scale heterogeneity at much finer scales. Therefore, we also investigate whether employing the Mosaic approach can still improve subgrid-scale heterogeneous representation and UHI simulations at these finer resolutions.

2 | Methods

2.1 | Model Description

2.1.1 | WRF

For this study, WRF version 3.9.1.1 was used. WRF is an open-source, three-dimensional, atmospheric-surface model developed and maintained by the National Centre for Atmospheric

Research (NCAR). WRF is a massive multi-modular code incorporating numerous parameterizations and numerical schemes for hydrometeorological processes. The Noah land surface model (LSM; Chen and Dudhia 2001a, 2001b), used in this study, is one of the LSMs provided in WRF, for simulation of land surface and subsurface energy and water fluxes. We also use the SLUCM to represent the urban part of the land surface in Noah. In recent years, WRF has employed LES schemes to represent small scale turbulence in the atmosphere (<the model filter scale), while eddies larger than the grid scale are explicitly resolved. This mode thus also allows the model to resolve the surface heterogeneity at neighbourhood scales (~150 m). Initially, WRF in LES mode was implemented and tested for ideal-case scenarios. Talbot, Bou-Zeid, and Smith (2012) evaluated WRF's nesting capability to run fine-scale LES of real-case scenarios forced by downscaled initial and boundary conditions through nested domains. Since then, numerous studies have utilised WRF-LES to study real-case scenarios (Kirkil et al. 2012; Daniels et al. 2016; Ronda et al. 2017; Wiersema, Lundquist, and Chow 2020; Talebpour, Welty, and Bou-Zeid 2021; Zhang et al. 2022).

Up to WRF version 3.7, only horizontal nesting was possible. All domains (if run concurrently) had to have identical vertical layers, which is problematic as mesoscale and microscale domains require different vertical depths to resolve the physical processes. Using one combination of vertical layers for all domains introduces numerical errors due to the aspect ratio (vertical-horizontal) of grid spacing. Daniels et al. (2016) and Mirocha and Lundquist (2017) discussed the errors and biases introduced into nested simulations due to the cell aspect ratio, and they developed and evaluated a new vertical nesting capability to address this issue. This concurrent vertical nesting capability was employed in this study to reduce numerical errors associated with aspect ratio.

2.1.2 | Domain Extent and Analysis Period

The Baltimore metropolitan area was chosen as the domain of study for several reasons. First, the Baltimore Office of Sustainability has developed and updated Baltimore's Disaster Preparedness and Planning Project (DP3; 2013, 2018, and 2021; <https://www.baltimoresustainability.org/>) to address the impact of climate change on the future of Baltimore City. Six natural hazards identified as threats to the future of the city are related to hydrometeorological processes, including extreme heat. Moreover, previous studies using WRF mesoscale simulations and urban temperature measurement networks have demonstrated that the combined effects of heat waves and UHIs severely impact the Baltimore metropolitan area (Li and Bou-Zeid 2013; Ramamurthy and Bou-Zeid 2017; Scott et al. 2017).

At the neighbourhood scale, the variability of heat exposure across cities, a phenomenon called intra-urban heat islands, has been investigated and reported before (Scott et al. 2017; Hoffman, Shandas, and Pendleton 2020; Wilson 2020; Shi et al. 2021). These studies have demonstrated how scattered zones of underserved communities are prone to higher LST and heat exposure due to several factors, including lower vegetation-to-impervious

surface ratio and higher population density. As a result, these zones have a higher hydrometeorological and socioeconomic vulnerability to extreme events such as heat waves. Therefore, Baltimore has many risk factors (hazard, exposure, and vulnerability) that compel the present investigation of fine-scale variations of the UHI effect.

In the horizontal plane, WRF was applied to the Baltimore metropolitan area for 97.2 km by 97.2 km (Figure 1b). The horizontal resolution of the innermost domain (domain 5) was 150 m. This resolution was selected to be (1) below the ~150 m threshold recommended by the WRF community based on recommendations provided by (Wynngaard 2004) for scales where the use of LES turbulence closure is suitable, (2) small enough to capture urban microclimate variability at the neighbourhood scale, and (3) large enough to make running the numerous scenarios on sensitivity analysis proposed in this study computationally feasible.

The simulation period for the innermost domain in this study was from Aug 19, 2017, 11:00 EDT to Aug 22, 2017, 12:00 EDT. The final time of the simulation contains one clear snapshot of Landsat 8 (Dwyer et al. 2018) provisional surface temperature data at a 30-m resolution that we use for model evaluation purposes. Moreover, this date was one of the hottest days observed by Landsat during the summer of 2017.

2.1.3 | Model Input

NCEP North American Regional Reanalysis (NARR; Mesinger et al. 2006) 3-hourly forcing data, at 0.25° (~32 km) resolution, were obtained and processed with the WRF Preprocessing System (WPS). Multi-scale Ultra-high Resolution (MUR) Sea Surface Temperature (MEaSURES-MUR; Chin, Vazquez-Cuervo, and Armstrong 2017; JPL MUR MEaSURES Project, 5/16/2024) data were also processed and used in this study. WPS also provides all other preprocessed essential dynamic and static input data required for WRF simulations, including land use and digital elevation model (DEM). The resolution of preprocessed DEM and land use data available in WPS is appropriate for domains 1–3 (Figure 1a). However, for domains 4 and 5, USGS DEM (<https://www.usgs.gov/core-science-systems/ngp/3dep>) and USGS 2011 National Land Cover Dataset (NLCD; Homer et al. 2015) were obtained and resampled at a higher resolution (~30 m).

2.1.4 | Nested Configurations

Four nested domains encompassed the smallest domain of interest (domain 5 at 150 m resolution), having grid resolutions of 12,150, 4050, 1350, and 450 m (Figure 1a). In all scenarios, domains 1–3 started in mesoscale mode earlier than domains 4 and 5. This spinup period allows the outer three RANS domains to develop smaller-scale dynamics before LES domains 4 and 5 are initialized by the boundary conditions interpolated from the parent domains (3 and 4). Due to their higher computational expense, starting the inner domains as late as possible is preferred, closer to the analysis period, but early enough to still allow the inner domains develop turbulent flow structures at the scales of their grid resolution.

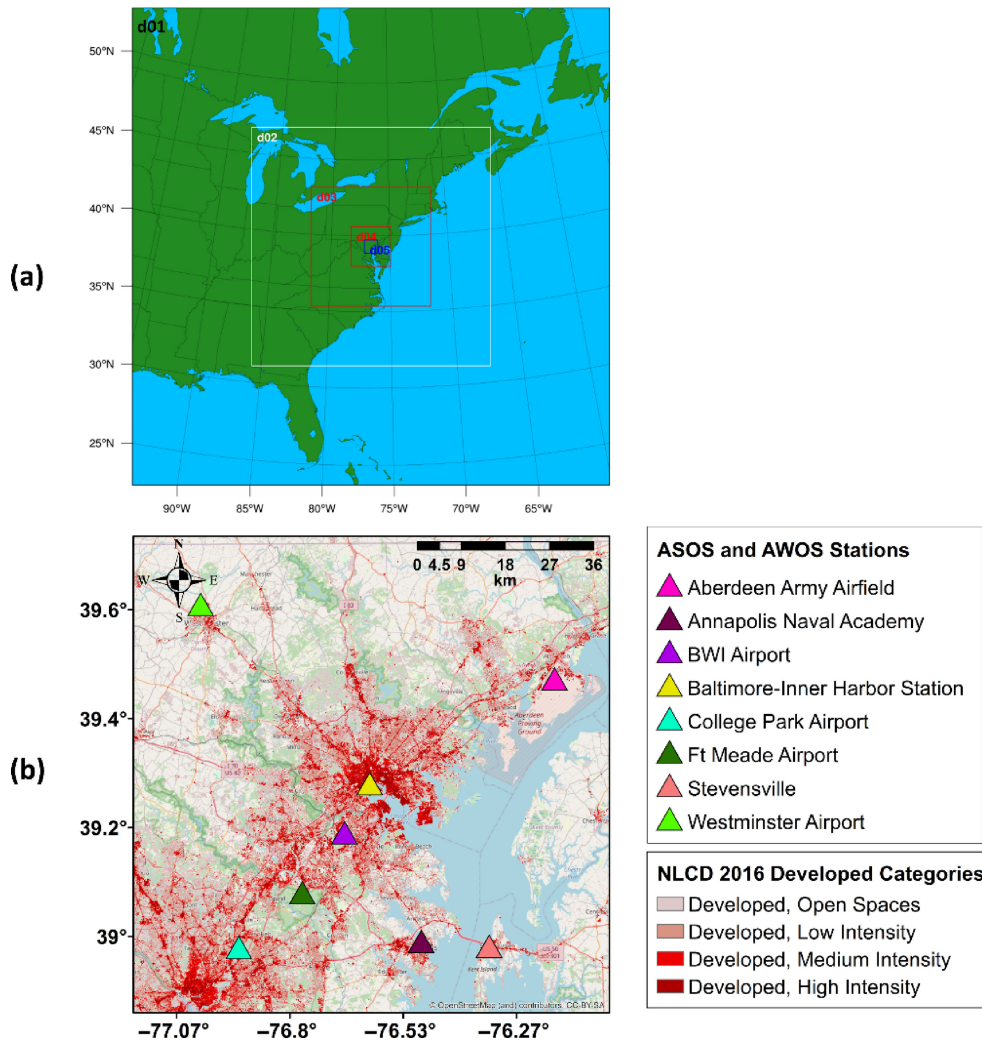


FIGURE 1 | (a) Nested four outer domains centring on domain 5 (innermost domain) of this study, (b) demonstrates the extent of the Baltimore domain simulated in this study. USGS World Topographic Map as base map overlaid by NLCD urban categories (<https://www.mrlc.gov>). The figure indicates eight AWOS and ASOS weather station sites for model air temperature evaluation. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

2.2 | Test Scenarios

We aim to address three research questions with our test scenarios.

2.2.1 | How Does the Length of the Spinup Period Influence the AT and LST Predictions?

To address the SMC initialization for the smaller domains, the outer three domains were run throughout rain events. Talebpour, Welty, and Bou-Zeid (2021) showed that intense rain events will ‘reset’ the SMC spatial distribution and remove the signature of low-resolution NARR SMC input. On the other hand, if the nested domains start much earlier before the analysis time, the atmospheric conditions tend to diverge from the observed input conditions since WRF here does not assimilate observational data like NARR (Ryu et al. 2016). Therefore, there is a trade-off between starting exactly before or much earlier than the analysis period. Table 1 describes scenarios used to evaluate differences in results between simulations starting at different times before the start of the analysis

window. In all three scenarios, domains 4 and 5 starting points are identical on Aug 19, 2017, 8:00 and Aug 19, 2017, 11:00. However, the outer three domains start together in each scenario on Aug 8, 2017, 11:00; Aug 14, 2017, 11:00; or Aug 17, 2017, 11:00.

2.2.2 | How Does the Mosaic Approach Influence AT or LST Predictions at Fine Scales?

Although 150 m is a fine-scale resolution compared to previous studies, urban heterogeneous land cover varies at a much finer scale. The 150-m by 150-m cells in this study are 25 times lower in resolution compared to NLCD land cover with 30-m resolution. We thus investigated whether incorporating the Mosaic approach, shown to improve the capability of WRF simulations at a 1-km resolution, can also improve the subgrid-scale parameterization and LST prediction where the scales between 150-m and 1-km are explicitly resolved. We ran three scenarios with three different starting times to investigate the Mosaic approach’s impact. These scenarios are called Mosaic group scenarios.

TABLE 1 | Simulation start time for domains 1–5 for each starting point scenario. All times are in EDT and military 24h format.

Initialization scenario	Domains 1–3 start point	Domain 4 start point	Domain 5 start point
START_08	Aug 08, 2017, 11:00	Aug 19, 2017, 8:00	Aug 19, 2017, 11:00
START_14	Aug 14, 2017, 11:00	Aug 19, 2017, 8:00	Aug 19, 2017, 11:00
START_17	Aug 17, 2017, 11:00	Aug 19, 2017, 8:00	Aug 19, 2017, 11:00

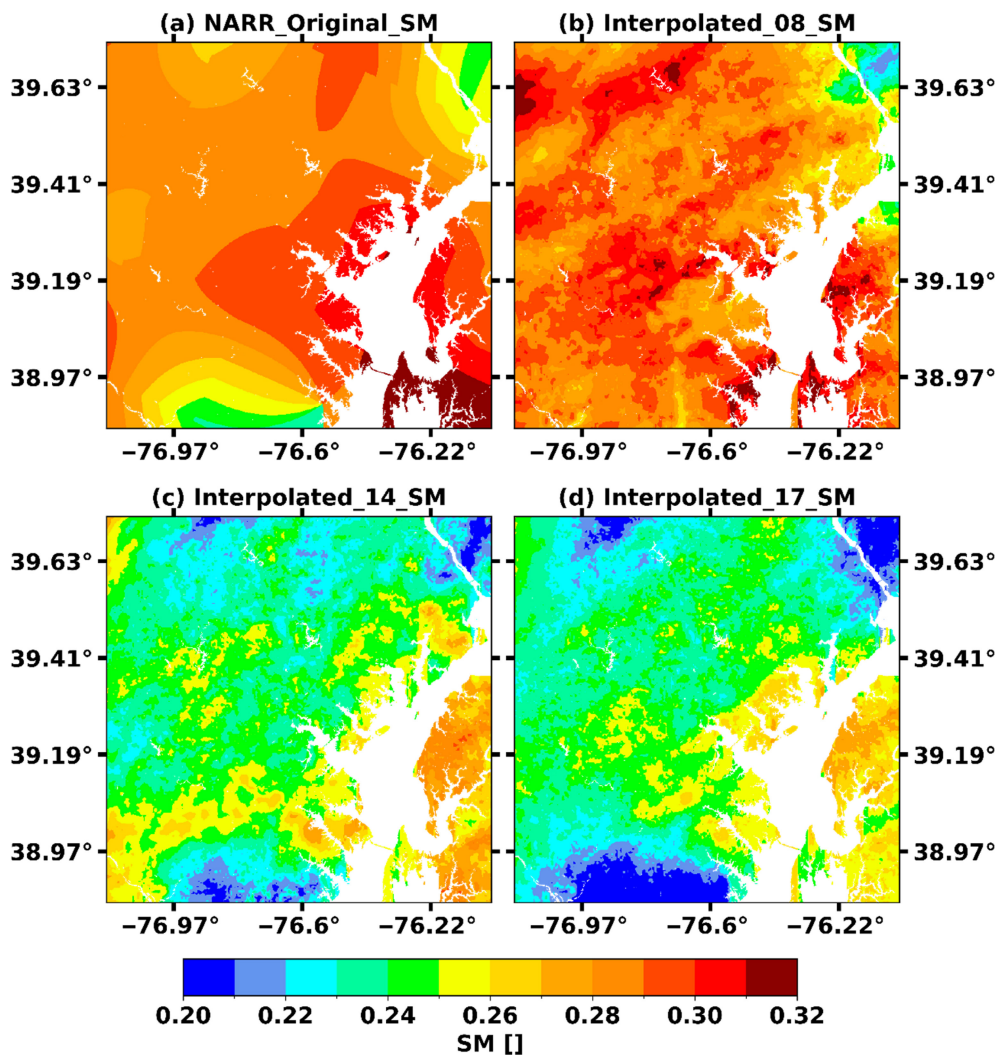


FIGURE 2 | SMC input for domain 5 (a) interpolation from NARR 32km resolution data; (b–d) interpolation from domain 4 for domain 5 in scenarios started on August 8, 14, and 17, 2017; with domain 4 itself initialized from domain 3 rather than NARR. [Colour figure can be viewed at wileyonlinelibrary.com]

2.2.3 | How Does Initializing the Innermost Domains (d4 and d5) With Interpolated Soil Moisture From the Parent Domain Affect LST and AT Evolution?

