remote sensing

MDPI

Article

Impacts of Remote-Sensing-Derived Terrain and Land-Use Datasets on Mesoscale-Model Simulations of an Extreme Heat Event over Henan, China: Sensitivity Experiments and Evaluation Against In Situ Observations and MODIS LST

Wei Deng [1,2], Changying Xue [1,3,*], Haibo Chen [1,3], Aifang Su [1,2], Hongwei Tian [1,3], Xiefei Zhi [4,5], Yiling Qi [1,2], Bingchen Li [6], Jie Wu [7], Mingjie Zhang [8,9], Haijiang Kong [1,2] and Yanling Zhang [10]

1 CMA/Henan Key Laboratory of Agrometeorological Support and Applied Technique, Zhengzhou 450003, China; dengthree1973@163.com (W.D.); chenhaibo1116@163.com (H.C.); suaf011@163.com (A.S.); thwenigma@163.com (H.T.); qiyiling710@163.com (Y.Q.); 15639791202@163.com (H.K.)
2 Henan Meteorological Observatory, Zhengzhou 450003, China
3 Henan Institute of Meteorological Sciences, Zhengzhou 450003, China
4 Key Laboratory of Meteorology Disaster, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044, China; zhi@nuist.edu.cn
5 Stellerus Technology Co., Ltd., Hong Kong SAR, China
6 Xingyang Meteorological Bureau, Xingyang 450100, China; lbc6@163.com
7 Zhengzhou Meteorological Bureau, Zhengzhou 450012, China; 13503991906@163.com
8 Anyang National Climate Observatory, Anyang 455000, China; 17372787182@163.com
9 Puyang Meteorological Bureau, Puyang 457000, China
10 Shangqiu Meteorological Bureau, Shangqiu 476000, China; nqzx2014@163.com
* Correspondence: xuecy9@163.com

Highlights

What are the main findings?

- Numerical model sensitivity experiments showed that terrain replacement produced hydrostatically consistent surface-pressure differences, while land-cover replacement produced stronger surface momentum and sensible heat exchange responses than terrain-only replacement. Direct interaction residuals indicated non-additive responses in simulated skin temperature (TSK) and 2 m air temperature (T2).
- For the examined event and model configuration, static land-surface datasets with higher nominal native resolution did not deliver consistent improvements in temperature simulation. Performance depended on the static-data configuration and differed between daytime and nighttime, with no single configuration consistently minimizing temperature errors across surface and near-surface evaluations.

What are the implications of the main findings?

- A diagnostic framework is proposed with two pathways for examining terrain-related and land-cover-related responses. The framework links WPS-mapped elevation (HGT_M) and dominant land-use category (LU_INDEX) differences with internal model responses, including TSK and T2. Combined with station and satellite evaluation, this framework distinguishes internal land–atmosphere responses from temperature performance and helps assess how remote-sensing-derived static inputs affect extreme heat simulations.

Academic Editor: Carmine Serio

Received: 24 August 2026
Revised: 29 October 2026
Accepted: 30 October 2026
Published: 31 October 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license.

- Stronger internal land–atmosphere responses did not necessarily improve agreement with station air-temperature observations or MODIS land-surface temperature. Remote-sensing-derived static datasets should therefore be selected based on their model-grid representation and evaluated temperature performance, rather than nominal native resolution alone. Applicability to other events and regions requires further testing.

Abstract

Remote-sensing-derived static terrain and land-cover datasets influence regional extreme heat simulations, but their effects on internal model responses and temperature performance require joint assessment. Five numerical experiments were conducted for the June 2022 Henan heatwave using five configurations drawn from three terrain datasets and four land-cover datasets. Atmospheric forcing and model physics were identical, and the inner grid spacing remained fixed at 3 km. Simulated skin temperature (TSK) was evaluated against Moderate Resolution Imaging Spectroradiometer (MODIS) land-surface temperature after pixel-level spatiotemporal matching. Simulated 2 m air temperature (T2) was evaluated against observations from 111 national meteorological stations. Two diagnostic pathways examined terrain-related and land-cover-related responses. Replacing MODIS15s-lake with MCD12Q1 changed the dominant land-use category in 22.2% of Henan grid cells. Daytime sensible heat flux difference fields had spatial root-mean-square (RMS) values exceeding 20 W/m^2. On the WRF Preprocessing System (WPS)-mapped 3 km grid, terrain fields derived from ASTER30 and GMTED5m had mean absolute differences of 7.5 and 22.8 m, respectively, relative to GMTED30s. Terrain-height and surface-pressure differences were strongly inversely correlated (R = −0.9990 to −0.9993), supporting hydrostatic consistency. Direct interaction residuals indicated non-additive TSK and T2 responses, and their full-period mean fields had spatial RMS values of 0.421 and 0.225 K, respectively. All configurations exhibited pooled TSK warm biases of 5.52–5.74 K, compared with an inter-configuration bias spread of approximately 0.22 K. No configuration consistently minimized temperature errors across variables and periods, and larger internal response magnitudes did not consistently correspond to smaller temperature errors. For this event and model configuration, products with higher nominal native resolution did not consistently improve temperature simulation; the independent resolution effect was not isolated. The framework combines internal diagnostics with station and satellite evaluation to assess remote-sensing-derived static inputs.

Keywords: remote-sensing-derived static datasets; digital elevation model (DEM); land use/land cover (LULC); land–atmosphere diagnostics; mesoscale-model simulations; extreme heat events

1. Introduction

Global warming has increased the frequency, duration, and magnitude of heatwaves worldwide, posing severe threats to terrestrial ecosystems, human health, and socioeconomic security [1,2]. Extreme high-temperature episodes have grown more frequent across China in recent decades, with the 2022 China Climate Bulletin recording the country's highest number of hot days (daily maximum temperature ≥ 35.0 °C) since 1961 [3]. Against this warming backdrop, Henan suffered a widespread, severe heatwave in mid-June 2022. This regional extreme high-temperature event provides a valuable case for investigating fine-scale heatwave processes and improving regional model simulations.

Existing research attributes the 2022 eastern China heatwave primarily to large-scale synoptic circulations, including an amplified western Pacific subtropical high and eastward-extended South Asian high [4,5]. However, comparatively less attention has been paid to how spatially heterogeneous land surfaces modulate regional extreme heat in model simulations. Biases in the representation of lower-boundary conditions remain poorly quantified, limiting efforts to accurately simulate the spatial patterns and intensity of regional extreme heat.

The Weather Research and Forecasting (WRF) model is widely adopted for heatwave simulations because of its capability to resolve regional circulations, planetary boundary layer (PBL) evolution, and land–atmosphere thermodynamic interactions [6,7]. Simulated temperatures are sensitive to physical parameterization schemes, soil moisture feedbacks, and prescribed surface properties. Previous studies have shown that combining remote-sensing data with data assimilation can improve air-temperature simulations during hot weather [8]. Even so, comparatively limited attention has been paid to static geographic inputs derived from remote sensing, which fundamentally define the surface environmental conditions represented within the model.

Remotely sensed digital elevation model (DEM) and land-cover products provide important static geographic inputs by representing terrain and land-surface heterogeneity. Terrain data determine model-grid elevation and topographic structure, while land-cover data prescribe surface roughness, albedo, and vegetation-related properties that regulate land–atmosphere momentum and energy exchanges [9–11]. Compared with conventional static geographic datasets, contemporary remote-sensing products can provide finer native spatial resolution and more recent information on terrain and land-surface conditions, which may improve the representation of grid-scale surface heterogeneity. However, simulation differences cannot be attributed to spatial resolution alone; they are also influenced by remote-sensing data quality, land-cover classification standards, and the model's representation of subgrid surface heterogeneity. Henan has pronounced topographic transitions and a fragmented mosaic of cropland and urban areas, making it a suitable region for examining how different remotely sensed terrain and land-cover datasets affect regional heatwave simulations.

Previous WRF studies have examined land-cover dataset sensitivity during heatwaves [12], while other work has investigated the effects of terrain and land-cover inputs on simulated meteorological variables [13]. Spatial aggregation and land-cover representation also influence how high-resolution products are incorporated into weather models [14]. While these studies provide a basis for examining static-data sensitivity beyond the nominal resolution of source products, further investigation is needed to distinguish the separate and non-additive temperature responses to terrain and land-cover dataset replacement under fixed model-grid settings. These responses should also be examined in relation to the geographic fields generated by the WRF Preprocessing System (WPS). This perspective can help assess static-data sensitivity during extreme heat events.

Station air-temperature observations and Moderate Resolution Imaging Spectroradiometer (MODIS) land-surface temperature (LST) provide complementary information for evaluating simulated thermal conditions. In situ air-temperature observations provide local measurements of near-surface temperature, whereas MODIS LST provides spatially distributed information on surface thermal patterns [15,16]. Simulated 2 m air temperature (T2) is therefore evaluated against station air-temperature observations, and simulated skin temperature (TSK) is evaluated against MODIS LST.

In this study, we investigate the sensitivity of heatwave simulations to remotely sensed terrain and land-cover datasets and examine the associated land–atmosphere thermodynamic and dynamical responses. One control run and four sensitivity experiments were conducted for the June 2022 Henan heatwave in order to assess the individual and

combined effects of replacing terrain and land-cover datasets. We compare the geographic fields generated from different static-data configurations and examine the corresponding land–atmosphere responses, relating differences in model-grid surface representation to simulated TSK and T2. Using station air-temperature observations and MODIS LST, we assess whether different remote-sensing static geographic inputs improve simulated thermal conditions. This work provides a physically based assessment of the applicability of different remote-sensing static datasets for regional extreme heat modeling.

2. Materials and Methods

2.1. Study Area and Extreme-Heat Event

Henan Province spans 31°23′–36°22′N and 110°21′–116°39′E and is located in the geographic and climatic transition zone connecting the North China Plain, Huang–Huai Plain, and Qinling–Funiu mountain ranges. Mountainous and hilly landscapes dominate western, southwestern, and northwestern Henan, while central and eastern areas are characterized by broad plains interspersed with croplands and urban patches. This surface heterogeneity makes Henan a suitable region for examining how terrain and land-cover datasets from different remote-sensing sources affect regional extreme heat simulations.