When nested inner domains start later than their parent domains (domains 4 and 5 in this study), WRF provides an option to initialize the inner domains by interpolating their parent hydrometeorological conditions. However, several input parameters, including soil moisture content, land surface temperature, and soil temperature, are not initialized from their parent domain output, even if the child domain starts later. In such instances, WPS will initialize the child domain with the same low-resolution observational or reanalysis data used to initialize the parent domain. Generally,

most simulations in this study start from the SMC distribution interpolated from the same NARR data used for domains 1–3. Therefore, the inner domains are initialized from low-resolution with almost homogeneous states of these parameters.

To examine whether using a high-resolution interpolated input has any influence on the prediction capability of the WRF model, we ran another group of scenarios with all three starting times for the parent domains 1 to 3. However, we bypassed the standard WRF initialization protocol in this group scenario. We interpolated SMC, soil water content, soil temperature, and relative SMC from the parent domains output 3 and 4 for initializing domains 4 and 5, respectively, at their start time; Figure 2a–d show four

different SMC inputs for domain 5 for different group scenarios. Figure 2b–d are the SMC distributions interpolated from domain 4 output in the scenarios started from August 8, 14, and 17, while Figure 2a shows the SMC distribution interpolated from NARR data. The potential benefits of interpolating from the parent domain are apparent, with domain 5 exhibiting much more variability at initial times when interpolating from domain 4 (compare Figure 2b–d to a). The influence of the parent domain spinup is also evident. The Interpolated_08 scenario has higher SMC values than Interpolated_14 and Interpolated_17 due to intense rain events captured in WRF’s RANS domains (1–3) between August 8–14, 2017. The NARR-interpolated SMC also has much higher values closer to the Start_08 distribution but with much less variability in spatial distribution.

2.2.4 | Summary of Group Scenarios

Table 2 summarises the case scenarios simulated in this study. There are nine scenarios with three different starting dates for domains 1–3. Three base scenarios run with different starting times all used NARR data in domains 4 and 5; this group is called the Original scenarios. Three scenarios with different starting dates all used the Mosaic approach (hence called the Mosaic group) to refine surface representation, as well as the NARR-interpolated data as in the Original group. Therefore, domain 5 SMC input for all six scenarios in the Original and Mosaic group is the same as in Figure 2a. The last group of scenarios, again with different starting times, incorporated the interpolated initial conditions from the parent domains but did not use the Mosaic approach. This last group is called the Interpolated group.

2.3 | Modified Zilitinkevich and LST Postprocessing

Li and Bou-Zeid (2014) demonstrated how the modified Zilitinkevich relationship developed and implemented into WRF (Chen and Zhang 2009) significantly reduced LST biases when evaluated using 1-km resolution satellite data (MODIS LST), in non-urban parts of their Baltimore-Washington domain. However, they showed that owing to incorrect calculations over

urban terrain, where WRF assigns thermal roughness parameters and uses aggregate heat fluxes to infer surface temperature, WRF still generates higher biases in LST. This bias is introduced into the calculation of the impervious surface temperature. To address this issue, they proposed an alternative equation that calculates LST for urban impervious surfaces:

$$T_{s(\text{impervious})} = f_{\text{roof}} T_r + (1 - f_{\text{roof}}) T_c \quad (1)$$

where $T_{s(\text{impervious})}$ is the surface temperature for impervious surfaces, and T_r and T_c are roof and canopy temperatures for each grid cell. T_r and T_c are computed directly by the SLUCM components in WRF over each cell. Following their postprocessing step, this study calculated all urban LST results from Equation (1). Li and Bou-Zeid (2014) provided detailed explanations of the biases in the Original $T_{s(\text{impervious})}$ calculated by WRF at runtime, evaluation of the alternative method, and justifications for using the alternative equation.

2.4 | Computational Resources

Most simulations for this study were carried out on the Cheyenne high-performance computing system provided by NCAR’s Computational and Information Systems Laboratory (CISL 2019). Some simulations conducted later to evaluate the impact of sea surface temperature were run on Derecho (CISL 2023), the successor of Cheyenne at CISL. Simulations used 9 and 36 computational nodes (36 cores per node) with 324 and 1296 cores at different simulation stages. Each scenario took ~96–120h of wall clock time from the start to the end. It is worth noting that the computing cost of Domain 5 is much higher than that of Domain 1 or other coarser domains. To simulate 24 physical hours, the computing cost in floating point operations scales with N^3/dt , where dt is the time step and N^3 is the total number of grid points in the domain. If the resolution of domain 1 is 12,150 m and that of domain 5 is 150 m, and assuming they use a comparable number of grid nodes N^3 , numerical stability (the Courant–Friedrichs–Lewy condition; Courant, Friedrichs, and Lewy 1967) requires domain 5 to have a dt that is about 150/12150 (0.012) smaller than that of domain 1. This consideration implies that the computing cost of domain 5, for example

TABLE 2 | Simulation case scenario description.

Scenario	Domains 4 and 5 input soil moisture	Domains 1–3 start time
Original_08	NARR Original	8/8/2017
Original_14	NARR Original	8/14/2017
Original_17	NARR Original	8/17/2017
Mosaic_08	NARR Original	8/8/2017
Mosaic_14	NARR Original	8/14/2017
Mosaic_17	NARR Original	8/17/2017
Interpolated_08	Interpolated from parent	8/8/2017
Interpolated_14	Interpolated from parent	8/14/2017
Interpolated_17	Interpolated from parent	8/17/2017

for simulating one day, will be $12,150/150=81$ times the computing cost of domain 1.

We conducted all the postprocessing and visualisation of output data on NCAR's JupyterHub on the Casper system provided by CISL (2019). Most pre- and post-processing was conducted by employing several software packages including Python, Scipy, Jupyter Notebook, wrf-python, xarray, NetCDF, and matplotlib (Rew and Davis 1990; Brown et al. 1993; Van Rossum and Drake 2009; Kluyver et al. 2016; Hoyer and Hamman 2017; Ladwig, William 2017; Virtanen et al. 2020; Caswell et al. 2023).

3 | Results and Discussion

3.1 | Soil Moisture

Figure 3a–d show the domain-averaged SMC values time series for all nine scenarios and four soil layers simulated in the Noah LSM component of WRF. Excluding the cells with water bodies (e.g., the Chesapeake Bay) was necessary as they were masked as permanent water bodies by WRF. WRF assigns values of 1.0 to the SMC output for these cells for the entire simulation period, and they skew domain-averaged values.

The average SMC value for all scenarios in the 4th (bottom) layer, with a thickness of 1 m, exhibits negligible changes for the entire

simulation and a SMC value of approximately 0.27. There are radical changes in area-averaged SMC values from the first to the second time step (not shown; Figure 3 only shows every 2-h intervals for clarity, skipping the second output interval), most recognisable in layers 1 and 3 and for the Original and Mosaic groups. These SMC values get overwritten in the WRF initialization module in the first timestep after the SMC is recalculated based on other parameters fed into the model. In Interpolated scenarios, SMC, soil temperature, and relative SMC values are interpolated from mesoscale parent domain values (domains 3–4 and 4–5), displaying a smoother transition from the first output interval to the next.

Precipitation input for all scenarios was negligible from Aug 19 to 22, 2017. For all scenarios, reduction in SMC due to evapotranspiration is noticeable during daytime, mainly in the first and second layers. We started all the Mosaic and Original group scenarios from identical high NARR-interpolated SMC values (Figure 2a). As seen in all layers, SMC remains almost identical for the Mosaic and Original group scenarios, with a slight divergence toward the end of the simulation. Nevertheless, as pointed out, there is a sudden drop (~ -0.02) and increase (~ 0.03) at the start of the simulation in average SMC values in layers 1 and 3 in these scenarios. Whereas the Mosaic and Original group scenarios started from higher SMC content on average in the top layer (~ 0.28), they decreased to ~ 0.22 toward the end ($\sim 21\%$ loss).

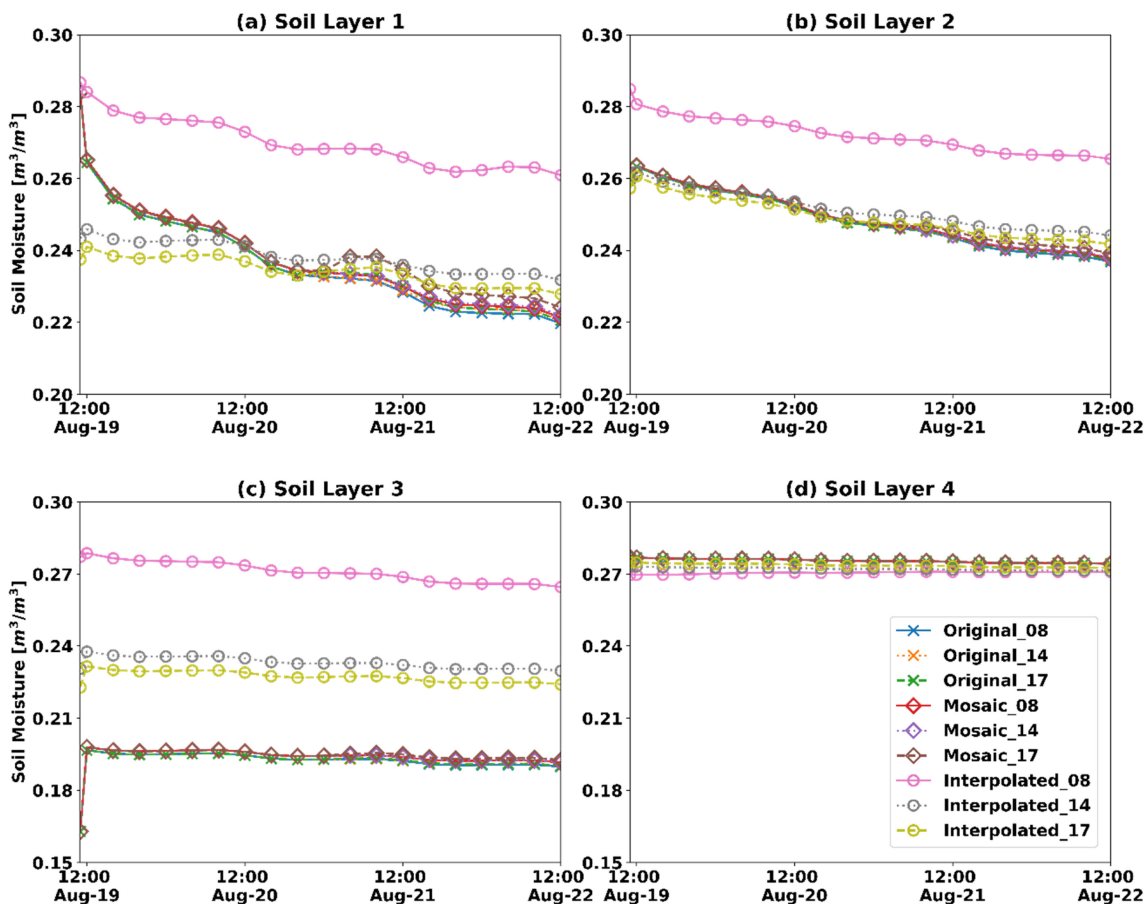


FIGURE 3 | Domain-average SMC values for the entire domain, excluding water body cells, for all nine scenarios, from Aug 19, 2017, 11 AM EST to Aug 22, 2017, 12 PM EST. Panels a-d show SMC for the four WRF soil layers. Markers are printed at 2-h intervals. [Colour figure can be viewed at wileyonlinelibrary.com]

The SMC of the Interpolated group scenarios had a different evolution. The Interpolated_08 scenario had the highest average SMC compared to all the scenarios in the top three layers, explaining the slower drop in the moisture of the top layer over time compared to the Original and Mosaic groups. Figure 2b shows the spatial distribution of the first SMC layer, interpolated from domains three and four, after nine days of simulation in WRF. Therefore, higher rain infiltrated the soils during Aug 8–14 in this scenario. The Interpolated_14 and Interpolated_17 had much lower average SMC values in layers 1–3 (compared to Interpolated_8, Figure 3a–c). Figure 3 shows domain-averaged SMC for the four WRF soil layers. While their values differ, the trends of change in average SMC values for the Interpolated scenarios are similar, indicating that the evapotranspiration might not differ.

In contrast with the Mosaic and Original group scenarios, the Interpolated_08 started from a similar area-averaged SMC value (~ 0.285), but it did not drop dramatically in the first interval. The cause might be that other than SMC content, relative SMC, soil liquid water, and soil temperature were also interpolated from domains 3–4 and then 4–5. Therefore, the water-energy balance recalculated at the first timestep was closer to an equilibrium state for the Interpolated group scenarios than the Mosaic and Original group scenarios.

3.2 | Land Surface Temperature Evaluation

We compared the LST spatial distribution results with Landsat LST observations on Aug 22, 2017, 12 PM EDT. Figure 4 shows the pseudocolour map of land surface temperatures for all scenarios and Landsat data on Aug 22, 2017, 12 PM EDT. All WRF scenarios model the entire Chesapeake Bay surface temperature with a spatially constant temperature. This constant bay temperature is due to the parameterization of water body temperature evolution in WRF. However, the Landsat (Figure 4j) has a temperature gradient over the Bay, with colder temperatures in the centre of the Bay in deeper water bodies, changing to warmer temperatures along the coastline. We also run two other simulations on domain 3, one with the Original NARR SST and another one with MUR SST as input. As shown in Appendix A, using higher resolution MUR SST, resulted in negligible differences with the simulation with the Original NARR SST. Therefore, the Bay cells were excluded from data analysis from all simulation scenarios and Landsat for statistical comparison.

3.2.1 | Bay-Excluded LST Statistics

Figure 5 depicts the spatial cell-wise difference between simulated scenarios LST and Landsat with the Bay masked. A quick view of the spatial difference indicates a sharp discrepancy between WRF performance over urban versus non-urban land covers. This can be explained by the fact that LST was calculated using fluxes computed by SLUCM and NOAH. Overall, in all scenarios WRF underpredicted LST over urban land covers and overpredicted over non-urban land covers. Start_08 scenarios with longer SMC spinup decreased the urban underprediction gap but exacerbated the overprediction over non-urban areas. On the other hand, Start_17 scenarios had lower bias over

non-urban areas compared to Landsat, but produced cooler temperature and larger biases over urban land tiles. Among all scenarios, Interpolated_08 had a lower bias in both non-urban and urban land covers, which might be due to higher soil moisture availability (shown in Figures 2 and 3) and longer spinup time. Nevertheless, spatial pseudocolour maps, while useful for showing spatial patterns, cannot explain all the trends without statistical analysis. Therefore, we discuss some statistical evaluation of scenarios compared to Landsat next.

Figure 6 shows the probability density functions (PDFs) of the cell-wise difference between simulated scenarios LST and Landsat with the Bay masked. A better model performance would result in narrower distributions of these different PDFs that are centred (mean or mode) around 0°C . However, moving from right to left, earlier starting times result in a larger departure of the PDF mode from zero, but the standard deviation (STD) becomes smaller (narrower PDF) with Interpolated_08 scenario having the narrowest distribution. A narrower PDF with shifted mode from zero means that starting earlier lowers the cell-wise errors for the entire domain but tends to shift to a warmer temperature bias on average for the entire domain. More specifically, the Start_08 scenarios (Figure 6, panels a, d, g) for all three groups (Original, Mosaic, and Interpolated) are characterised by PDFs of difference from Landsat ranging from -8 to 10°C with peaks at $\approx 1^{\circ}\text{C}$. The scenarios started on Aug 17 (Figure 6, panels c, f, i) exhibit PDFs ranging from -10 to 10°C , with peaks at $\approx 0^{\circ}\text{C}$.

Table 3 summarises the LST statistics for all scenario groups compared to Landsat over the entire domain, excluding the Bay. Figure 7 demonstrate a statistical comparison between scenarios and Landsat over the entire domain, again excluding the Bay. Figure 7a shows the Taylor diagram (Taylor 2001) for evaluating scenario performances by comparing correlation and STD values between scenarios and Landsat and showing their unbiased RMSE (CRMSE) on a single diagram. The Landsat values, excluding the Bay, had an STD value of 3.55°C marked by a star on the bottom STD axis. The smaller the distance between a scenario marker and this Landsat point, the closer its overall predictions (in terms of mean, variance, and spatial patterns) are to the Landsat values.

In Figure 7a, all scenarios spread along the STD line of $\approx 2.5^{\circ}\text{C}$, which is $\approx 1^{\circ}\text{C}$ smaller than Landsat STD of 3.55°C , indicating lower simulated spatial variability compared to Landsat. However, this variability alone cannot fully describe model skill since scenarios might have similar STDs but still contain significant local errors. The spatial correlation values can help to evaluate scenarios' performances in capturing the spatial distribution or patterns of predicted values compared to Landsat. Correlation values range from about 0.4 for Start_17 scenarios to about 0.67 for the Interpolated_08 scenario. The Start_17 scenarios cluster around a 0.45 correlation line.

Moving from the star marker depicting Landsat toward the scenarios, the values of CRMSE, showing the unbiased error from Landsat, increase. It is obvious that start time had the most substantial influence on model performance, with Start_17 scenarios having lower correlation and higher CRMSE values. While the rest of the scenarios are all clustered, Interpolated_08 with a red diamond performs slightly better by demonstrating the highest correlation of 0.67 and lowest CRMSE (2.65°C).

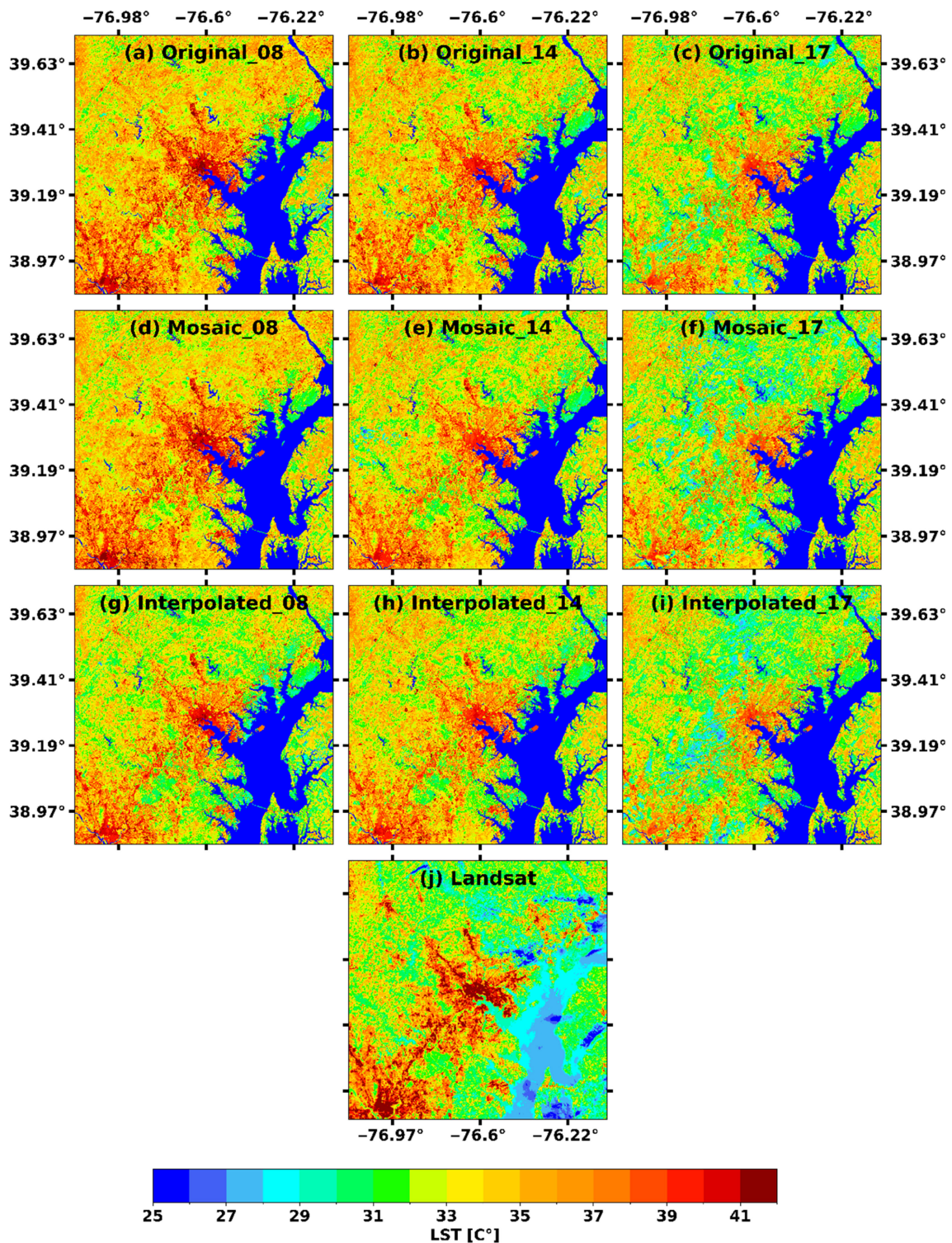


FIGURE 4 | Pseudocolour plots of LST for all case scenarios and Landsat for the entire domain on Aug 22, 2017, 12 PM ET. [Colour figure can be viewed at wileyonlinelibrary.com]

Although the Taylor diagram is a powerful tool to compare different model performances to observation, it lacks information on the bias of the scenarios. Figure 7b shows the Bias-Error diagram demonstrating the Mean Errors (ME; the bias) on the horizontal and CRMSE on the vertical axis. The shaded, orange-coloured circular wedges show the range of RMSE from 0°C to 3.5°C in increments of 0.5°C. Similar to the Taylor diagram, the smaller the distance from the star (depicting the Landsat

point at the centre along the ME axis), the better the model's performance.

The Bias-Error diagram further demonstrates differences between the scenarios. Once again, Interpolated_08 is the closest to Landsat, indicating the lowest combined bias and error. The Start_17 group scenarios performed worse than the Start_14 and Start_08. The starting point seems to have the most substantial

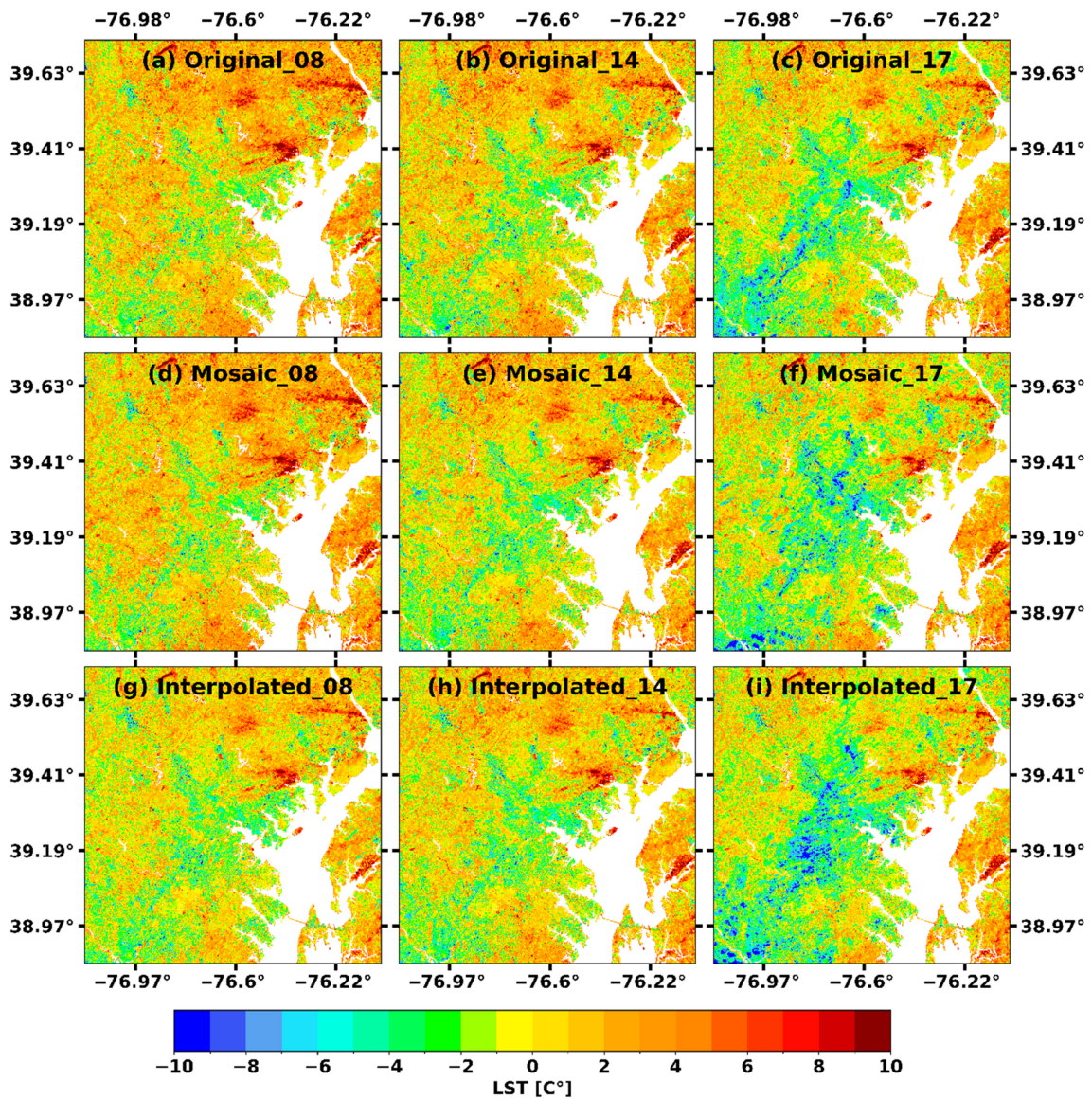


FIGURE 5 | Pseudocolour plots of difference between WRF scenarios LST minus Landsat LST with the Bay masked on Aug 22, 2017, 12 PM ET. [Colour figure can be viewed at wileyonlinelibrary.com]

influence on scenarios. Most scenarios with similar starting points are clustered, with Start_17 showing the lowest bias (ME) but the highest CRMSE and RMSE. Similarly for Start_14 scenarios, clustering at lower CRMSE and RMSE demonstrates higher ME. The only exception here is Interpolated_08, which has the lowest CRMSE and RMSE, and its ME is the third lowest (0.37°C), better than the other Start_08 scenarios (Original_08 and Mosaic_08) with respective ME values of 1.39°C and 1.42°C.

The combination of the Taylor diagram and Bias-Error diagram shown in Figure 7 emphasises that the starting point of spinup most strongly influences the model's performance in predicting LST compared to Landsat over the entire domain, excluding the Bay. The longer the spinup period, the more significant the spatial correlation between scenarios and the Landsat observation and lower local errors from observation (CRMSE). However, as shown in both Figures 6 and 7b, the longer the spinup, the larger the ME and the overprediction of LST over the entire domain. The Interpolated_08 scenario is an exception. While Interpolated_08 has the lowest cell-wise

local error (CRMSE) due to its longer spinup, it has the third lowest ME (0.37°C). The Interpolated_08 scenario benefited from a more extended spinup period and improved spatial distribution of soil moisture provided by interpolation from the parent domain.

3.2.2 | Urban Land Cover LST Statistics

We compared the performances of different scenarios for urban and non-urban areas to determine the influence of different methods used in each scenario on LST prediction over specific land uses. This analysis examines how urban versus rural energy balance models perform since the SLUCM handles these balances and fluxes in WRF over urban terrain while Noah simulates vegetated surfaces. Moreover, it aids in reaching one of the objectives of this study, which was to evaluate the performance of SLUCM and LES mode together in representing fine-scale UHI. This section presents the urban LST results for each scenario compared to Landsat.

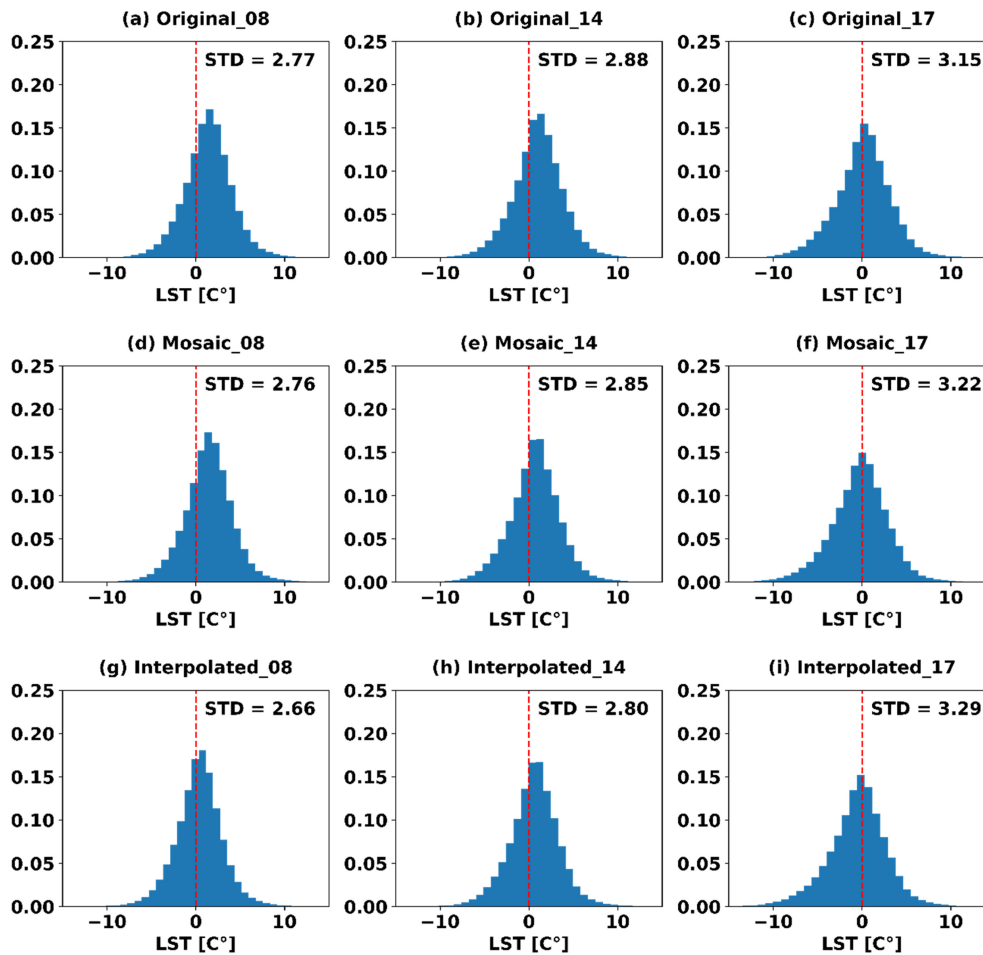


FIGURE 6 | Probability density function (PDF) of cell-wise difference between WRF scenarios LST minus Landsat LST, with the Bay masked. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 | LST statistics for all scenarios and Landsat data, excluding Bay cells. All values are in degrees Celsius except for the Pearson column showing the correlation values.