Daily maximum temperature records from 111 national meteorological stations were used to define the evaluation period and representative high-temperature day. Meteorological days follow the China Meteorological Administration (CMA) standard, from 21:00 Beijing Time (BJT; UTC + 8) to 20:00 BJT of the following calendar day. The 8-day period from 12 to 19 June 2022 was selected to cover the onset, intensification, peak, and recession of the regional extreme heat event. Among these dates, 16 June recorded the largest number of stations exceeding the 35 °C hot-day threshold and was therefore selected as the representative high-temperature day for spatial analysis.

2.2. Datasets and Static-Data Processing

This study integrates National Centers for Environmental Prediction (NCEP) Final (FNL) Operational Global Analysis data, three terrain datasets, four land-cover datasets, air-temperature observations from 111 national meteorological stations, and MODIS land-surface temperature (LST). For clarity, the original terrain and land-cover rasters are termed static datasets, whereas the grid variables generated after WPS/geogrid resampling are referred to as static geographic fields. Because resampling and aggregation modify the original remote-sensing information at the model-grid scale, terrain height (HGT_M) and dominant land-use category (LU_INDEX) are used as the primary indicators of terrain and land-use differences.

Six-hourly NCEP FNL Operational Global Analysis data at 1° × 1° resolution [17] were used to provide identical initial and lateral boundary conditions for all simulations.

Three terrain datasets were used for the inner D02 domain: the model-built-in 30 arc-second GMTED30s dataset [18], the customized 1 arc-second (~30 m) ASTER30 dataset derived from ASTER GDEM V3 [19], and the model-built-in 5 arc-minute GMTED5m dataset [18]. The land-cover datasets included three WPS built-in International Geosphere-Biosphere Program (IGBP)-MODIS layers: MODIS30s-lake, MODIS15s-lake, and MODIS5m-lake. Their native resolutions are 30 arc-seconds, 15 arc-seconds, and 5 arc-minutes, respectively [20,21]. A customized dataset derived from the LC_Type1 layer of the 2019 MCD12Q1 V061 product was also included [22,23]. LC_Type1 follows the IGBP classification system, with classes 1–17 corresponding to the first 17 categories of the WPS MODIS land-use scheme [22,24]. Satellite-derived land-cover datasets of this type have been widely used to update model land-use representations [25]. The MCD12Q1 input retained LC_Type1 codes 1–17, consistent with the corresponding category numbering in

the WRF IGBP-modified MODIS scheme. Class 17 denotes water bodies; no separate lake category (class 21) was introduced during preprocessing. Unclassified pixels were encoded as 255 and designated as missing values. Therefore, differences associated with MCD12Q1 reflect combined effects of data source, acquisition year, spatial resolution, and classification scheme rather than a single factor. For categorical land-use inputs, geogrid uses nearest_neighbor to calculate category fractions (LANDUSEF) from source pixels within each model cell; dominant = LU_INDEX assigns the dominant category subject to land–water handling. The 21-layer MODIS5m-lake input is defined as continuous, so nearest_neighbor is applied separately to each layer. The analysis focuses on LU_INDEX: geogrid-stage categories and their initialization-stage treatment are documented in Table S17, while land–water mask differences among experiments are reported in Table S18.

ASTER GDEM V3 incorporates void filling during product generation using supplementary elevation datasets. The downloaded tiles were mosaicked, clipped to the study domain, and processed in the EPSG:4326 geographic coordinate system before conversion into continuous binary tiles compatible with WPS. For MCD12Q1, LC_Type1 was extracted from each HDF tile, and the extracted tiles were mosaicked in the native sinusoidal projection. The mosaic was then reprojected to EPSG:4326 using nearest-neighbor resampling and cropped before conversion to categorical WPS binary tiles. The input-grid specifications and WPS settings for the six D02 datasets are summarized in Supplementary Tables S1–S3, and the corresponding index and case-specific GEOGRID. TBL and namelist.wps files are available in the shared dataset. For each of the five experiments, geogrid.exe was run with the corresponding namelist.wps, GEOGRID.TBL, and dataset-specific index files to generate a case-specific geo_em.d02.nc. HGT_M and LU_INDEX were then extracted as the geogrid-stage terrain-height and dominant land-use fields, respectively. The preprocessing workflow is shown in Figure 1. Raster preprocessing was performed in a Python 3.8 environment using pyModis 2.4.1, GDAL 3.4.3, and PROJ 8.2.1 [26–28], while WPS/geogrid performed model-grid mapping and static-field generation. The source ASTER and MCD12Q1 datasets covered an area larger than D02 and its buffer zone, ensuring complete coverage during WPS processing.

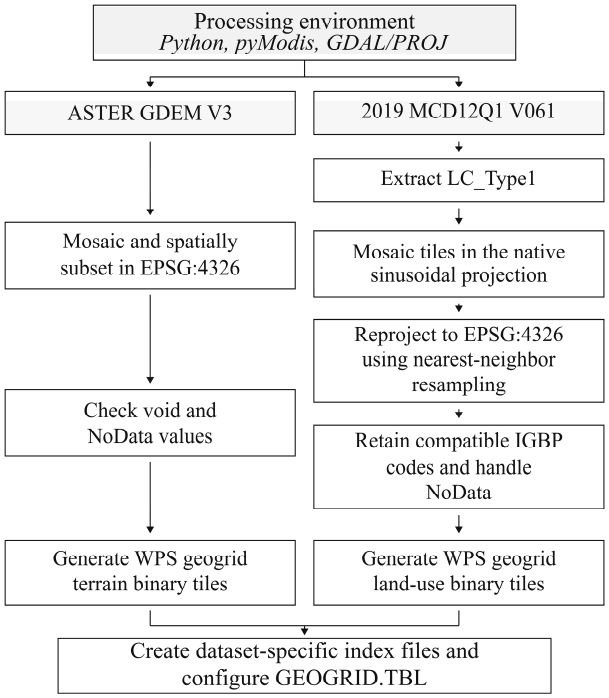


Figure 1. Workflow for generating static terrain and land-use datasets from ASTER and MCD12Q1 for use in a mesoscale model. Arrows indicate the processing sequence and data flow.

Air-temperature observations, including hourly air temperature and daily maximum and minimum temperatures, were obtained from 111 national meteorological stations in Henan Province through the Henan Meteorological Big Data Cloud Platform, Tianqing. Invalid and missing records were excluded. Hourly observations were converted to UTC, while Beijing Time (BJT) was retained for meteorological-day statistics. Air temperature was observed at 1.5 m above ground level, and daily maximum and minimum temperatures were taken directly from the operational daily extreme records.

The MODIS MOD11A1 V061 daily LST product was used to evaluate simulated skin temperature (TSK) from 12 to 19 June 2022. The product has a spatial resolution of approximately 1 km and was obtained from the NASA Land Processes Distributed Active Archive Center (LP DAAC) [29]. The extracted layers included LST_Day_1km, LST_Night_1km, QC_Day, QC_Night, Day_view_time, Night_view_time, Day_view_angl, and Night_view_angl. Pixels within Henan Province were retained according to the administrative boundary, with raw LST values converted using the scale factor of 0.02 K. Quality control was performed using QC_Day and QC_Night, and pixels were retained only if the QC flags indicated Mandatory QA = 0, Data Quality = 0, emissivity error ≤ 0.01, and LST error ≤ 1 K. Retained pixels also required LST values between 200 and 400 K and valid ViewTime and ViewAngle values. No additional absolute ViewAngle threshold was applied. The subsequent spatial and temporal matching between MODIS LST and simulated TSK is described in Section 2.4.

2.3. Model Configuration and Experimental Design

WRF-ARW v4.4.1 and WPS v4.4 were used to simulate the June 2022 extreme heat event over Henan Province. The simulations quantified temperature responses to differences in terrain and land-cover inputs on the model grid. Simulations ran from 00:00 UTC 9 June to 00:00 UTC 20 June; the first 72 h served as spin-up and were discarded. Valid hourly outputs from 12 to 19 June were used for subsequent analysis.

A two-way nested dual-domain scheme was applied: the outer domain (D01), with a grid spacing of 9 km and dimensions of 200 × 190 grid points, provided atmospheric forcing for the inner domain (D02), which had a grid spacing of 3 km and dimensions of 199 × 190 grid points. D02 fully covered Henan Province, and the nesting ratio was 3. All domains employed a Lambert conformal conic projection centered at 34.50°N, 112.50°E, with standard parallels at 30.0°N and 60.0°N. The outer domain adopted a 45 s time step, and the vertical configuration included 45 hybrid terrain-following layers extending to 50 hPa. The spatial layout of the nested simulation domains is shown in Figure 2. Minimum projected distances to the D02 boundary, defined by the outermost mass-grid centers, are 7.32 km (2.44 grid intervals) for the Henan polygon and 33.11 km (11.04 grid intervals) for the 111 evaluation stations. These geometric separations do not establish that lateral boundary effects are negligible.

All experiments shared identical physical parameterizations so that inter-experiment differences could be attributed to differences in the static terrain and land-cover inputs. The unified configuration included WSM3 microphysics, RRTMG radiation, Noah-MP land surface, YSU planetary boundary layer, and revised MM5 surface-layer schemes [30–34]. T2 is a diagnostic 2 m air temperature, not the temperature of the lowest model level. In this configuration, T2 is diagnosed using surface-layer similarity theory as part of the coupled surface-layer and land-surface calculations. The native WRF T2 output was used without reconstruction or empirical adjustment. The Kain–Fritsch cumulus scheme was enabled for the outer D01 domain and turned off in the convection-permitting D02 domain [35]. A fixed 21-category land-use setup without urban canopy parameterization was adopted. Vertical velocity damping (w_damping = 1) and an epssm value of 0.3 were applied to enhance numerical stability [36].

One control run (CTRL) and four sensitivity experiments (DEM, LU, LU + DEM, and C5m) were conducted to quantify the individual and combined effects of differences among the terrain and land-cover datasets (Table 1). All simulations used identical atmospheric forcing, physical schemes, and D01 static inputs; only the terrain and/or land-cover datasets in D02 were varied among experiments.