Scenarios	Mean	STD	ME	RMSE	CRMSE	Correlation
Original_08	34.84	2.6	1.39	3.1	2.77	0.63
Original_14	34.46	2.5	1.02	3.05	2.88	0.60
Original_17	33.51	2.39	0.06	3.15	3.15	0.50
Mosaic_08	34.87	2.50	1.42	3.11	2.76	0.63
Mosaic_14	34.12	2.46	0.67	2.93	2.85	0.60
Mosaic_17	33.16	2.55	-0.28	3.23	3.22	0.48
Interpolated_08	33.81	2.57	0.37	2.68	2.66	0.67
Interpolated_14	34.04	2.47	0.59	2.86	2.80	0.62
Interpolated_17	32.88	2.47	-0.57	3.34	3.29	0.45
Landsat	33.45	3.55	0	0	0	1

Table 4 summarises the LST statistics for all group scenarios compared to Landsat over urban cells. We again used the Taylor and Bias-Error diagrams to compare the scenarios against Landsat (Figure 8). A quick overview of the Taylor diagram in Figure 8a shows a similar pattern to Figure 7a for the Bay-excluded results

for the scenarios. Overall, the most substantial influence again comes from the spinup duration, where Start_17 scenarios clustered at correlation values ranging from 0.42 to 0.49 and CRMSE values ranging from 3.25°C to 3.5°C. On the other hand, the rest of the scenarios clustered closer to the Landsat point (star marker).

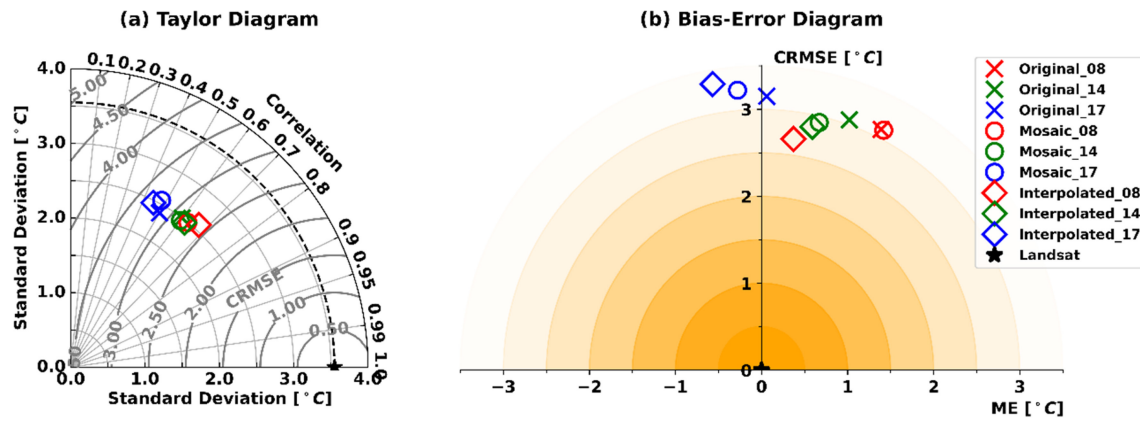


FIGURE 7 | Statistical comparison between scenarios and Landsat over the entire domain excluding the Bay on two diagrams: (a) Taylor diagram comparing the correlation between scenarios and Landsat, the standard deviation of scenarios and the Landsat (dashed line), and unbiased RMSE (CRMSE); and (b) Bias-Error diagram acting as a complimentary to the Taylor diagram to show the error and bias from Landsat for all scenarios. The shaded orange wedges are 0.5 increments of RMSE values ($RMSE^2 = CRMSE^2 + ME^2$). [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 | LST statistics for all scenarios and Landsat data only over urban cells. All values are in degrees Celsius except for the Pearson column showing the correlation values.

Scenarios	Mean	STD	ME	RMSE	CRMSE	Correlation
Original_08	36.83	2.60	0.22	2.90	2.89	0.61
Original_14	36.19	2.52	-0.41	2.99	2.96	0.59
Original_17	34.90	2.53	-1.70	3.67	3.25	0.49
Mosaic_08	36.85	2.53	0.24	2.93	2.92	0.60
Mosaic_14	35.92	2.47	-0.69	3.00	2.92	0.60
Mosaic_17	34.63	2.66	-1.97	3.88	3.34	0.48
Interpolated_08	35.95	2.65	-0.66	2.91	2.83	0.62
Interpolated_14	35.84	2.54	-0.77	3.00	2.90	0.60
Interpolated_17	34.22	2.66	-2.38	4.23	3.50	0.42
Landsat	36.61	3.64	0	0	0	1

Although all scenarios had similar Taylor diagram for urban versus all land covers, the performance of the scenarios is quite different in the Bias-Error diagram (Figure 8b). All scenarios except for Original_08 and Mosaic_08 show cooler bias compared to Landsat over urban areas on average. This indicates that the low bias in Figure 7 for all land uses resulted from an error cancellation with a warm bias over natural terrain. The Start_17 scenarios are much farther from Landsat than the Bay-excluded comparison (Figure 7b). The ME values for the Start_17 scenarios range from -2.38 to -1.7°C . All other scenarios have similar distances to the Landsat point, indicating similar RMSE values. Original_08 has the lowest RMSE and ME values of 2.90°C and 0.22°C . Based on the overall statistics, it is evident that the SLUCM results are also sensitive to spinup duration.

The sensitivity of the SLUCM to starting points can be detected from the clustering of different colours. Interpolated_08 again is an exception to this rule because the interpolation from domain 3 to domain 4 and domain 4 to domain 5 dramatically changed SMC at their initial start time. Therefore, the Interpolated_08

scenario acted differently from the Original_08 and Mosaic_08 and predicted cooler temperatures. As explained above, all interpolated scenarios' SMC were calculated from their parent domains. Scenarios started from August 08, featured significant and long precipitation events between August 10 and August 14, resulting in significant increase in SMC for these scenarios. Therefore, Interpolated_08 started with much higher SMC and latent heating, and cooler temperatures, at the end compared to any other simulation. This observation is consistent with Talebpour, Welty, and Bou-Zeid's (2021) findings that Noah LSM allows high infiltration rates during precipitation events, leading to higher latent heating and cooler temperatures over impervious surfaces.

The statistics shown in Figure 8 and Table 4 are evaluated for all urban land cover categories (i.e., the four urban categories in NLCD land cover). Figure 9 shows the MEs per each NLCD urban category and illustrates how each scenario performed in the prediction of LST in each of these urban categories. The values printed in blue at the centre of each figure are the MAEs for the entire urban land cover for each scenario. In most

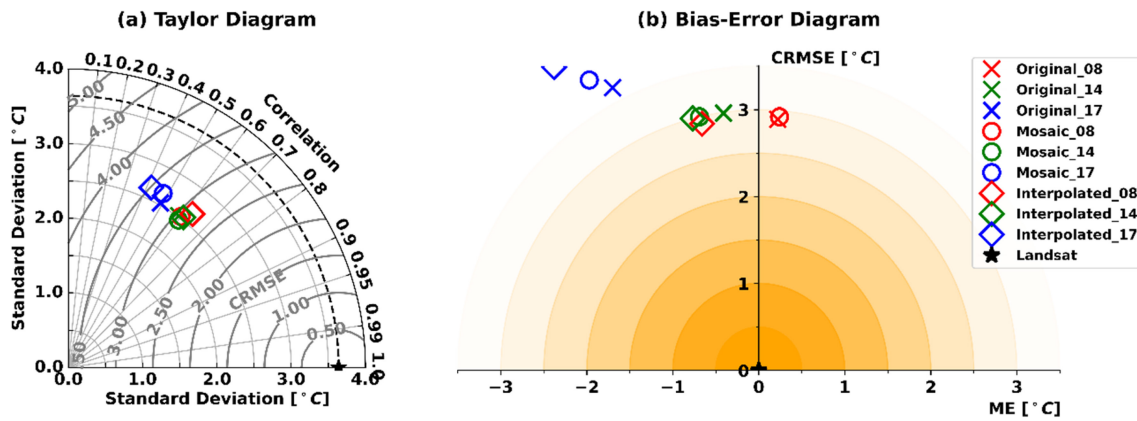


FIGURE 8 | Statistical comparison between scenarios and Landsat only over urban land cover on two diagrams: (a) Taylor diagram comparing the correlation between scenarios and Landsat, the standard deviation of scenarios and the Landsat (dashed line), and unbiased RMSE (CRMSE); and (b) Bias-Error diagram acting as a complimentary to the Taylor diagram shows the error and bias from Landsat for all scenarios. The shaded orange wedges are 0.5 increments of RMSE values ($RMSE^2 = CRMSE^2 + ME^2$). [Colour figure can be viewed at wileyonlinelibrary.com]

scenarios, the MAE increases from the Developed-Open Space category with $\approx 20\%$ imperviousness toward Developed-High Intensity with $\approx 95\%$ imperviousness. However, the difference between Developed-Open Space and Developed-Low Intensity is negligible. Figure 4 illustrates how most scenarios could not capture the hottest spots in the Landsat map, which overlaps with Developed-High Intensity land cover category.

3.2.3 | Non-Urban Land Cover LST Statistics

Table 5 summarises the LST statistics for all group scenarios compared to Landsat over non-urban cells. Figure 10 shows the corresponding Taylor and Bias-Error diagrams for non-urban land covers. While for the urban terrain Taylor diagram, the patterns were similar to the Bay-excluded patterns, the non-urban diagram is different. The spatial correlation values, between 0.17 and 0.28, are much lower than urban values. Regarding the scenarios' performances implied from the Taylor diagram, the scenarios had similar performance, with starting points strongly influencing the clusters. More specifically, Start_17 scenarios show higher CRMSE values and are farther from Landsat than the other scenarios. Interpolated_08 again demonstrates the lowest RMSE and shows significantly more predictive skill than other Start_08 scenarios.

In the Bias-Error diagram, the clusters are similar to those found in the Bay-excluded diagram; however, they spread out more along the ME axis. The Start_17 group scenarios and the Interpolated_08 scenario have the lowest ME values. Considering all the previous observations, it is evident that the Interpolated_08 scenario benefits from both longer spinup and interpolation of SMC, resulting in the lowest biases and errors and the highest correlation values. While Start_17 scenarios produce lower ME, their spatial correlation with Landsat is much lower than other scenarios, and they produce higher local cell-wise errors (RMSE). Finally, the lower correlation and higher ME values observed in the non-urban versus urban areas indicate that parameterizations in the SLUCM predict surface energy balance and LST better than the parameterizations in Noah over non-urban land cover.

3.3 | Air Temperature Evaluation

We compared a time series of AT values from all scenarios with eight ASOS and AWOS stations across the domain (see map in Figure 1b). The comparisons included time series plots of AT for Original and Interpolated scenarios over three stations: Aberdeen, BWI, and College Park. Mosaic scenarios showed no improvement over the Original scenarios, so they are excluded here. The plots are shown in Figure 11. These three stations were chosen because they span three domain regions (north-east, centre, and southwest) and have different surrounding land cover. During nighttime (8 PM to 6 AM; shaded area in the plots), the difference between all WRF scenarios and the observations reaches the highest values for each station.

Overall, the daytime model performance was better at all stations. This might be due to the difficulty in parameterizing the nighttime boundary layer and the stable atmospheric regimes, as also observed and commented on by Talbot, Bou-Zeid, and Smith (2012). The Original_14 and Interpolated_14 scenarios have the closest match to NOAA observations, as shown in Figure 11. The Original_17 and Interpolated_17, on the other hand, have the most considerable biases at all three stations. The difference between Start_14 and Start_08 scenarios during daytime is almost negligible.

All the statistical values (e.g., RMSE and STD) were calculated for 72 hourly points over the entire simulation. Figure 12 shows the Taylor diagrams for these comparisons for each of the 8 station locations in the domain identified in Figure 1a. The Taylor diagrams reveal that at all stations, the correlation values for all scenarios are above 0.8 and, in some scenarios, reach 0.95. At BWI and Westminster locations, the scenarios are all clustered together.