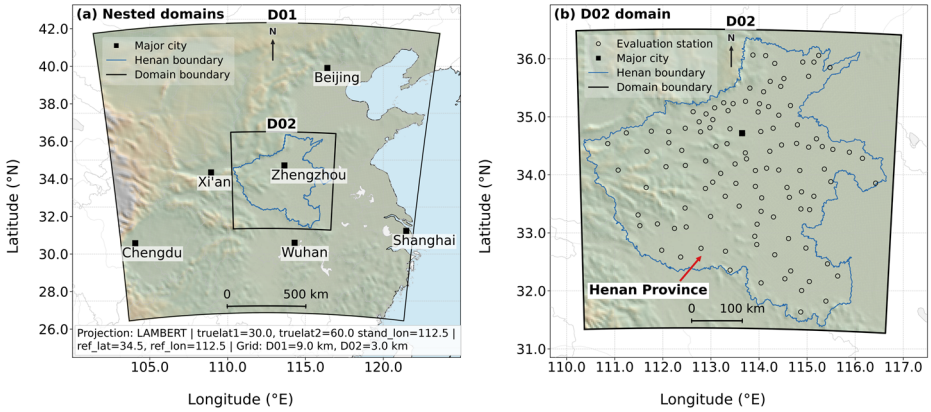


Figure 2. WRF domains and evaluation stations: (**a**) D01 and nested D02; (**b**) enlarged D02 and Henan Province. Open circles denote the 111 evaluation stations; black squares, major cities; blue lines, the Henan boundary; black outlines, domain boundaries; and thin gray lines, other administrative boundaries. Each panel has its own scale bar. Terrain shading is from the CTRL WPS geogrid output. Physical basemap layers are from Natural Earth (https://www.naturalearthdata.com/ (accessed on 1 October 2026)). Administrative boundaries were derived from the standard map provided by the Ministry of Natural Resources of China (map approval No. GS(2019)1822; http://bzdt.ch.mnr.gov.cn/ (accessed on 1 October 2026)).

Table 1. Experimental design of the static-data configurations.

Case	D01 Terrain and Land-Use Data	D02 Terrain and Land-Use Data	Purpose
CTRL	GMTED30s + MODIS30s-lake	GMTED30s + MODIS15s-lake	Baseline simulation using WPS model-built-in static geographic datasets
DEM	Same as CTRL D01	ASTER30 + MODIS15s-lake	Evaluates the impact of terrain-data replacement
LU	Same as CTRL D01	GMTED30s + MCD12Q1	Evaluates the impact of land-use-data replacement
LU + DEM	Same as CTRL D01	ASTER30 + MCD12Q1	Evaluates the combined effects of terrain- and land-use-data replacement
C5m	Same as CTRL D01	GMTED5m + MODIS5m-lake	Evaluates the integrated effect of a coarser WPS model-built-in static geographic configuration in D02

Note: ASTER30: terrain from ASTER GDEM V3; MCD12Q1: land use from 2019 MCD12Q1 V061; C5m: "5 m" denotes 5 arc-minutes. All other settings are identical across experiments except D02 terrain/land use. Land-cover replacement includes associated changes in the land–water mask and available initialization surface fields; these effects are not independently isolated.

CTRL served as the baseline using model-built-in WPS static datasets. The DEM experiment replaced the baseline terrain dataset with high-resolution ASTER30 while retaining MODIS15s-lake land cover, thereby isolating the effects of terrain-data replacement. The LU experiment retained GMTED30s terrain but replaced MODIS15s-lake with

the customized MCD12Q1 dataset to examine the effects of land-use-data replacement. The LU + DEM experiment simultaneously replaced both terrain and land-cover datasets with ASTER30 and MCD12Q1, respectively, to examine their combined effects. The C5m experiment adopted the coarser model-built-in GMTED5m terrain and MODIS5m-lake land-cover datasets in D02 to evaluate the integrated effects of a coarser static geographic configuration. Thus, C5m tests joint terrain–land-cover replacement rather than a DEM-only resolution change.

2.4. Evaluation and Diagnostic Methods

Two independent observational datasets, station air-temperature observations and MODIS land-surface temperature (LST), were used to evaluate simulated thermal conditions under different static-data configurations. Simulated 2 m air temperature (T2) was evaluated against station air-temperature observations, whereas simulated skin temperature (TSK) was evaluated against MODIS LST. Common statistical metrics, including bias, mean absolute error (MAE), root-mean-square error (RMSE), Pearson correlation coefficient (R), and valid-sample count (N), were applied to both evaluations.

For the MODIS evaluation, a pixel-level spatiotemporal matching framework was adopted based on quality-controlled MODIS LST pixels. Satellite local solar time was converted to UTC to provide a consistent temporal reference:

$$t_{\mathrm{UTC}} = t_{\mathrm{local}} - \frac{\lambda}{15} \tag{1}$$

where t_{local} is the local solar time (h), and λ is the longitude of the pixel center expressed in degrees east ($^{\circ}E$). Each valid MODIS pixel was spatially paired with the nearest D02 grid cell within a 3 km threshold. For temporal matching, linear interpolation was performed between adjacent hourly model outputs to match the actual satellite overpass time:

$$\mathrm{TSK}(t) = (1 - w)\mathrm{TSK}(t_0) + w\mathrm{TSK}(t_1), \tag{2}$$

where

$$w = \frac{t - t_0}{t_1 - t_0}. \tag{3}$$

Here, $\mathrm{TSK}(t_0)$ and $\mathrm{TSK}(t_1)$ denote the surface skin temperatures of the corresponding D02 grid cell at the two adjacent model output times, and w is the linear interpolation weight. Multiple valid MODIS pixels assigned to each D02 grid cell were averaged to obtain one grid-level paired sample. The baseline (0%) retained records with at least one common valid pixel across all five experiments within the Henan D02-center mask. Within-cell coverage was calculated as the number of common valid pixels divided by the total number of candidate pixel positions assigned to that cell within Henan. The denominator included positions rejected by quality control. For boundary cells, only positions within Henan were counted. This pixel-count proportion is not an exact area fraction, and the denominator was not fixed at nine. Sensitivity tests used minimum coverage thresholds of 0%, 25%, 50%, and 75% (Tables S4 and S5). MODIS LST is a satellite-retrieved radiometric surface temperature, whereas WRF TSK is the model surface skin temperature. Differences in physical representation and sampling may affect their comparability even after spatiotemporal matching.

For each date and MODIS overpass type, D02-cell spatial coverage was calculated as the percentage of the 17,871 Henan D02 cells with valid paired records. These cells had centers within or on the Henan boundary. Coverage was based on cell counts rather than area (Table S6). Figures S1 and S2 show daily coverage and valid-date frequency by cell.

Terrain and land-cover representation were summarized using CTRL HGT_M and LU_INDEX (Table S7A–D); the ratios are descriptive.

For station-based evaluation, station coordinates were converted to the D02 grid projection, and valid observation–model pairs were matched by station ID and UTC time. Simulated T2 was bilinearly interpolated to station locations and compared directly with air temperature observed at 1.5 m, without height adjustment. The 0.5 m height mismatch introduces uncertainty that depends on the near-surface temperature gradient; its sign and magnitude cannot be estimated from the available single-level observations. Daily maximum and minimum air temperatures were taken directly from the operational daily extreme records, whereas simulated T2max and T2min were derived from complete 24 h hourly outputs for each meteorological day. For spatial visualization of the representative high-temperature day, cubic radial basis function interpolation was applied to station data and clipped to the Henan administrative boundary. All quantitative evaluations were based on the original paired samples, with spatial interpolation used only for visualization.

For hourly station evaluation, three sample subsets were defined: All, Day, and Night. All covers the full 192 h evaluation period from 00:00 UTC 12 June to 23:00 UTC 19 June 2022; Day was defined as 08:00 ≤ BJT < 20:00; and Night was defined as BJT ≥ 20:00 or BJT < 08:00. Tmax/Tmin evaluation for the representative high-temperature day followed the China Meteorological Administration (CMA) meteorological-day definition described in Section 2.1. For MODIS LST evaluation, Pooled, Day, and Night samples were used. Day and Night correspond to MOD11A1 daytime and nighttime overpass retrievals, respectively, whereas Pooled denotes the combined valid Day and Night samples. Thus, MODIS Pooled samples are not equivalent to the All sample used for hourly station evaluation.

The evaluation metrics were calculated as

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i), \tag{4}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|M_i - O_i|, \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2}, \tag{6}$$

$$R = \frac{\sum_{i=1}^{N}(M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{N}(M_i - \bar{M})^2 \sum_{i=1}^{N}(O_i - \bar{O})^2}}, \tag{7}$$

where M_i and O_i denote the simulated and corresponding reference values, respectively, and $\bar{M}$ and $\bar{O}$ are their sample means. For the station evaluation, N represents the number of valid station–time paired records. For the MODIS evaluation, N represents the number of valid date–overpass-type–D02-grid-cell paired records common to all five experiments after applying the Henan D02-center mask and, where applicable, the specified coverage threshold. The baseline comprised 183,964 records (88,415 Day and 95,549 Night). Bias was calculated as simulation minus observation or reference value; therefore, a positive temperature bias corresponds to a warm bias. Mean squared error was decomposed as MSE = Bias^2 + Var(e), where e is model minus observation and Var(e) is the population variance (Table S15). This statistical decomposition does not imply physical attribution.

Cluster-based nonparametric bootstrap resampling with 20,000 replicates and a fixed random seed of 20220616 was used to quantify uncertainty in the TSK–MODIS LST statistics. Resampling was clustered by observation date, with all valid records for each selected date resampled together. The 95% confidence interval (CI) was determined from the 2.5th and 97.5th percentiles of the bootstrap distribution [37,38]. Because only eight date clusters were available, the resulting CI represents finite-sample uncertainty rather than a formal significance test [39]. These approximate CIs are conditional on the available paired samples and resampling assumptions and do not account for selection by cloud screening, quality filtering, or unobserved locations or times. The date-clustered bootstrap does not explicitly preserve adjacent-date dependence, so positive cross-date error correlation may lead to underestimated uncertainty.

To standardize internal process diagnosis, differences between each sensitivity experiment and the control run were defined as

$$\Delta X = X_{\mathrm{EXP}} - X_{\mathrm{CTRL}}, \tag{8}$$

where X_{EXP} and X_{CTRL} denote the corresponding variables in the sensitivity experiment and control run, respectively. The diagnostic variables include terrain height (HGT_M), surface pressure (PSFC), friction velocity (UST), sensible heat flux (HFX), skin temperature (TSK), and 2 m air temperature (T2). Model-response statistics were calculated over the 17,871 D02 grid cells whose centers fell within or on the Henan administrative boundary; the same spatial mask was applied to the grid-level MODIS evaluation. The separate geogrid-to-initialization and cross-experiment field audits in Tables S17 and S18 cover the full D02 mass grid. For each experiment, hourly grid differences were temporally averaged, and spatial statistics including the regional mean difference (ΔMean), mean absolute difference (Mean|Δ|), and root-mean-square difference (RMS) were calculated. The RMS values characterize the spatial heterogeneity of the time-averaged physical fields rather than integrated spatiotemporal errors. Internal Day and Night diagnostics were defined as 08:00–20:00 and 20:00–08:00 BJT, respectively.