As noted before, a good complement for the Taylor diagram is the Error-Bias diagram that we plot in Figure 13. The Bias-Error diagrams reveal that at the Annapolis, Inner Harbour, Stevensville, and Westminster stations, all scenarios have near zero or negative MEs, indicating underprediction of air temperature at these

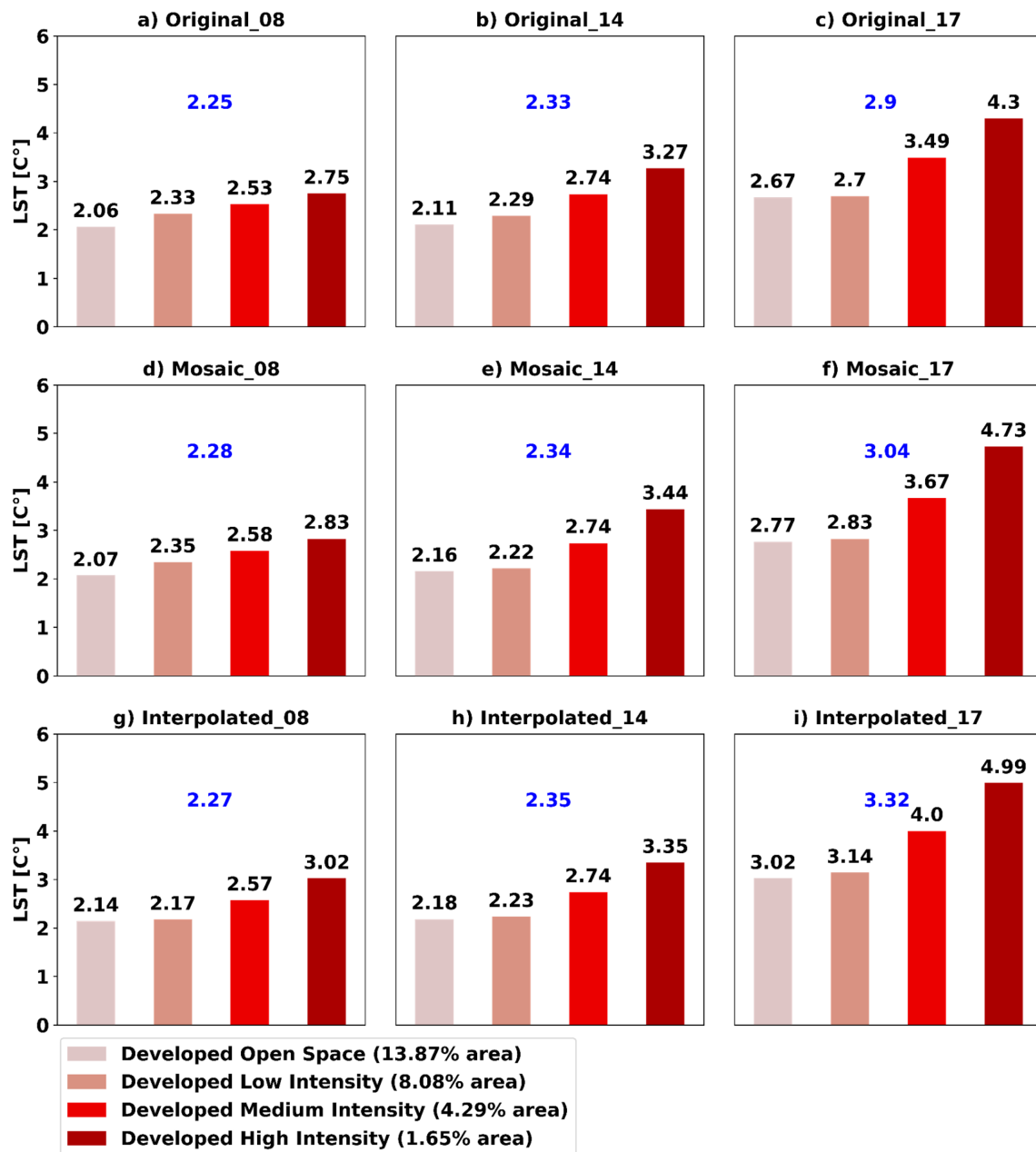


FIGURE 9 | MEs from Landsat LST for each NLCD urban category for all scenarios. The blue number in each figure represents the ME for all urban land cover for each scenario, as shown in Table 4. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

stations by WRF compared to the observations (Figure 13b,f-h). Three stations are on the Chesapeake Bay coast (Annapolis, Inner Harbour, and Stevensville). As shown in Figure 4, all model scenarios treat water body temperatures as constant. In this case, the modelled scenarios were cooler by more than 2°C–3°C difference, compared to Landsat, over most Bay cells. These cooler water temperatures in the model can result in strong sea breezes, leading to cooler temperatures or biases in the scenarios near the coastlines. However, for the Westminster station, the comparison is different. The Westminster station is far away from any water body. Therefore, the Bay breeze cannot cause the cooler bias. The most identifiable characteristic of this station is that croplands and pasture/hay land covers surround it, unlike other stations. Delving into why all scenarios underpredicted AT at the Westminster station is beyond the scope of this paper as it needs an examination of the parameterizations of the

croplands and pasture/hay land covers in WRF and how any irrigation may affect the results.

All scenarios overpredicted AT at the other four stations: Aberdeen, BWI, College Park and Fort Meade (Figure 13a,c-e). All these stations are located farther from water bodies. BWI has lower RMSE and ME values compared to the other stations. This could be because BWI is closer to the Chesapeake Bay and is surrounded by large forested areas.

The spinup period again has the most substantial impact on scenario prediction capabilities. Start_14 scenarios predicted lower AT at all stations than Start_08 and Start_17 scenarios. The Interpolated_08 scenario again has closer values to the Start_14 scenarios as it incorporates longer spinup for atmospheric processes and soil moisture than other Start_08

TABLE 5 | LST statistics for all scenarios and Landsat data over non-urban land covers. All values are in degrees Celsius except for the Pearson column showing the correlation values.

Scenarios	Mean	STD	ME	RMSE	CRMSE	Correlation
Original_08	33.85	1.97	1.97	3.20	2.52	0.28
Original_14	33.61	2.00	1.72	3.08	2.55	0.27
Original_17	32.82	1.99	0.94	2.86	2.70	0.18
Mosaic_08	33.89	1.81	2.01	3.19	2.48	0.25
Mosaic_14	33.23	1.90	1.35	2.89	2.56	0.24
Mosaic_17	32.44	2.15	0.55	2.86	2.81	0.18
Interpolated_08	32.76	1.74	0.87	2.56	2.41	0.28
Interpolated_14	33.14	1.88	1.26	2.78	2.48	0.28
Interpolated_17	32.21	2.07	0.33	2.79	2.77	0.17
Landsat	31.88	2.22	0	0	0	1

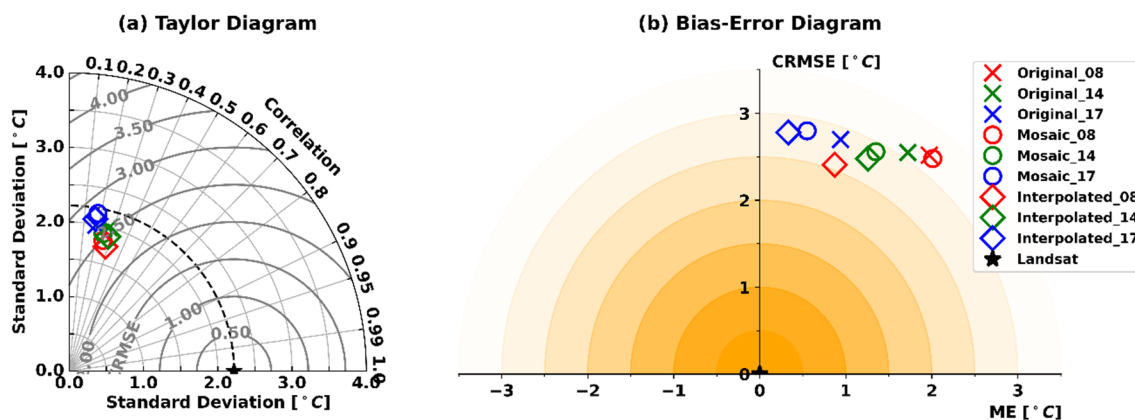


FIGURE 10 | Statistical comparison between scenarios and Landsat only over non-urban land cover on two diagrams: (a) Taylor diagram comparing the correlation between scenarios and Landsat, the standard deviation of scenarios and the Landsat (dashed line), and unbiased RMSE (CRMSE); and (b) Bias-Error diagram acting as a complimentary to the Taylor diagram, shows the error and bias from Landsat for all scenarios. The shaded orange wedges are 0.5 increments of RMSE values ($RMSE^2 = CRMSE^2 + ME^2$). [Colour figure can be viewed at wileyonlinelibrary.com]

scenarios. Considering the difference in prediction of AT over urban, non-urban, and coastline, it is hard to come up with a general conclusion for which scenario performed better in AT prediction over urban areas. However, since 50% of stations are located in urbanised areas, and their pattern is similar, Start_14 group scenarios and Interpolated_08 had better predictions than other scenarios when compared to ASOS and AWOS stations' observations.

Overall, spinup time had the strongest influence on AT evolution compared to incorporating the Mosaic approach or interpolating soil moisture and temperatures from the parent domain in the Interpolated scenarios. The Annapolis, Inner Harbour, and Stevensville stations were close to large water bodies. When we looked at the time series of AT for all scenarios and NOAA observations at these stations, we noticed considerable cool biases compared to the observations. As discussed earlier, WRF simulates the entire surface temperature of water bodies with an error of $\approx 2.5^\circ\text{C}$ compared to Landsat. Therefore, the inaccurate representations of water

temperatures may have led to cooler biases in AT near water bodies.

3.4 | Discussion and Synthesis

Section 2.2 proposed three questions to evaluate the influence of different initialization methods and incorporation of the Mosaic approach on fine-resolution simulations of the UHI effect using WRF-LES and SLUCM. Here, we answer those questions based on a synthesis of the comparison of model results to Landsat LST and ASOS and AWOS AT observations.

3.4.1 | How Does the Length of the Spinup Period Influence the AT and LST Predictions?

The length of the spinup period strongly influenced the predictions of LST and AT across scenarios. The more extended spinup period in the Start_08 scenarios, which captured several

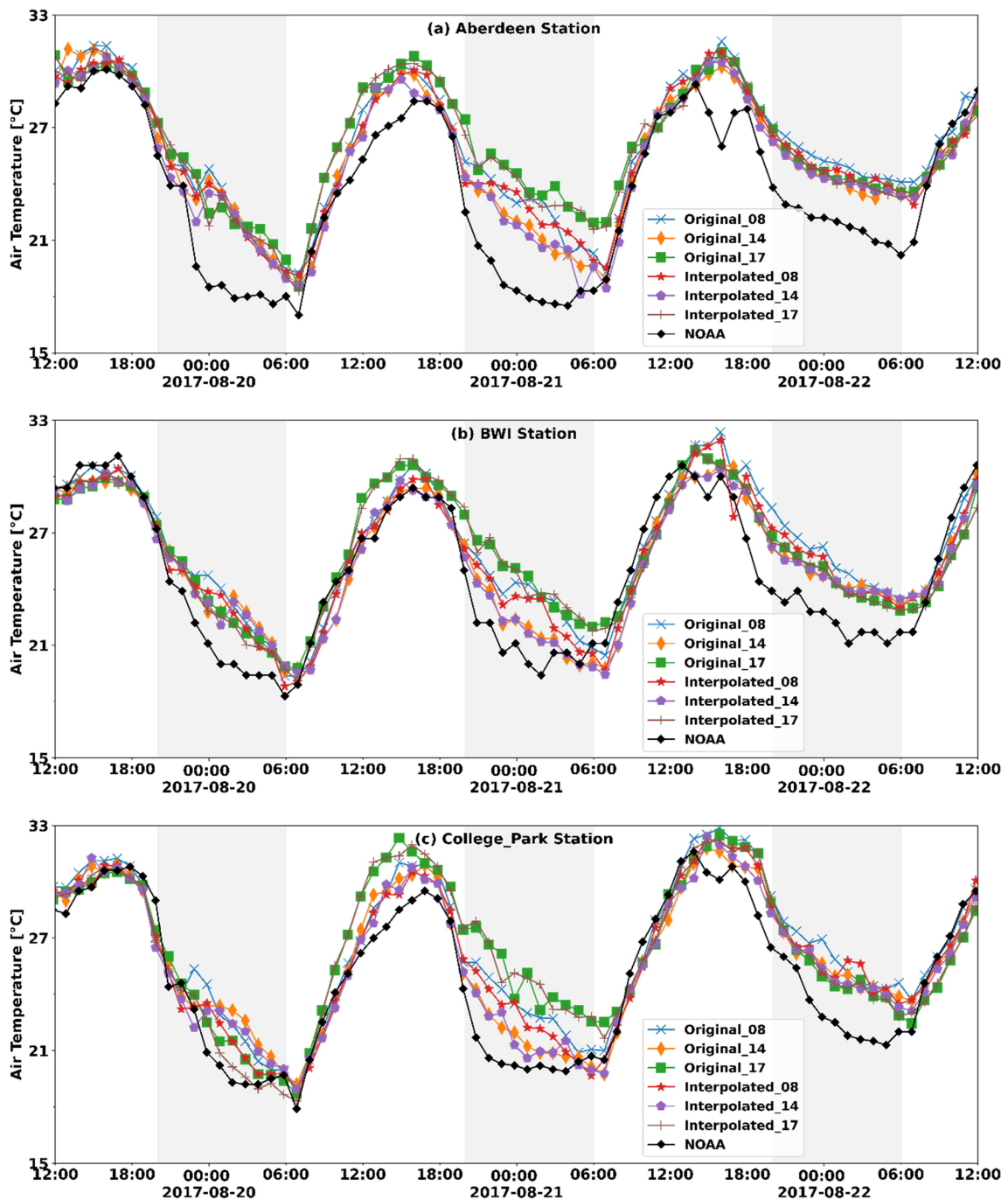


FIGURE 11 | Time series of 2-m air temperature at (a) Aberdeen ASOS station, (b) BWI airport ASOS station, and (c) College Park airport AWOS station from Aug 19, 2017, 12:00 PM to Aug 22, 2017, 12:00 PM. The figure shows only the Original and Interpolated scenarios and NOAA ASOS and AWOS stations. [Colour figure can be viewed at wileyonlinelibrary.com]

intense rain events, improved LST prediction in all three groups (Original, Mosaic, and Interpolated). Start₁₄ scenarios, despite having almost half of the leading time, had a closer performance to Start₀₈ scenarios in LST prediction. Start₁₇ scenarios performed poorly in LST prediction.

In predicting AT evolution from Aug 19, 2017, 11:00 AM EST to Aug 22, 2017, 12:00 AM EST, compared to eight ASOS and AWOS stations across the domain, Start₁₄ scenarios performed better than Start₀₈ and Start₁₇ scenarios. Start₁₇ scenarios performed worse than the other scenarios (at five stations) except at two stations (Inner Harbour and Annapolis). The

two-day spinup period in the Start₁₇ scenarios thus appear to be too short for adequate representation of the initial conditions, while the 5 and 11 days spinup in the other two sets had closer agreement.

3.4.2 | How Does the Mosaic Approach Influence AT or LST Predictions at Fine Scales?