For $\boldsymbol{X}$ = TSK or T2, the interaction residual and two equally weighted RMS measures were defined as

$$I_X = X_{\mathrm{LU+DEM}} - X_{\mathrm{LU}} - X_{\mathrm{DEM}} + X_{\mathrm{CTRL}}, \tag{9}$$

$$\mathrm{RMS}_{\mathrm{mean}}(Y) = \left[\frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{T}\sum_{t=1}^{T} Y_{it}\right)^2\right]^{\frac{1}{2}}, \tag{10}$$

$$\mathrm{RMS}_{\mathrm{st}}(Y) = \left[\frac{1}{MT}\sum_{i=1}^{M}\sum_{t=1}^{T} Y_{it}^2\right]^{\frac{1}{2}}, \tag{11}$$

where Y denotes either I_X or an experiment-minus-CTRL response, M = 17,871 is the number of Henan D02 grid cells, and T is the number of hours in the selected period. $\mathrm{RMS}_{\mathrm{mean}}$ is the spatial RMS of the time-mean field, while $\mathrm{RMS}_{\mathrm{st}}$ is the RMS of the hourly field over space and time. Diagnostics covered 12 June 00:00–19 June 23:00 UTC (192 h), with Day defined as 08:00 ≤ BJT < 20:00 and Night as the remaining hours (96 h each). C5m was excluded from the interaction calculation. Daily diagnostics and leave-one-UTC-date-out $\mathrm{RMS}_{\mathrm{st}}$ checks assessed sensitivity to individual dates (Tables S12 and S13), rather than statistical significance. Hourly differences from CTRL in TSK, T2, HFX, UST, and PBLH were summarized over the same cells and period using $\mathrm{RMS}_{\mathrm{mean}}$ and $\mathrm{RMS}_{\mathrm{st}}$ for the full period and Day/Night subsets. Beijing-time-hour composites were obtained by averaging

the eight fields corresponding to each hour of the day (Figures S6 and S7; Table S11). These diagnostics are descriptive and do not isolate individual process contributions.

Terrain-induced pressure responses were interpreted using the linearized hydrostatic pressure–height relationship:

$$\frac{\Delta p_{\mathrm{s}}}{p_{\mathrm{s}}} \approx -\frac{g\Delta h}{R_{\mathrm{d}}\overline{T_{\mathrm{v}}}}, \tag{12}$$

where p_{s} denotes the CTRL surface pressure used as the reference state, Δp_{s} represents the surface-pressure perturbation between sensitivity experiments and CTRL, Δh is the corresponding terrain-height perturbation derived from ΔHGT_M, $\overline{T_{\mathrm{v}}}$ is the mean near-surface virtual temperature, g is gravitational acceleration, and R_{d} is the specific gas constant for dry air [40]. This relationship was used to assess whether surface-pressure responses were physically consistent with terrain-height perturbations. It does not establish a one-to-one causal relationship.

The response of near-surface momentum exchange to land-use perturbations was interpreted using Monin–Obukhov similarity theory:

$$U(z) = \frac{u_*}{\kappa}\left[\ln\left(\frac{z}{z_{0\mathrm{m}}}\right) - \psi_{\mathrm{m}}\left(\frac{z}{L}\right) + \psi_{\mathrm{m}}\left(\frac{z_{0\mathrm{m}}}{L}\right)\right], \tag{13}$$

where $U(z)$ is the mean wind speed at height z, u_* is the friction velocity, κ is the von Kármán constant, $z_{0\mathrm{m}}$ is the momentum roughness length, L is the Monin–Obukhov length, and ψ_{m} is the stability correction function for momentum [41]. This relation provides the physical basis for diagnosing how land-use-related changes in surface roughness and atmospheric stability modify surface momentum exchange and UST.

Surface thermal responses were further interpreted using a simplified surface energy-balance framework:

$$R_{\mathrm{n}} = H + LE + G + \varepsilon, \tag{14}$$

where R_{n} is net surface radiation, H is sensible heat flux represented by HFX in this study, LE is latent heat flux, and G is soil heat flux. For conceptual purposes, ε collectively represents storage, phase-change, and numerical-closure terms that are not explicitly expanded in this simplified formulation. The land-surface energy exchange processes represented in Noah-MP follow the coupled physical formulations described in the Noah-MP technical documentation [42]. Therefore, land-surface parameters associated with different LU_INDEX categories can modify the partitioning of available surface energy among sensible, latent, and soil heat fluxes. HFX was used to assess the physical consistency between land-use perturbations and changes in sensible heat exchange. A complete surface energy-budget closure analysis was not performed, so the contributions of individual energy-budget terms could not be fully quantified.

3. Results

3.1. Static Geographic-Field Differences

Remote-sensing terrain and land-cover datasets produced distinct static geographic fields after WPS resampling to the 3 km D02 grid. Grid-scale terrain height (HGT_M) and dominant land-use category (LU_INDEX) were compared over 17,871 valid grid cells within Henan Province, using GMTED30s and MODIS15s-lake as the respective baselines.

Figure 3 shows the terrain-height distributions and corresponding differences among GMTED30s, ASTER30, and GMTED5m. All three datasets reproduce the major topographic pattern of Henan, while differences among terrain sources and native resolutions produce localized grid-scale deviations concentrated mainly in mountainous and transition zones. Relative to GMTED30s, ASTER30 produced comparatively small terrain-

height differences, ranging from −120.2 to 108.7 m. The mean difference, MAE, and RMSD were −3.1, 7.5, and 12.1 m, respectively. Grid cells with absolute differences exceeding 10, 30, and 50 m accounted for 20.7%, 3.7%, and 0.9%, respectively. These differences were concentrated mainly in western and northwestern complex-terrain regions, whereas plain areas were only weakly affected.

In contrast, GMTED5m showed much larger localized terrain differences, with a mean difference of 0.6 m, MAE of 22.8 m, RMSD of 47.6 m, and a range from −434.0 to 338.4 m. Grid cells with absolute differences exceeding 10, 30, and 50 m accounted for 38.0%, 23.1%, and 15.1%, respectively. Thus, the small regional mean difference in GMTED5m masks substantial local deviations, particularly in mountainous and transition regions. These statistics describe differences in WPS-generated HGT_M relative to GMTED30s on the 3 km D02 grid over Henan, not the absolute accuracy of the original DEM products.

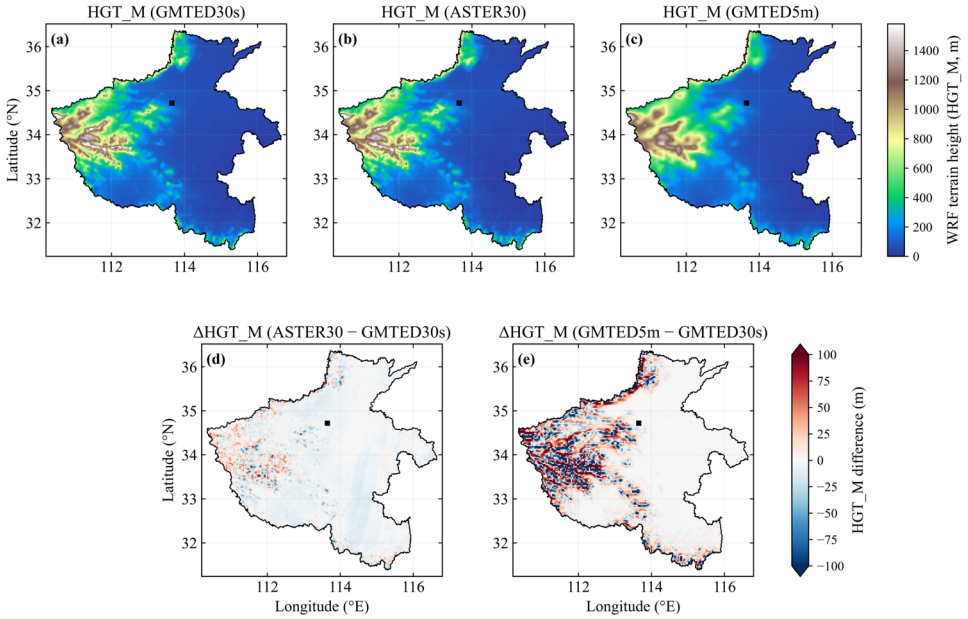


Figure 3. D02 terrain height (HGT_M) over Henan Province: (**a**) GMTED30s; (**b**) ASTER30; (**c**) GMTED5m; (**d**) ASTER30 − GMTED30s; and (**e**) GMTED5m − GMTED30s. Black squares mark Zhengzhou.

Land-cover differences were more spatially discrete and strongly associated with category transitions (Figure 4). Relative to MODIS15s-lake, MCD12Q1 changed LU_INDEX in 22.2% of valid grid cells, compared with 9.4% for MODIS5m-lake. MCD12Q1 reduced the cropland fraction from 85.4% to 75.0% and increased the urban fraction from 1.2% to 4.2%. By comparison, MODIS5m-lake increased the cropland fraction to 90.4% and reduced mixed-forest coverage to 7.7%.

Spatially, the MCD12Q1-related changes were more extensive across mountainous, transition, and urban areas, whereas MODIS5m-lake differences were more locally concentrated. These differences reflect the combined effects of dataset source, representative period, native spatial resolution, and classification scheme. They describe differences in grid-scale land-use representation rather than absolute classification errors.

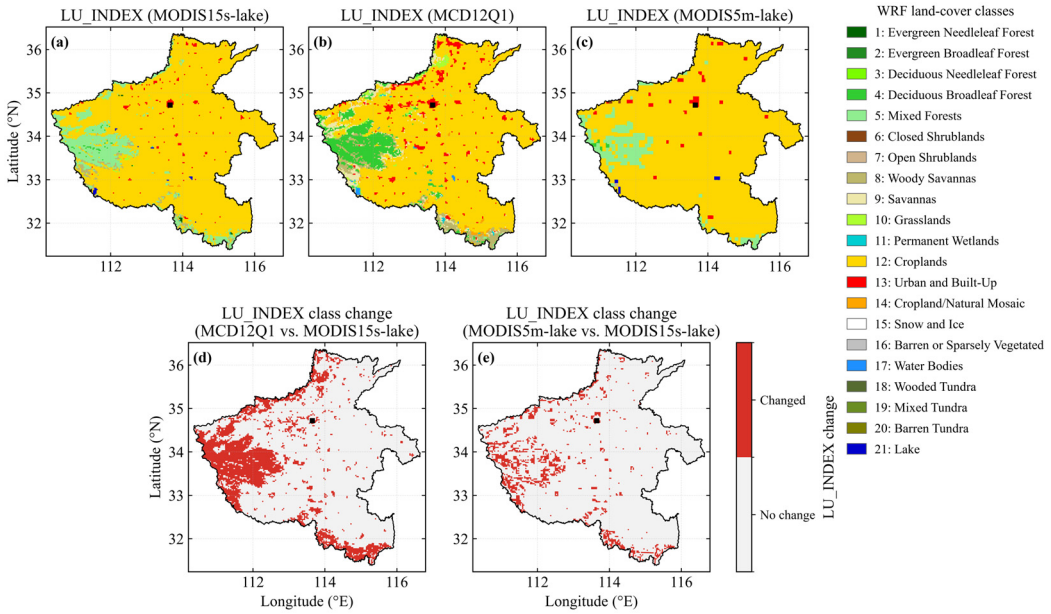


Figure 4. D02 LU_INDEX over Henan Province derived from (**a**) MODIS15s-lake, (**b**) MCD12Q1, and (**c**) MODIS5m-lake, and corresponding land-cover class changes relative to MODIS15s-lake for (**d**) MCD12Q1 and (**e**) MODIS5m-lake. Black squares mark Zhengzhou.

3.2. Temperature Evaluation Against MODIS and Station Observations

MODIS LST and station air-temperature observations provide complementary evaluations of simulated surface and near-surface thermal conditions. Simulated TSK was evaluated against time-matched MODIS LST, whereas simulated T2 was directly compared with 1.5 m station air-temperature observations following the procedures described in Section 2.4.

Common paired samples totaled 88,415 for Day and 95,549 for Night. Mean daily D02-cell spatial coverage was 61.84% for Day (range: 0.04–98.19%) and 66.83% for Night (range: 31.35–95.55%). Day coverage was particularly low on 12 and 19 June, at 0.04% and 5.05%, respectively (Table S6). At least one sample was available over the full period for 99.87% of cells by Day and 99.90% by Night, despite incomplete daily coverage (Figures S1 and S2). Descriptive sample composition by elevation and land cover is summarized in Table S7A–D; corresponding stratified errors are reported in Tables S19 and S20.

All five configurations exhibited a TSK warm bias relative to MODIS LST during 12–19 June 2022 on the common Henan-masked sample (Figure 5; Table S4). Pooled bias, MAE, and RMSE ranged from 5.520 to 5.742 K, 5.850 to 6.020 K, and 6.758 to 6.896 K, respectively. C5m reduced these metrics relative to CTRL by approximately 0.222, 0.157, and 0.129 K. Daytime bias, MAE, and RMSE ranged from 5.200 to 5.596 K, 5.952 to 6.122 K, and 7.221 to 7.382 K, with LU producing the smallest errors. Nighttime ranges were 5.719–5.965 K, 5.735–5.984 K, and 6.255–6.464 K, respectively, with C5m producing the smallest errors. Daytime R ranged from 0.498 to 0.542, nighttime R from 0.772 to 0.777, and pooled R from 0.889 to 0.899. The larger pooled correlations reflect the broader temperature range obtained by combining daytime and nighttime samples, and do not imply equally strong correlations within the separate daytime and nighttime subsets. The modest error differences relative to the shared warm bias do not establish universal configuration superiority.

At the 75% coverage threshold, 91.70% of baseline records were retained. From the 0% to 75% threshold, the maximum absolute changes were 0.0906 K for bias, 0.0802 K for MAE, 0.0645 K for RMSE, and 0.00550 for R. Pooled and nighttime MAE/RMSE rankings were unchanged, while some closely ranked daytime configurations changed order. Full

results and daily sample counts are provided in Tables S4 and S5; these comparisons are based on point estimates.

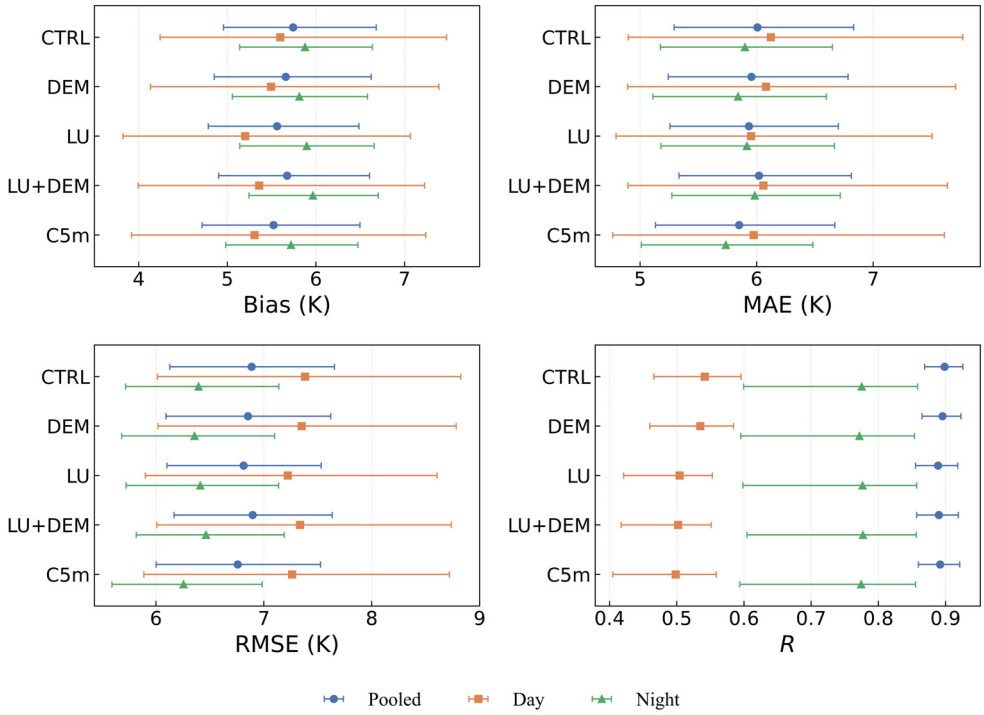


Figure 5. Evaluation of model-simulated TSK against MODIS LST for the five static-data experiments for 12–19 June 2022, based on the common Henan-masked sample (183,964 Pooled records; 88,415 Day and 95,549 Night). Error bars denote the 95% date-clustered bootstrap confidence intervals.

Daily errors and changes from CTRL are summarized in Table S8A–C and Figure S3. LU had lower daytime MAE and RMSE than CTRL on all eight dates. C5m had lower pooled MAE and RMSE on all eight dates, lower nighttime MAE on all eight dates, and lower nighttime RMSE on seven dates. In the leave-one-date-out analysis (Table S9A–C), C5m ranked first for pooled and nighttime MAE and RMSE in all eight runs, whereas the daytime ranking changed when 15 June was omitted. These are descriptive sensitivity results; they do not establish statistical significance or remove cross-day dependence.

The spatial distribution of TSK–MODIS LST differences on 16 June further demonstrates the strong diurnal contrast (Figure 6). Daytime warm bias was pronounced over northern and north-central Henan, whereas parts of southern and southeastern Henan showed weak or locally negative differences. Nighttime differences were more spatially homogeneous and were dominated by warm bias across most of the province. All five configurations retained similar large-scale spatial patterns, although local bias magnitude and patch boundaries varied among experiments. Thus, the static-data replacements primarily modified local bias intensity rather than reversing the regional pattern of TSK–LST differences.

Station observations likewise indicate a persistent warm tendency in simulated near-surface air temperature. Over the full 192 h evaluation period, bias ranged from 2.36 to 2.76 °C, MAE from 2.88 to 3.16 °C, and RMSE from 3.43 to 3.68 °C. Correlations between simulated T2 and observed air temperature were approximately 0.91 for all configurations (Figure 7 and Table 2). C5m produced the smallest full-period bias, MAE, and RMSE, with values of 2.36, 2.88, and 3.43 °C, respectively. Nevertheless, the differences among configurations remained modest relative to the warm bias shared by all experiments.