While previous studies have shown significant simulation improvement by incorporating the Mosaic approach at high mesoscale resolutions (1 km), this approach did not improve the

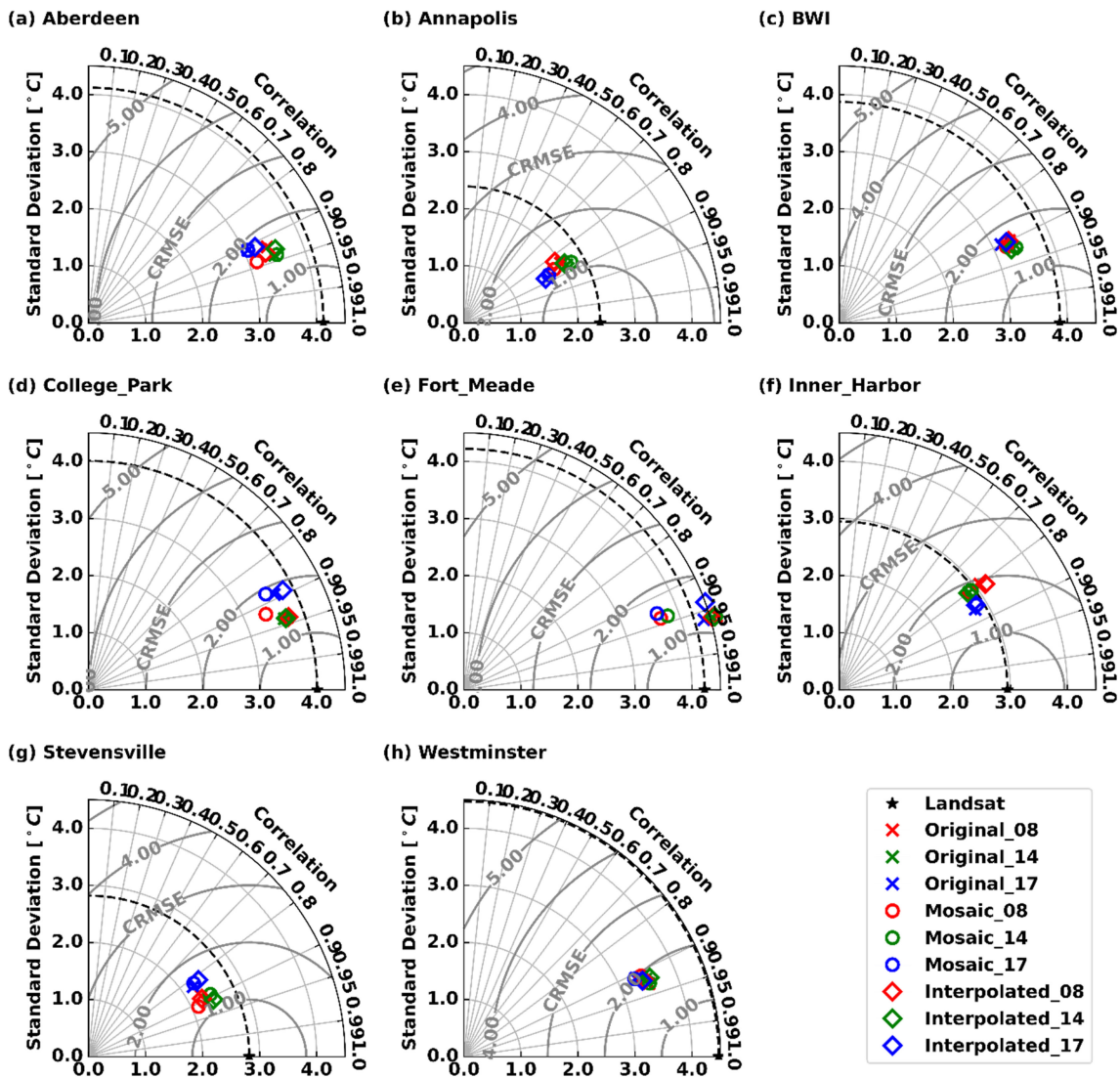


FIGURE 12 | Taylor diagrams depicting statistical comparisons of 2-m air temperature between scenarios and NOAA AWOS and ASOS stations. The stations are located across the domain shown in Figure 1b. The bottom axis is the standard deviation in degrees Celsius removed from the subfigures to save space. The dashed line shows the STD value for the NOAA stations. [Colour figure can be viewed at wileyonlinelibrary.com]

LST or AT predictions appreciably at fine-scale WRF-LES resolutions of 150 m. Although at a 150 m resolution, there are still urban areas where land cover is highly variable in a single computational cell, the subgrid-scale parameterization of these heterogeneous energy and water fluxes represented by the Mosaic approach does not appear to influence the predictions of WRF significantly. This could be due to compensating errors across cells at such fine-scale resolutions. For example, if some urban terrain is ignored in a given cell without the Mosaic approach, this may be compensated for by an overrepresentation of urban terrain in adjacent cells.

3.4.3 | How Does Initializing the Innermost Domains (d4 and d5) With Interpolated Soil Moisture From the Parent Domain Affect LST and AT Evolution?

As shown in section 3.1, interpolating soil moisture and temperature from parent domains (3 and 4) to domains 4 and 5 at their start time on Aug 19, 2017, 8:00 AM EST and Aug 19,

2017, 11:00 AM EST, resulted in significant changes in spatio-temporal soil moisture evolution between Interpolated scenarios on the one hand, and Original and Mosaic scenarios on the other hand. The interpolation method yielded differences among Interpolated scenarios as well. The Interpolated_08 had a much higher SMC than all other scenarios over the simulation period. Interpolated_14 and Interpolated_17 started similarly from much less SMC than other scenarios at the first layer. Interpolation improved the RMSE and Pearson correlation values in Interpolated_14 and Interpolated_08 compared to all other scenarios compared to Landsat LST. For AT evolution, interpolated_08 had the lowest ME values.

4 | Conclusions

Our research underscores the sensitivity of high-resolution WRF-LES coupled with single-layer urban canopy models (SLUCM) to soil moisture initialization in simulating urban microclimates and the urban heat island (UHI) effect. We

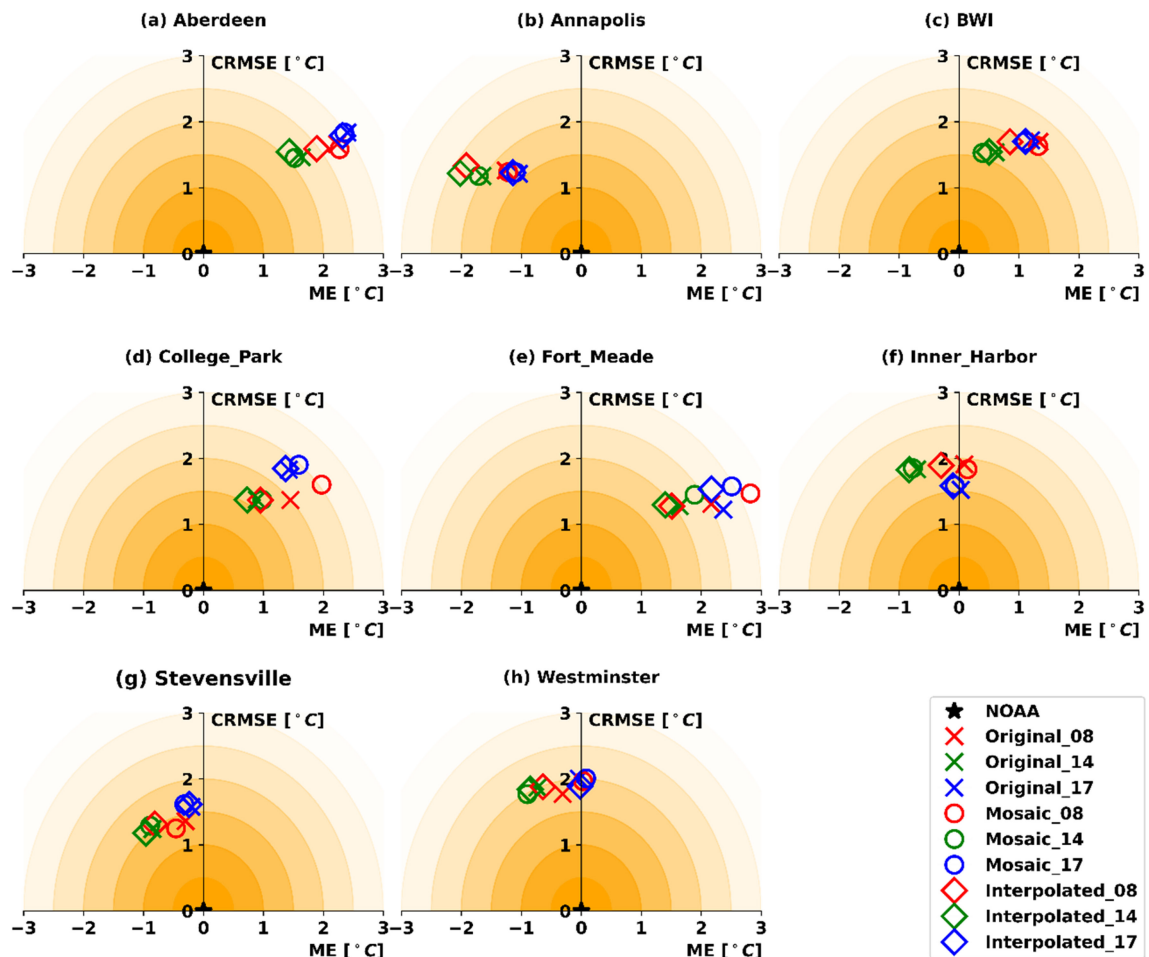


FIGURE 13 | Error-Bias diagrams depicting statistical comparisons of 2-m air temperature between scenarios and NOAA AWOS and ASOS stations. The stations are located across the domain shown in Figure 1b. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.com)]

demonstrated the critical influence of spinup lead time and soil moisture interpolation on land surface temperature (LST) and air temperature (AT) predictions, highlighting the importance of accurate initialization for capturing the dynamics of urban thermal environments under varying climatic conditions.

While the Mosaic approach offered limited benefits at the fine-scale resolution of our study, it remains a valuable tool for resolving subgrid-scale processes in coarser simulations. Our findings also underscore the need for further refinement of WRF’s representation of surface water bodies, a crucial component for accurately simulating urban microclimates.

Our comparison of high-resolution SST data with the Original NARR SST revealed negligible differences between simulations, despite significant SST variations. This contradicts expectations that such SST discrepancies should manifest in more pronounced atmospheric responses. Future research could examine the effect of improved representation of water body physics and temperatures, and exchanges with the air aloft, on atmospheric dynamics within WRF to better capture the complex interactions between these systems, ensuring more accurate and reliable simulations of urban microclimates and their response to climate change for coastal areas.

Ultimately, this study advances our understanding of the complex interactions between urban surfaces, atmospheric processes, and soil moisture, providing valuable insights for refining urban hydrometeorological models. Despite the fact that initialization is a lesser concern in hydroclimatological models (that run for a long time) compared to their hydrometeorological counterparts, the findings also have two climatological implications. First, climate models will still miss the significant urban surface variability that emerged in our fine scale resolution, and hence Mosaic and similar approaches are crucial for climatological applications over cities. Second, climatological downscaling of extreme events of short duration, which are essentially aiming to capture extreme weather, need to carefully design the initialization of the model for the downscaling. By improving the accuracy and reliability of such downscaling models, we can better assess the impact of climate change on urban environments, develop targeted mitigation strategies, and inform urban planning decisions to enhance resilience and sustainability in the face of escalating extreme heat events.

Author Contributions

Mahdad Talebpour: conceptualization, methodology, software, data curation, formal analysis, validation, investigation, visualization, project administration, writing – review and editing, writing – original draft. **Elie Bou-Zeid:** conceptualization, methodology, investigation,

supervision, funding acquisition, project administration, writing – review and editing, writing – original draft, resources, formal analysis. **Claire Welty:** funding acquisition, project administration, resources, writing – review and editing, supervision, conceptualization, writing – review and editing, writing – original draft. **Dan Li:** methodology, software, writing – review and editing. **Benjamin Zaitchik:** conceptualization, writing – review and editing, project administration, resources, supervision, funding acquisition.

Acknowledgements

This work was funded under multiple projects, including the (1) Urban Water Innovation Network project under National Science Foundation Cooperative Agreement #1444758, (2) Johns Hopkins University Glenadore and Howard L. Pim Postdoctoral Fellowship in Global Change, (3) National Science Foundation Cooperative Agreement #2012340, and (4) Baltimore Social-Environmental Collaborative Urban Integrated Field Laboratory under Department of Energy Agreement #DE-SC0023218. We also want to acknowledge UMBC's large Google Drive for backup and storage. We specifically thank Darryn Waugh for his helpful discussions.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Allouche, M., E. Bou-Zeid, and J. Iipponen. 2023. “The Influence of Synoptic Wind on Land–Sea Breezes.” *Quarterly Journal of the Royal Meteorological Society* 149, no. 757: 3198–3219. <https://doi.org/10.1002/qj.4552>.

Arnfield, A. J. 2003. “Two Decades of Urban Climate Research: A Review of Turbulence, Exchanges of Energy and Water, and the Urban Heat Island.” *International Journal of Climatology* 23, no. 1: 1–26. <https://doi.org/10.1002/joc.859>.

Bassett, R., X. Cai, L. Chapman, C. Heaviside, and J. E. Thornes. 2019. “Semi-Idealized Urban Heat Advection Simulations Using the Weather Research and Forecasting Mesoscale Model.” *International Journal of Climatology* 39, no. 3: 1345–1358. <https://doi.org/10.1002/joc.5885>.

Brown, S. A., M. Folk, G. Goucher, R. Rew, and P. F. Dubois. 1993. “Software for Portable Scientific Data Management.” *Computers in Physics* 7, no. 3: 304–308. <https://doi.org/10.1063/1.4823180>.

Caswell, T. A., E. S. de Andrade, A. Lee, et al. 2023. *matplotlib/matplotlib: REL: v3.7.4*. Zenodo. <https://doi.org/10.5281/zenodo.10152802>.

Chen, F., and J. Dudhia. 2001a. “Coupling an Advanced Land Surface–Hydrology Model With the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity.” *Monthly Weather Review* 129, no. 4: 569–585. [https://doi.org/10.1175/1520-0493\(2001\)129<0569:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2).

Chen, F., and J. Dudhia. 2001b. “Coupling an Advanced Land Surface–Hydrology Model With the Penn State–NCAR MM5 Modeling System. Part II: Preliminary Model Validation.” *Monthly Weather Review* 129, no. 4: 587–604. <https://doi.org/10.1175/1520-0493129<0587:CAALSH>2.0.CO;2>.

Chen, F., H. Kusaka, R. Bornstein, et al. 2011. “The Integrated WRF/Urban Modelling System: Development, Evaluation, and Applications to Urban Environmental Problems.” *International Journal of Climatology* 31, no. 2: 273–288. <https://doi.org/10.1002/joc.2158>.

Chen, F., and Y. Zhang. 2009. “On the Coupling Strength Between the Land Surface and the Atmosphere: From Viewpoint of Surface

Exchange Coefficients.” *Geophysical Research Letters* 36, no. 10. <https://doi.org/10.1029/2009GL037980>.

Chin, T. M., J. Vazquez-Cuervo, and E. M. Armstrong. 2017. “A Multi-Scale High-Resolution Analysis of Global Sea Surface Temperature.” *Remote Sensing of Environment* 200: 154–169. <https://doi.org/10.1016/j.rse.2017.07.029>.

Computational and Information Systems Laboratory. 2019. *Cheyenne: HPE/SGI ICE XA System*. Boulder: University Community Computing, CO: NSF National Center for Atmospheric Research. <https://doi.org/10.5065/D6RX99HX>.

Computational and Information Systems Laboratory. 2023. *Derecho: HPE/SGI ICE XA System*. Boulder: University Community Computing, CO: NSF National Center for Atmospheric Research. <https://doi.org/10.5065/qx9a-pg09>.

Courant, R., K. Friedrichs, and H. Lewy. 1967. “On the Partial Difference Equations of Mathematical Physics.” *IBM Journal of Research and Development* 11, no. 2: 215–234. <https://doi.org/10.1147/rd.112.0215>.

Daniels, M. H., K. A. Lundquist, J. D. Mirocha, D. J. Wiersema, and F. K. Chow. 2016. “A New Vertical Grid Nesting Capability in the Weather Research and Forecasting (WRF) Model.” *Monthly Weather Review* 144, no. 10: 3725–3747. <https://doi.org/10.1175/MWR-D-16-0049.1>.