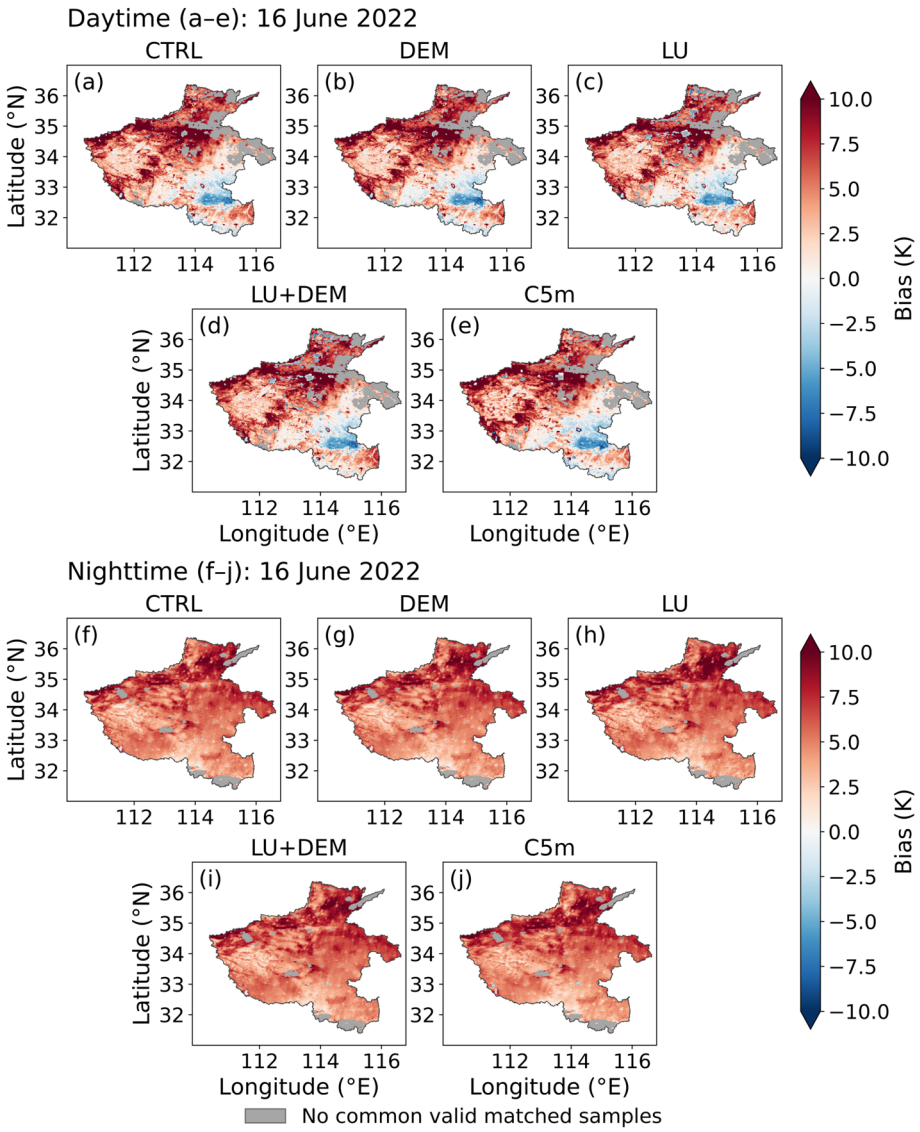


Figure 6. Spatial distribution of TSK bias (model TSK − MODIS LST) during MODIS daytime and nighttime overpasses on 16 June 2022. Panels (**a–e**) show daytime overpasses, and panels (**f–j**) show nighttime overpasses. TSK was interpolated to the MODIS acquisition times, and valid MODIS pixels were averaged within their matched D02 cells. Only grid cells whose centers fell within or on the Henan boundary were retained. Gray denotes cells within this evaluation mask without common valid samples; near-white colors represent near-zero bias. Areas outside the evaluation mask are left blank.

Table 2. Summary of the lowest temperature-error metrics and highest correlation coefficients for station-based air-temperature evaluation during the All, Day, and Night periods.

Period	Bias (°C)	MAE (°C)	RMSE (°C)	*R*
All	C5m (2.36)	C5m (2.88)	C5m (3.43)	All configurations (~0.91)
Day	C5m (2.47)	C5m (3.09)	C5m (3.61)	CTRL (0.87)
Night	C5m (2.25)	C5m (2.66)	C5m (3.24)	C5m (0.86)

Note: Values in parentheses are the corresponding metric values. Best bias denotes the smallest absolute bias, with the sign retained; best MAE and RMSE denote the minimum values, and best *R* denotes the maximum value. For the All period, *R* was approximately 0.91 for all configurations at the reported precision. Day and Night are defined in Section 2.4. Rankings are descriptive and were not significance-tested.

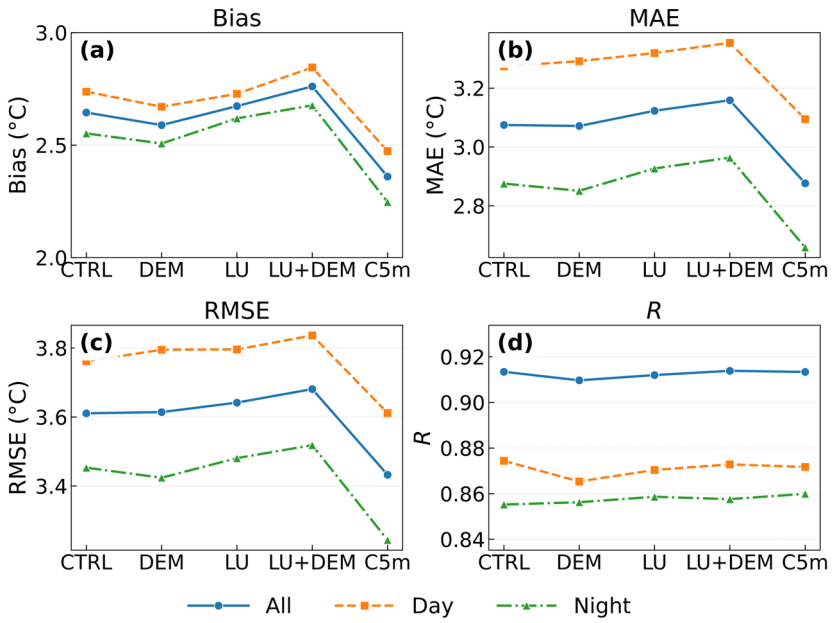


Figure 7. Performance statistics of simulated 2 m air temperature against station observations during the 192 h evaluation period (12–19 June 2022) for All, Day, and Night samples: (**a**) bias; (**b**) mean absolute error (MAE); (**c**) root-mean-square error (RMSE); and (**d**) Pearson correlation coefficient (*R*). Bias = model − observation.

The station-based temperature errors also exhibited diurnal differences (Figure 7 and Table 2). C5m produced the smallest absolute bias, MAE, and RMSE for both Day and Night samples. Its daytime bias, MAE, and RMSE were 2.47, 3.09, and 3.61 °C, respectively, whereas the corresponding nighttime values were 2.25, 2.66, and 3.24 °C. The highest daytime correlation was obtained by CTRL (R = 0.87), whereas C5m showed the highest nighttime correlation (R = 0.86). For the All period, the reported correlations were approximately 0.91 for all configurations. C5m yielded the lowest station-based T2 absolute bias, MAE, and RMSE for All, Day, and Night, while the highest correlation depended on the sampling period (Table 2). For the baseline MODIS comparison, LU had the lowest daytime TSK errors and C5m the lowest pooled and nighttime errors (Figure 5; Table S4). These descriptive rankings indicate trade-offs across variables, metrics, and sampling periods rather than a universally optimal configuration; no statistical significance is inferred. Representative hourly station comparisons show a general warm bias, with cold biases at some stations and times (Figures S8 and S9; Table S14). Daily extreme temperatures on the representative high-temperature day provide an additional spatial evaluation (Figure 8). All 111 stations recorded Tmax above 35 °C on 16 June, including 98 stations above 38 °C and 56 stations above 40 °C. Observed Tmax ranged from 35.3 to 42.0 °C, with a mean of 39.62 °C, and the highest temperatures were concentrated in central, northern, and eastern Henan. Simulated T2max means ranged from 42.66 to 42.92 °C, indicating an overall overestimation of daytime extreme temperature, while spatial correlations remained moderate (R = 0.68–0.69).

Observed Tmin ranged from 18.4 to 27.6 °C, with a mean of 23.37 °C. Simulated T2min means were higher, ranging from 25.84 to 26.39 °C, and their spatial correlations were substantially lower (R = 0.27–0.33). Compared with Tmax, the simulated Tmin fields were spatially smoother and reproduced localized nighttime cool areas less effectively. The contrast between Tmax and Tmin indicates that the broad daytime thermal gradient was reproduced more successfully than the fine-scale nighttime temperature structure.

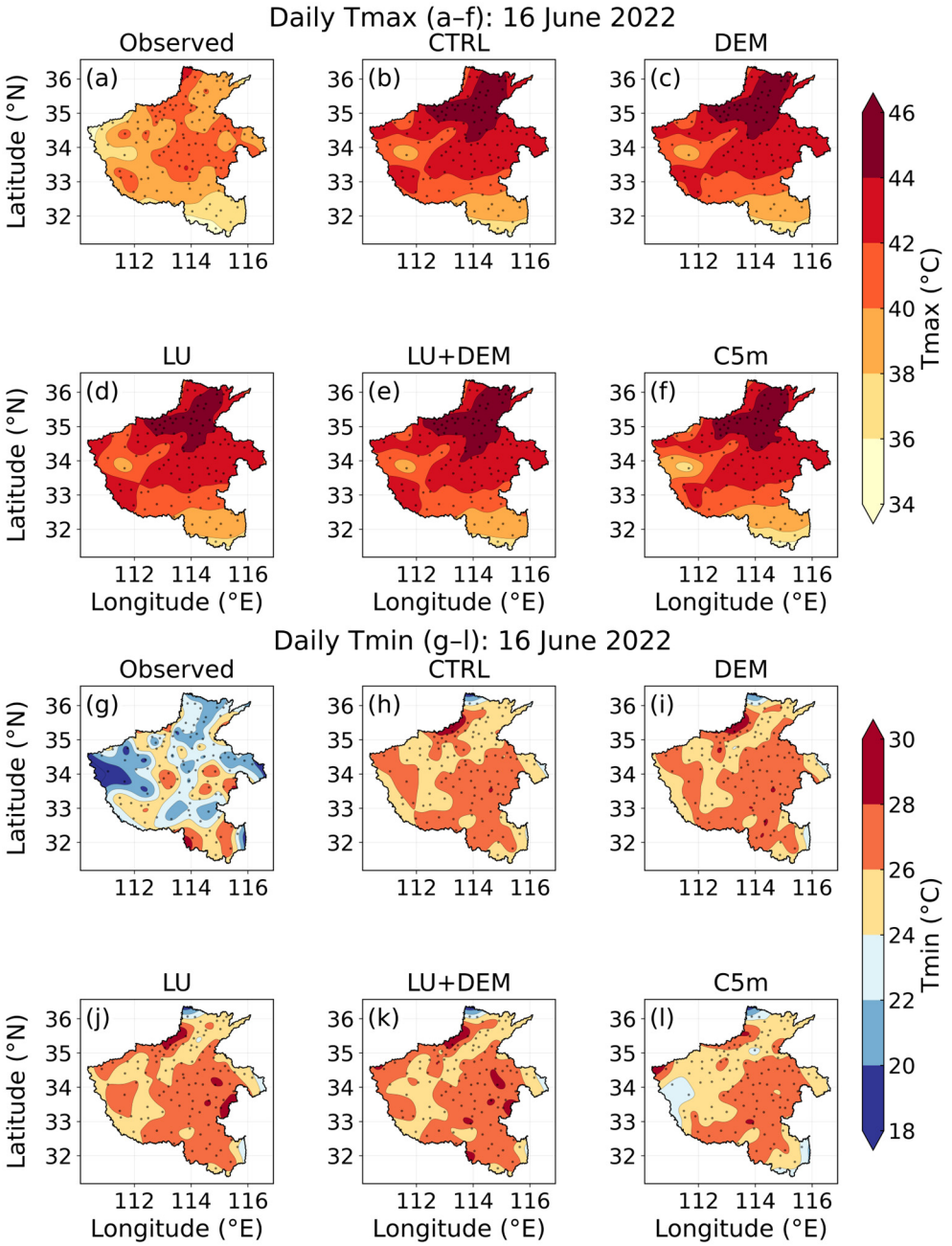


Figure 8. Observed and simulated daily Tmax (**a**–**f**) and Tmin (**g**–**l**) on 16 June 2022. Panels (**a**,**g**), (**b**,**h**), (**c**,**i**), (**d**,**j**), (**e**,**k**), and (**f**,**l**) show observations, CTRL, DEM, LU, LU+DEM, and C5m, respectively. Simulated daily extremes were derived from 24 hourly T2 outputs within the same meteorological day. Dots denote stations; cubic RBF interpolation is used for visualization only.

3.3. Physical Responses to Terrain and Land-Use Perturbations

The internal diagnostics show distinct responses of near-surface dynamical, surface-exchange, and thermal fields to terrain and land-use perturbations. All differences are defined relative to CTRL according to Section 2.4.

Terrain-related experiments produced clear surface-pressure responses. Figure 9a shows the relationship between ΔHGT_M and the 192 h mean ΔPSFC for the DEM experiment, while Figure 10 summarizes the regional PSFC responses across the sensitivity experiments. Spatial differences in HGT_M and PSFC were strongly inversely related, with correlation coefficients of −0.9990 to −0.9993 and more than 92% of affected grid cells

showing opposite signs. The LU experiment retained the CTRL terrain and showed negligible PSFC perturbations. Experiments with modified terrain exhibited substantially larger local pressure responses, with the greatest spatial variability in C5m.

Land-use-related experiments produced more pronounced spatial responses in surface momentum and sensible heat exchange than terrain replacement alone. The UST responses associated with major land-use transitions are shown in Figure 9b–d, while Figure 11a–c summarize their regional All/Day/Night statistics. Regional-mean UST changes were small, but their spatial variability was considerably greater in the experiments involving land-use replacement, particularly during daytime.

Sensible heat flux showed an even stronger diurnal contrast. Figure 9e–g show HFX responses associated with major land-use transitions, while Figure 11d–f summarize the regional statistics. Land-use-related experiments exhibited daytime HFX RMS values exceeding 20 W/m^2, whereas nighttime responses were much weaker, indicating pronounced daytime spatial heterogeneity in surface sensible heat exchange.

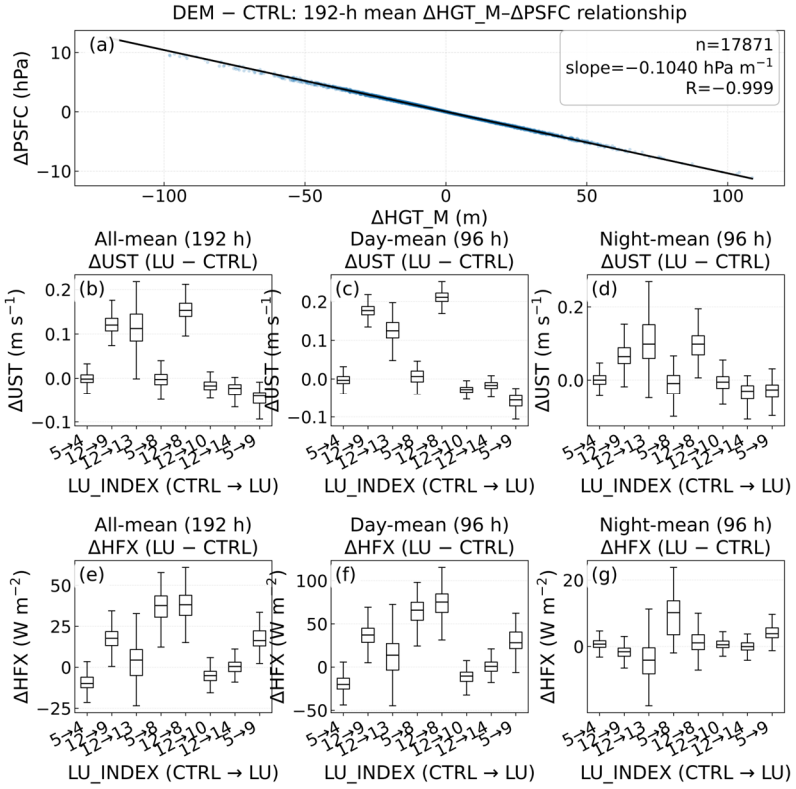


Figure 9. Model responses to static geographic-data perturbations. Panel (**a**) shows ΔHGT_M versus the 192 h mean ΔPSFC for DEM relative to CTRL. Panels (**b**–**d**) show ΔUST and panels (**e**–**g**) show ΔHFX for major land-use transitions from CTRL to LU, with All, Day, and Night periods shown in that order within each group. Differences are calculated as DEM − CTRL in (**a**) and LU − CTRL in (**b**–**g**). Day is defined as 08:00 ≤ BJT < 20:00, and Night comprises the remaining hours.

The dynamical and surface-exchange responses were accompanied by changes in TSK and T2. Figure 12 shows the spatial patterns of the 192 h mean ΔTSK and ΔT2, and Figure 13 summarizes their regional All/Day/Night statistics. Both figures indicate stronger and more spatially heterogeneous responses in TSK than in T2.

For TSK, the full-period ΔMean values for DEM, LU, LU + DEM, and C5m were −0.083, −0.110, approximately 0, and −0.199 °C, respectively, with corresponding RMS values of 0.289, 0.846, 0.838, and 0.984 °C (Figure 13a–c). The near-zero regional mean for LU + DEM, together with its relatively large RMS, indicates strong spatial cancelation between localized warming and cooling rather than an absence of thermal response. This

spatial compensation is also evident in the alternating positive and negative ΔTSK patterns in Figure 12a–d.

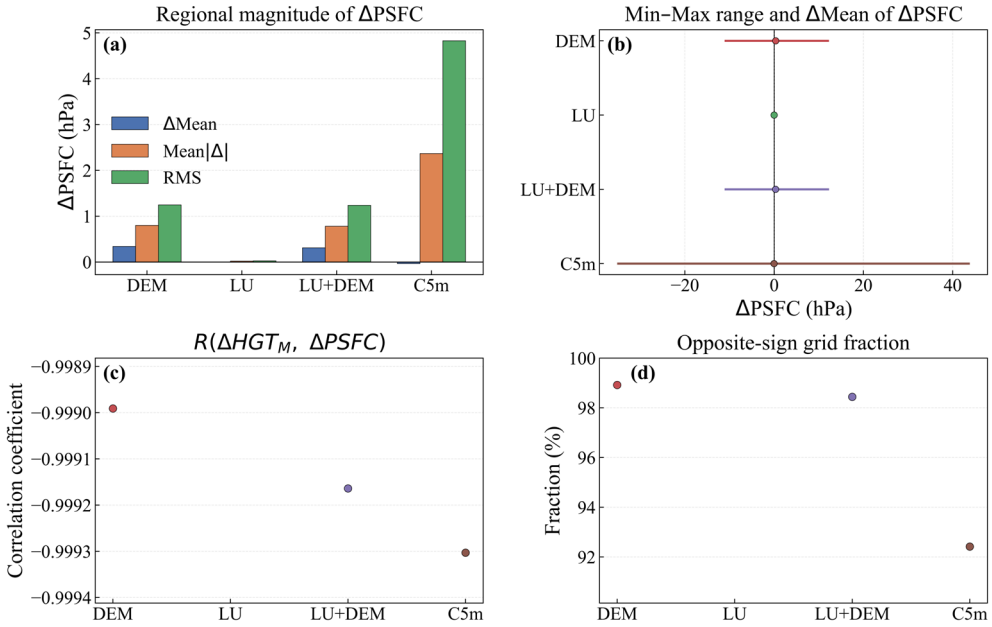


Figure 10. Regional ΔPSFC responses to static geographic-data perturbations over Henan during the 192 h period: (**a**) ΔMean, Mean|Δ|, and RMS; (**b**) minimum–maximum range (horizontal lines) and ΔMean (dots); (**c**) spatial correlation between ΔHGT_M and ΔPSFC; and (**d**) fraction of grid cells where ΔHGT_M and ΔPSFC have opposite signs. Differences are calculated as EXP − CTRL; the correlation and opposite-sign statistics are omitted for LU because HGT_M is unchanged.

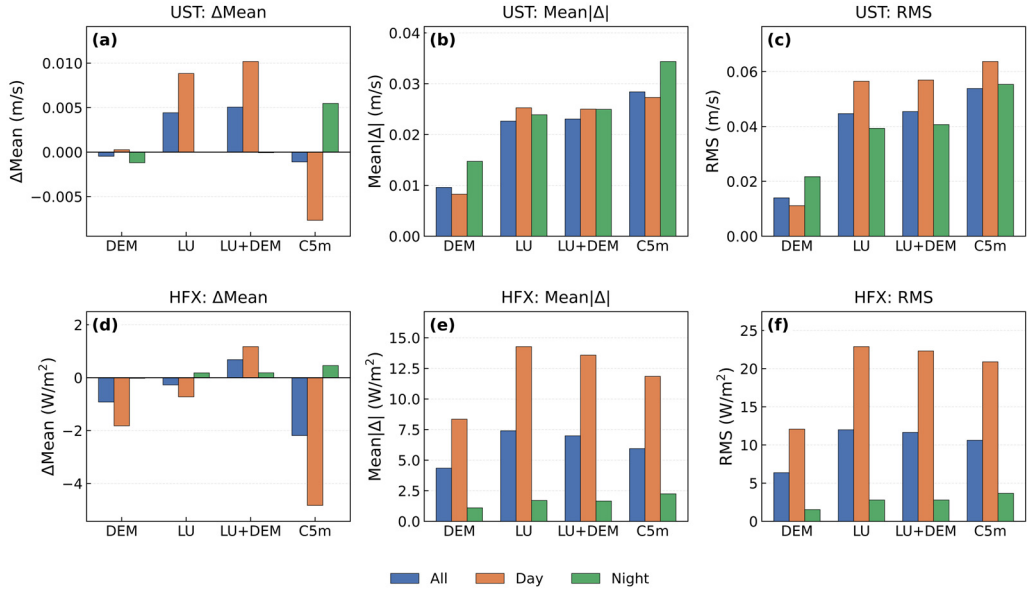


Figure 11. Regional UST and HFX responses to static geographic-data perturbations during All, Day, and Night periods. Panels (**a–c**) show ΔMean, Mean|Δ|, and RMS of UST; (**d–f**) show the corresponding HFX statistics. Differences are calculated as EXP − CTRL.