Dennis, E. J., and E. H. Berbery. 2021. “The Role of Soil Texture in Local Land Surface–Atmosphere Coupling and Regional Climate.” *Journal of Hydrometeorology* 22, no. 2: 313–330. <https://doi.org/10.1175/JHM-D-20-0047.1>.

Dwyer, J. L., D. P. Roy, B. Sauer, C. B. Jenkerson, H. K. Zhang, and L. Lymburner. 2018. “Analysis Ready Data: Enabling Analysis of the Landsat Archive.” *Remote Sensing* 10, no. 9: 1–19. <https://doi.org/10.3390/rs10091363>.

Dy, C. Y., and J. C. H. Fung. 2016. “Updated Global Soil Map for the Weather Research and Forecasting Model and Soil Moisture Initialization for the Noah Land Surface Model.” *Journal of Geophysical Research* 121, no. 15: 8777–8800. <https://doi.org/10.1002/2015JD024558>.

Georgescu, M., M. Moustauoui, A. Mahalov, and J. Dudhia. 2011. “An Alternative Explanation of the Semiarid Urban Area “Oasis Effect”.” *Journal of Geophysical Research* 116: D24113. <https://doi.org/10.1029/2011JD016720>.

Hall, T. W., L. Blunn, S. Grimmond, et al. 2024. “Utility of Thermal Remote Sensing for Evaluation of a High-Resolution Weather Model in a City.” *Quarterly Journal of the Royal Meteorological Society* 150, no. 760: 1771–1790. <https://doi.org/10.1002/qj.4669>.

Hoffman, J. S., V. Shandas, and N. Pendleton. 2020. “The Effects of Historical Housing Policies on Resident Exposure to Intra-Urban Heat: A Study of 108 US Urban Areas.” *Climate* 8, no. 1: 1–15. <https://doi.org/10.3390/cli8010012>.

Homer, C., J. Dewitz, L. Yang, et al. 2015. “Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a Decade of Land Cover Change Information.” *Photogrammetric Engineering and Remote Sensing* 81, no. 5: 345–354.

Hoyer, S., and J. Hamman. 2017. “Xarray: N-D Labeled Arrays and Datasets in Python.” *Journal of Open Research Software* 5, no. 1: 10. <https://doi.org/10.5334/jors.148>.

Hsu, A., G. Sheriff, T. Chakraborty, and D. Manya. 2021. “Disproportionate Exposure to Urban Heat Island Intensity Across Major US Cities.” *Nature*. *Communications* 12, no. 1: 2721. <https://doi.org/10.1038/s41467-021-22799-5>.

Jandaghian, Z., and U. Berardi. 2020. “Comparing Urban Canopy Models for Microclimate Simulations in Weather Research and Forecasting Models.” *Sustainable Cities and Society* 55: 102025. <https://doi.org/10.1016/j.scs.2020.102025>.

- JPL MUR MEaSURES Project. 2020. "GHRSSST Level 4 MUR Global Foundation Sea Surface Temperature Analysis (v4.1)." <https://doi.org/10.5067/GHGMR-4FJ04>.
- Kamal, S., H.-P. Huang, and S. W. Myint. 2015. "The Influence of Urbanization on the Climate of the Las Vegas Metropolitan Area: A Numerical Study." *Journal of Applied Meteorology and Climatology* 54, no. 11: 2157–2177. <https://doi.org/10.1175/JAMC-D-15-0003.1>.
- Binita, K. C., J. M. Shepherd, and C. J. Gaither. 2015. "Climate Change Vulnerability Assessment in Georgia." *Applied Geography* 62: 62–74. <https://doi.org/10.1016/j.apgeog.2015.04.007>.
- Kirkil, G., J. Mirocha, E. Bou-Zeid, F. K. Chow, and B. Kosović. 2012. "Implementation and Evaluation of Dynamic Subfilter-Scale Stress Models for Large-Eddy Simulation Using WRF*." *Monthly Weather Review* 140, no. 1: 266–284. <https://doi.org/10.1175/MWR-D-11-00037.1>.
- Kluyver, T., B. Ragan-Kelley, F. Pérez, et al. 2016. "Jupyter Notebooks – A Publishing Format for Reproducible Computational Workflows," *In Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited by F. Loizides and B. Schmidt, 87–90. Amsterdam: IOS Press. <https://doi.org/10.3233/978-1-61499-649-1-87>.
- Kusaka, H., H. Kondo, Y. Kikegawa, and F. Kimura. 2001. "A Simple Single-Layer Urban Canopy Model for Atmospheric Models: Comparison With Multi-Layer and Slab Models." *Boundary-Layer Meteorology* 101, no. 3: 329–358. <https://doi.org/10.1023/A:1019207923078>.
- Ladwig, William. 2017. "wrf-python (Version 1.3.3)." In *Software*. Colorado: UCAR/NCAR: Boulder. <https://doi.org/10.5065/D6W094P1>.
- Li, D., and E. Bou-Zeid. 2013. "Synergistic Interactions Between Urban Heat Islands and Heat Waves: The Impact in Cities Is Larger Than the Sum of Its Parts." *Journal of Applied Meteorology and Climatology* 52, no. 9: 2051–2064. <https://doi.org/10.1175/JAMC-D-13-02.1>.
- Li, D., and E. Bou-Zeid. 2014. "Quality and Sensitivity of High-Resolution Numerical Simulation of Urban Heat Islands." *Environmental Research Letters* 9, no. 5. <https://doi.org/10.1088/1748-9326/9/5/055001>.
- Li, D., E. Bou-Zeid, M. Barlage, F. Chen, and J. A. Smith. 2013. "Development and Evaluation of a Mosaic Approach in the WRF-Noah Framework." *Journal of Geophysical Research: Atmospheres* 118, no. 21: 11,918–11,935. <https://doi.org/10.1002/2013JD020657>.
- Li, D., E. Bou-Zeid, and M. Oppenheimer. 2014. "The Effectiveness of Cool and Green Roofs as Urban Heat Island Mitigation Strategies." *Environmental Research Letters* 9, no. 5: 55002. <https://doi.org/10.1088/1748-9326/9/5/055002>.
- Li, H., Y. Zhou, X. Wang, X. Zhou, H. Zhang, and S. Sodoudi. 2019. "Quantifying Urban Heat Island Intensity and Its Physical Mechanism Using WRF/UCM." *Science of the Total Environment* 650: 3110–3119. <https://doi.org/10.1016/j.scitotenv.2018.10.025>.
- Liu, N., and L. Morawska. 2020. "Modeling the Urban Heat Island Mitigation Effect of Cool Coatings in Realistic Urban Morphology." *Journal of Cleaner Production* 264: 121560. <https://doi.org/10.1016/j.jclepro.2020.121560>.
- Mesinger, F., G. DiMego, E. Kalnay, et al. 2006. "North American Regional Reanalysis." *Bulletin of the American Meteorological Society* 87, no. 3: 343–360. <https://doi.org/10.1175/BAMS-87-3-343>.
- Miao, S., F. Chen, M. A. LeMone, M. Tewari, Q. Li, and Y. Wang. 2009. "An Observational and Modeling Study of Characteristics of Urban Heat Island and Boundary Layer Structures in Beijing." 48: 484–501. <https://doi.org/10.1175/2008JAMC1909.1>.
- Ming, Y., Y. Liu, X. Liu, and Z. Tian. 2024. "Demographic Disparity in Diurnal Surface Urban Heat Island Exposure Across Local Climate Zones: A Case Study of Chongqing, China." *Science of the Total Environment* 923: 171203. <https://doi.org/10.1016/j.scitotenv.2024.171203>.
- Mirocha, J. D., and K. A. Lundquist. 2017. "Assessment of Vertical Mesh Refinement in Concurrently Nested Large-Eddy Simulations Using the Weather Research and Forecasting Model." *Monthly Weather Review* 145, no. 8: 3025–3048. <https://doi.org/10.1175/MWR-D-16-0347.1>.
- Oke, T. R., G. Mills, A. Christen, and J. A. Voogt. 2017. *Urban Climates*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/9781139016476>.
- Ortiz, L. E., J. E. González, R. Horton, et al. 2019. "High-Resolution Projections of Extreme Heat in New York City." *International Journal of Climatology* 39, no. 12: 4721–4735. <https://doi.org/10.1002/joc.6102>.
- Ramamurthy, P., and E. Bou-Zeid. 2014. "Contribution of Impervious Surfaces to Urban Evaporation." *Water Resources Research* 50, no. 4: 2889–2902. <https://doi.org/10.1002/2013WR013909>.
- Ramamurthy, P., and E. Bou-Zeid. 2017. "Heatwaves and Urban Heat Islands: A Comparative Analysis of Multiple Cities." *Journal of Geophysical Research: Atmospheres* 122, no. 1: 168–178. <https://doi.org/10.1002/2016JD025357>.
- Ramamurthy, P., E. Bou-Zeid, J. A. Smith, et al. 2014. "Influence of Subfacet Heterogeneity and Material Properties on the Urban Surface Energy Budget." *Journal of Applied Meteorology and Climatology* 53, no. 9: 2114–2129. <https://doi.org/10.1175/JAMC-D-13-0286.1>.
- Rew, R., and G. Davis. 1990. "NetCDF: An Interface for Scientific Data Access." *IEEE Computer Graphics and Applications* 10, no. 4: 76–82. <https://doi.org/10.1109/38.56302>.
- Ronda, R. J., G. J. Steeneveld, B. G. Heusinkveld, J. J. Attema, and A. A. M. Holtslag. 2017. "Urban Finescale Forecasting Reveals Weather Conditions With Unprecedented Detail." *Bulletin of the American Meteorological Society* 98, no. 12: 2675–2688. <https://doi.org/10.1175/BAMS-D-16-0297.1>.
- Rosenzweig, C., W. Solecki, and R. Slosberg. 2006. "Mitigating New York City's Heat Island With Urban Forestry, Living Roofs, and Light Surfaces." *86th AMS Annual Meeting*.
- Ryu, Y.-H., E. Bou-Zeid, Z. H. Wang, and J. A. Smith. 2016. "Realistic Representation of Trees in an Urban Canopy Model." *Boundary-Layer Meteorology* 159, no. 2: 193–220. <https://doi.org/10.1007/s10546-015-0120-y>.
- Salamanca, F., A. Martilli, and C. Yagüe. 2012. "A Numerical Study of the Urban Heat Island Over Madrid During the DESIREX (2008) Campaign With WRF and an Evaluation of Simple Mitigation Strategies." *International Journal of Climatology* 32, no. 15: 2372–2386. <https://doi.org/10.1002/joc.3398>.
- Santanello, J. A., P. Lawston, S. Kumar, and E. Dennis. 2019. "Understanding the Impacts of Soil Moisture Initial Conditions on NWP in the Context of Land-Atmosphere Coupling." *Journal of Hydrometeorology* 20, no. 5: 793–819. <https://doi.org/10.1175/JHM-D-18-0186.1>.
- Scott, A. A., B. Zaitchik, D. W. Waugh, and K. O'Meara. 2017. "Intraurban Temperature Variability in Baltimore." *Journal of Applied Meteorology and Climatology* 56, no. 1: 159–171. <https://doi.org/10.1175/JAMC-D-16-0232.1>.
- Shaffer, S. R., W. T. L. Chow, M. Georgescu, et al. 2015. "Multiscale Modeling and Evaluation of Urban Surface Energy Balance in the Phoenix Metropolitan Area." *Journal of Applied Meteorology and Climatology* 54, no. 2: 322–338. <https://doi.org/10.1175/JAMC-D-14-0051.1>.
- Shahmohamadi, P., A. I. Che-Ani, K. N. A. Maulud, N. M. Tawil, and N. A. G. Abdullah. 2011. "The Impact of Anthropogenic Heat on Formation of Urban Heat Island and Energy Consumption Balance." *Urban Studies Research* 2011: 1–9. <https://doi.org/10.1155/2011/497524>.
- Sharma, A., P. Conry, H. J. S. Fernando, A. F. Hamlet, J. J. Hellmann, and F. Chen. 2016. "Green and Cool Roofs to Mitigate Urban Heat Island Effects in the Chicago Metropolitan Area: Evaluation With a Regional Climate Model." *Environmental Research Letters* 11, no. 6: 064004. <https://doi.org/10.1088/1748-9326/11/6/064004>.
- Sharma, A., H. J. S. Fernando, A. F. Hamlet, J. J. Hellmann, M. Barlage, and F. Chen. 2017. "Urban Meteorological Modeling Using WRF: A

- Sensitivity Study.” *International Journal of Climatology* 37, no. 4: 1885–1900. <https://doi.org/10.1002/joc.4819>.
- Shi, R., B. F. Hobbs, B. F. Zaitchik, D. W. Waugh, A. A. Scott, and Y. Zhang. 2021. “Monitoring Intra-Urban Temperature With Dense Sensor Networks: Fixed or Mobile? An empirical study in Baltimore, MD.” *Urban Climate* 39: 100979. <https://doi.org/10.1016/j.uclim.2021.100979>.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, et al. 2008. “A Description of the Advanced Research WRF Version 3.” *NCAR Technical Note*. <https://doi.org/10.5065/D68S4MVH>.
- Sobstyl, J. M., T. Emig, M. J. A. Qomi, F. J. Ulm, and R. J. M. Pellenq. 2018. “Role of City Texture in Urban Heat Islands at Nighttime.” *Physical Review Letters* 120. <https://doi.org/10.1103/PhysRevLett.120.108701>.
- Talbot, C., E. Bou-Zeid, and J. Smith. 2012. “Nested Mesoscale Large-Eddy Simulations With WRF: Performance in Real Test Cases.” *Journal of Hydrometeorology* 13, no. 5: 1421–1441. <https://doi.org/10.1175/JHM-D-11-048.1>.
- Talebpour, M., C. Welty, and E. Bou-Zeid. 2021. “Development and Testing of a Fully-Coupled Subsurface-Land Surface-Atmosphere Hydrometeorological Model: High-Resolution Application in Urban Terrains.” *Urban Climate* 40: 100985. <https://doi.org/10.1016/j.uclim.2021.100985>.
- Taylor, K. E. 2001. “Summarizing Multiple Aspects of Model Performance in a Single Diagram.” *Journal of Geophysical Research: Atmospheres* 106, no. D7: 7183–7192. <https://doi.org/10.1029/2000JD900719>.
- Van Rossum, G., and F. L. Drake. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- Venter, Z. S., H. Figari, O. Krange, and V. Gundersen. 2023. “Environmental Justice in a Very Green City: Spatial Inequality in Exposure to Urban Nature, Air Pollution and Heat in Oslo, Norway.” *Science of the Total Environment* 858: 160193. <https://doi.org/10.1016/j.scitotenv.2022.160193>.
- Virtanen, P., R. Gommers, T. E. Oliphant, et al. 2020. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.” *Nature Methods* 17, no. 3: 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
- Wang, J., and X.-M. Hu. 2021. “Evaluating the Performance of WRF Urban Schemes and PBL Schemes Over Dallas–Fort Worth During a Dry Summer and a Wet Summer.” *Journal of Applied Meteorology and Climatology* 60: 779–798. <https://doi.org/10.1175/JAMC-D-19-0195.1>.
- Wiersema, D. J., K. A. Lundquist, and F. K. Chow. 2020. “Mesoscale to Microscale Simulations Over Complex Terrain With the Immersed Boundary Method in the Weather Research and Forecasting Model.” *Monthly Weather Review* 148, no. 2: 577–595. <https://doi.org/10.1175/mwr-d-19-0071.1>.
- Wilson, B. 2020. “Urban Heat Management and the Legacy of Redlining.” *Journal of the American Planning Association* 86, no. 4: 443–457. <https://doi.org/10.1080/01944363.2020.1759127>.
- Wong, K. V., A. Paddon, and A. Jimenez. 2013. “Review of World Urban Heat Islands: Many Linked to Increased Mortality.” *Journal of Energy Resources Technology* 135, no. 2: 22101–22102. <https://doi.org/10.1115/1.4023176/365904>.
- Wynngaard, J. C. 2004. “Toward Numerical Modeling in the “Terra Incognita”.” *Journal of the Atmospheric Sciences* 61, no. 14: 1816–1826. [https://doi.org/10.1175/1520-0469\(2004\)061<1816:TNMITT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<1816:TNMITT>2.0.CO;2).
- Yang, J., and E. Bou-Zeid. 2019. “Scale Dependence of the Benefits and Efficiency of Green and Cool Roofs.” *Landscape and Urban Planning* 185: 127–140. <https://doi.org/10.1016/J.LANDURBPLAN.2019.02.004>.
- Yang, J., and Z.-H. Wang. 2014. “Physical Parameterization and Sensitivity of Urban Hydrological Models: Application to Green Roof Systems.” *Building and Environment* 75: 250–263. <https://doi.org/10.1016/J.BUILDENV.2014.02.006>.
- Yang, J., Z. H. Wang, F. Chen, et al. 2015. “Enhancing Hydrologic Modelling in the Coupled Weather Research and Forecasting–Urban Modelling System.” *Boundary-Layer Meteorology* 155, no. 1: 87–109. <https://doi.org/10.1007/s10546-014-9991-6>.
- Yang, L., F. Qian, D. X. Song, and K. J. Zheng. 2016. “Research on Urban Heat-Island Effect.” *Procedia Engineering* 169: 11–18. <https://doi.org/10.1016/J.PROENG.2016.10.002>.
- Zhang, W., J. Yang, L. Yang, and D. Niyogi. 2022. “Impacts of City Shape on Rainfall in Inland and Coastal Environments’, Earth’s.” *Futures* 10, no. 5: e2022EF002654. <https://doi.org/10.1029/2022EF002654>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Appendix A

Sea Surface Temperature

A.1 | Sea Surface Temperature Representation in WRF

The Weather Research and Forecasting (WRF) model typically uses a fixed initial sea surface temperature (SST) for the entire simulation period, suitable for short-term forecasts where SST changes are minimal. However, for longer simulations, this approach can introduce inaccuracies due to the dynamic nature of SST.

In our study, we investigated the impact of SST resolution on WRF simulations within domain 3 (1350 m resolution). We compared two scenarios: one using the coarser resolution 32 km NARR SST data and another using the higher resolution 1 km MUR SST (JPL MUR MEASURES Project 2020) data. The comparison points to a low sensitivity of our results to varying SST datasets, even over large bodies of water like the Chesapeake Bay where the coarser NARR data resulted in uniform SST values.

A.2 | Comparing the Results With Different SST Products

We analysed the hourly spatial evolution of wind fields across two scenarios with differing SST inputs. Both scenarios shared identical initializations for domains 1–3 and ran from August 8th, 2017, 11 AM to August 22, 2017, 12 ET. Figure A1 presents three wind pseudocolour maps overlaid with wind vectors, capturing snapshots at the end of the simulation period: (1) August 21, 4 ET, (2) August 21, 20 ET, and (3) August 22, 12 ET to illustrate diurnal variations. To gauge the maximum divergence between scenarios, we selected snapshot toward the end of the simulation. Despite minor differences in wind speed patterns over the Atlantic Ocean, the overall wind field remains largely consistent across both scenarios after 14 days of simulation.

To further investigate SST’s influence, we compared wind vector plots overlaid on surface temperature maps at the same time points (Figure S1). A notable $\approx 2^\circ\text{C}$ difference in SST emerged between scenarios, particularly over the Southeastern Atlantic Ocean, resulting in localised shifts in mesoscale wind patterns. However, the broader surface temperature and wind patterns remained similar throughout the simulation period.

To assess if the slight shifts in large-scale wind pattern between the two simulations with different SST can have a significant impact on microscale circulations and atmospheric properties, we also compared 2-m air temperature (AT) between simulations. Figure A2 presents time series of AT for eight ASOS/AWOS stations (Figure 1b) from August 19th to 22nd, 2017. Both scenarios generally captured the observed AT trends, with minor discrepancies mainly occurring at night. Nighttime differences in AT were negligible across most stations for the entire period, except for a few nighttime hours at Aberdeen, Annapolis, BWI, and Fort Meade. The differences between the two simulations are, in general, smaller than the differences of either compared to observations.

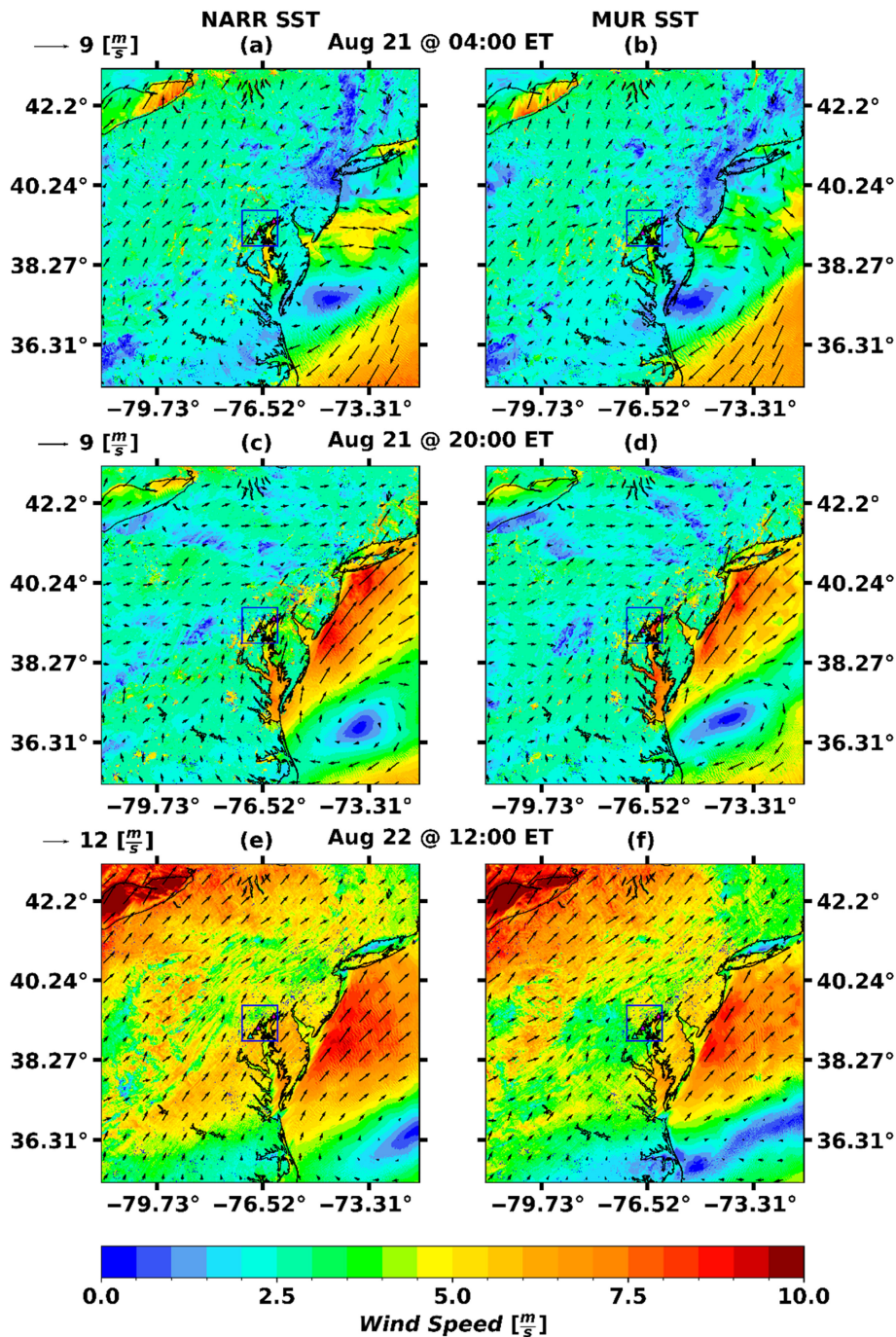


FIGURE A1 | Wind field vector overlaying the wind speed map over for domain 3. The left and right panels correspond to simulations initialized by NARR and MUR SST. The three timestamps were chosen deliberately to demonstrate wind fields at different times. Blue box identifies domain 5 and the coloured triangles show the ASOS and AWOS stations similar to Figure 1b. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

To investigate the underlying mechanisms, we also examined the wind fields throughout the simulation period. Figure S2 zooms in on domain 5 (Figure A1), focusing on the ASOS/AWOS stations. While minor differences in wind patterns emerged, particularly in the final snapshot (Figure S1e-f), the overall patterns remained consistent between scenarios across most time points.

Figure S3 illustrates wind vectors over surface temperature maps zoomed in on domain 5. Notably, the Chesapeake Bay consistently exhibited a $\approx 2^\circ\text{C}$ higher temperature in the MUR SST simulation compared to the Original NARR SST simulation. We analysed hourly snapshots of these overlays to understand the nighttime AT discrepancies observed at Aberdeen, Annapolis, BWI, and Fort Meade stations (Figure A2).

While space limitations preclude inclusion of all hourly snapshots, our analysis revealed that the combination of (1) surface temperature differences and (2) late afternoon wind speed variations between the two scenarios may explain the observed AT discrepancies. For instance, at Aberdeen station between August 21st, 20 ET, and August 22nd, 6 ET (Figure A2, third shaded area), the MUR SST simulation showed a $\approx 2^\circ\text{C}$ higher AT than the Original NARR SST. Concurrently, wind direction persisted from southwest to northeast, with the southwest region near Aberdeen exhibiting a $\approx 2^\circ\text{C}$ cooler LST in the Original NARR SST simulation. A similar pattern, albeit with reversed AT and LST differences, was observed near Annapolis station during the same period.

Despite these local (in space and time) differences at some stations associated with wind patterns shifts, throughout the simulation, both

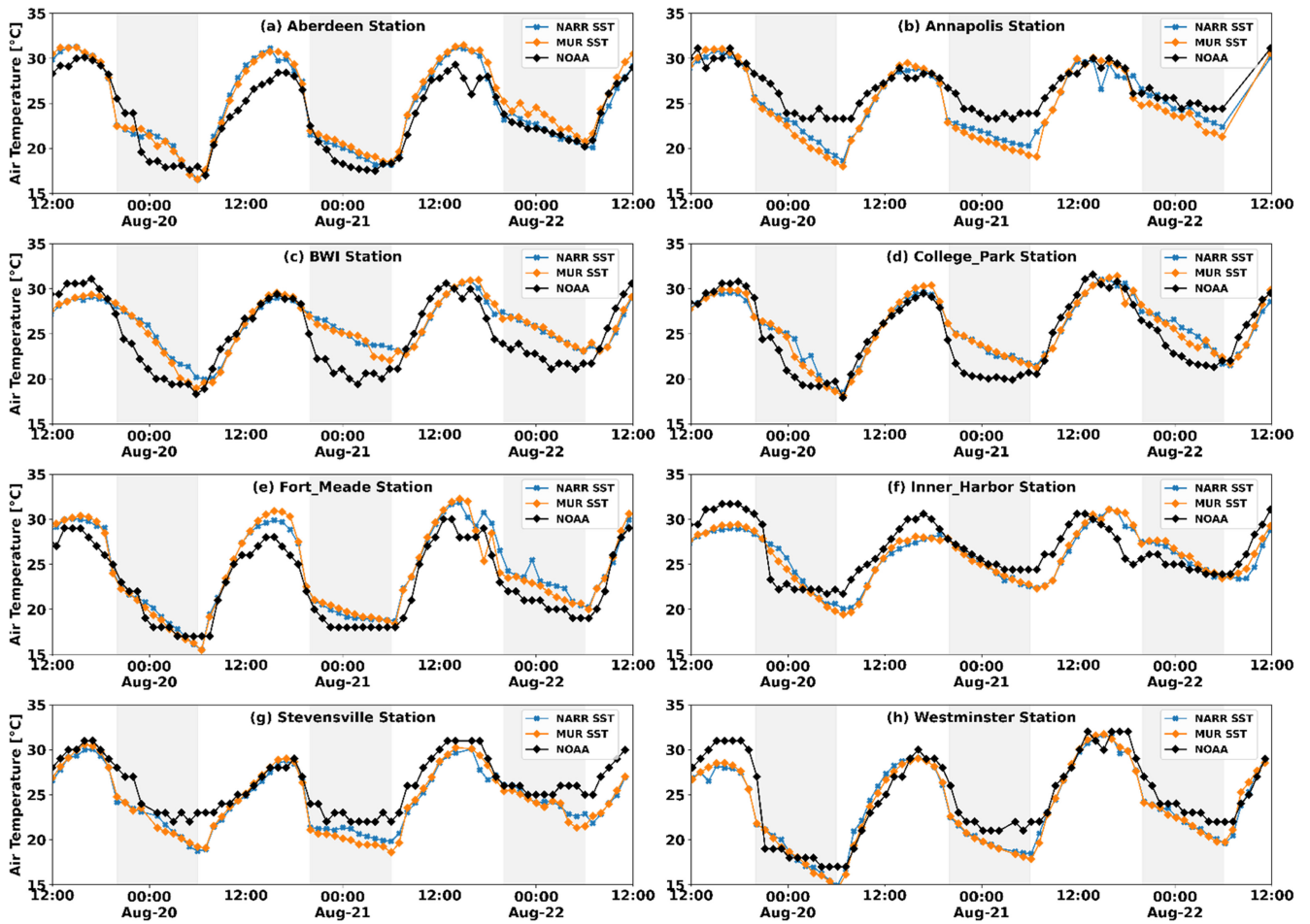


FIGURE A2 | Time series of 2-m air temperature, at all 8 ASOS and AWOS stations shown in Figure 1b, from Aug 19, 2017, 12:00 PM to Aug 22, 2017, 12:00 PM. The figure shows Original NARR SST and MUR SST scenarios. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.com)]

scenarios exhibited nearly identical air temperatures (AT) at all stations, with minor differences limited to a few hours, despite the MUR simulation's higher Chesapeake Bay temperature. This, however, may not be the case when the synoptic wind conditions are weaker. The development of land-sea breezes that would be affected by a 2°C change

in SST competes with synoptic pressure forcing that works to obliterate these circulations (Allouche, Bou-Zeid, and Iipponen 2023). The synoptic wind in our analysed periods is quite strong at over 6 m s^{-1} over the bay, and under such conditions SST may have a limited impact on the atmospheric dynamics.