Daytime TSK perturbations were consistently stronger than nighttime perturbations (Figure 13a–c): daytime RMS values were 0.432, 1.597, 1.581, and 1.687 °C for DEM, LU, LU + DEM, and C5m, respectively, compared with nighttime values of 0.191, 0.375, 0.384,

and 0.530 °C. The stronger daytime spatial responses were particularly evident in experiments involving land-use replacement.

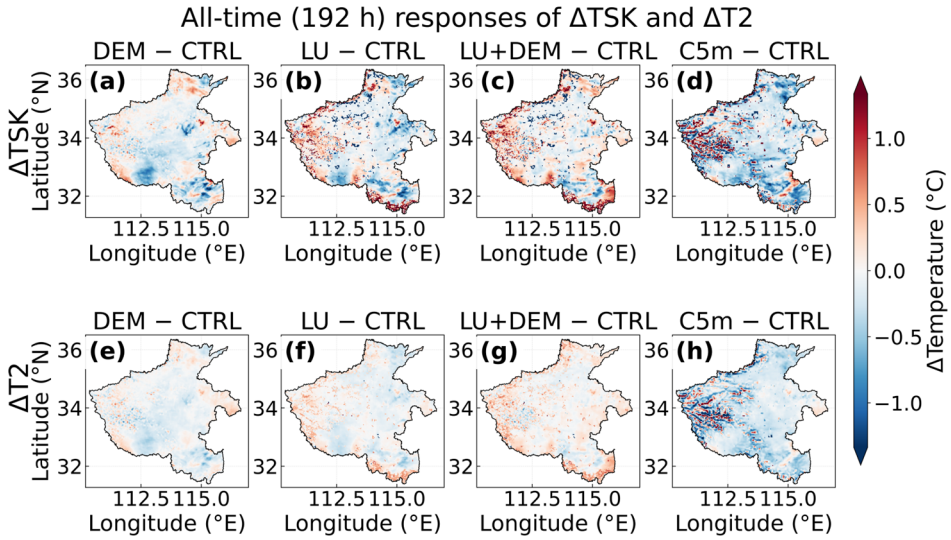


Figure 12. Spatial patterns of 192 h mean ΔTSK (**a**–**d**) and ΔT2 (**e**–**h**) for DEM, LU, LU + DEM, and C5m relative to CTRL. Δ = EXP − CTRL; red and blue indicate warming and cooling, respectively.

The corresponding T2 responses were generally weaker and spatially smoother than those of TSK. Full-period ΔMean T2 values for DEM, LU, LU + DEM, and C5m were −0.055, −0.013, +0.069, and −0.217 °C, respectively, with corresponding RMS values of 0.164, 0.234, 0.251, and 0.510 °C. All T2 RMS values were lower than the corresponding TSK values. For C5m, nighttime mean T2 cooling (−0.245 °C) was greater than daytime cooling (−0.189 °C), whereas daytime RMS was larger than nighttime RMS (0.595 vs. 0.489 °C).

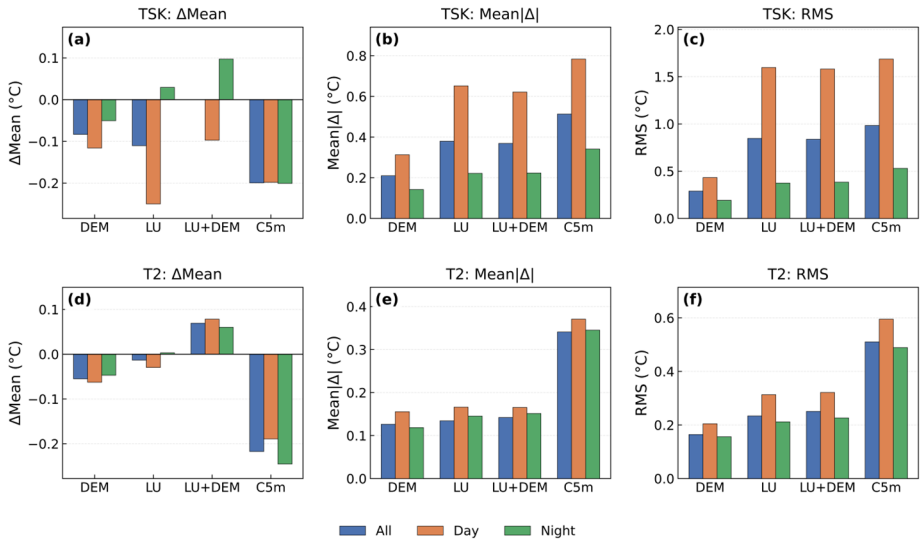


Figure 13. Regional TSK and T2 responses to static geographic-data perturbations during All, Day, and Night periods. Panels (**a**–**c**) show ΔMean, Mean|Δ|, and RMS of TSK; panels (**d**–**f**) show the corresponding T2 statistics. Differences are calculated as EXP − CTRL.

Overall, terrain perturbations were associated with clear HGT_M–PSFC responses, whereas land-use perturbations produced stronger spatial variability in UST and HFX. These internal responses were accompanied by more pronounced changes in TSK than in

T2, with particularly strong daytime TSK variability in the experiments involving land-use replacement. Possible mechanisms linking these intermediate responses to simulated temperature changes are discussed in Section 4.

4. Diagnostic Analysis and Possible Mechanisms

While extreme heat simulation studies often emphasize quantitative error evaluation, fewer studies systematically examine how terrain and land-cover datasets from different remote-sensing sources change model surface conditions and simulated temperatures. Building on the evaluation and internal-response results in Section 3, this section organizes terrain-related and land-cover-related responses into two diagnostic pathways. The hydrostatic, surface-exchange, and energy-balance relationships described in Section 2.4 guide their physical interpretation. The framework connects static inputs and internal model diagnostics with temperature evaluation, without isolating each mediating process or establishing a complete causal chain.

4.1. Land-Surface Perturbations Induced by Remote-Sensing Data Grid Mapping

WPS grid resampling transforms raw remote-sensing datasets into grid-scale land-surface fields, whose differences constitute the initial perturbations among the model experiments. Remotely sensed DEM and land-cover products were mapped to the same 3 km D02 grid. Differences in their sources, representative periods, native resolutions, and classification characteristics resulted in distinct terrain-height (HGT_M) and land-use category (LU_INDEX) fields.

Relative to the baseline GMTED30s dataset, ASTER30 produced comparatively small terrain differences, with a mean difference of −3.1 m, MAE of 7.5 m, and RMSD of 12.1 m. In contrast, GMTED5m produced much larger localized differences, with an MAE of 22.8 m and RMSD of 47.6 m despite a small regional mean difference of 0.6 m. The larger deviations were concentrated mainly in mountainous and terrain-transition regions, indicating that regional means alone can mask substantial grid-scale terrain perturbations. These quantities describe relative differences among the model-grid terrain fields rather than the absolute accuracy of the original DEM products.

Land-cover datasets likewise exhibited evident grid-scale inconsistencies after WPS processing. Relative to MODIS15s-lake, MCD12Q1 changed LU_INDEX in 22.2% of valid grid cells, whereas MODIS5m-lake differed in 9.4%. MCD12Q1 reduced cropland coverage from 85.4% to 75.0% and increased the urban fraction from 1.2% to 4.2%. These differences reflect combined effects of dataset source, representative period, native spatial resolution, and classification scheme rather than a single resolution effect. After WPS mapping, differences in HGT_M and LU_INDEX define the lower-boundary perturbations used to examine terrain-related and land-cover-related model responses.

4.2. Terrain-Related Surface-Pressure Responses

Terrain-modified experiments showed a strong inverse spatial relationship between ΔHGT_M and ΔPSFC (R = −0.9990 to −0.9993), with opposite signs in more than 92% of affected grid cells. This relationship is consistent with the linearized hydrostatic pressure–height relationship described in Section 2.4, but does not by itself demonstrate circulation adjustments. PSFC is evaluated at each experiment's own surface elevation, so these differences are not pressure perturbations at a common height.

The LU experiment, which retained the CTRL terrain, showed a small PSFC response (RMS = 0.0260 hPa). Among the terrain-modified experiments, C5m had a PSFC RMS of 4.8273 hPa and a range of −35.171 to 43.828 hPa, with its near-zero regional mean reflecting compensation between positive and negative local differences.

ASTER30 produced smaller terrain differences and more modest PSFC differences than GMTED5m. These results describe the correspondence between terrain and surface-pressure differences but do not separate elevation-related effects from circulation adjustments.

Although the strong ΔHGT_M–ΔPSFC relationship supports hydrostatic consistency within the terrain-related diagnostic pathway, it does not establish the contribution of pressure differences or circulation changes to TSK and T2. PSFC is an internal model diagnostic rather than an independent observational evaluation target. Surface exchange, atmospheric stability, and turbulent mixing provide possible process interpretations, whose individual contributions were not isolated. C5m changes both terrain and land-cover inputs, so its temperature response cannot be attributed to terrain alone.

4.3. Land-Use Perturbations and Boundary-Layer Momentum and Thermal Exchange

In contrast to the terrain-related surface-pressure responses, land-cover perturbations are associated with changes in land-surface momentum and energy exchange. Grid-scale differences in LU_INDEX modify surface properties represented within the Noah-MP land-surface scheme, including vegetation, roughness, radiative characteristics, and related exchange parameters. These changes are reflected in friction velocity (UST) and sensible heat flux (HFX), with pronounced spatial and diurnal heterogeneity.

Land-use perturbations produced relatively small regional-mean UST changes but much stronger local spatial variability. Full-period UST RMS values for DEM, LU, LU + DEM, and C5m were 0.0140, 0.0447, 0.0454, and 0.0538 m/s, respectively. Thus, LU and LU + DEM produced roughly three times the UST spatial response of the terrain-only DEM experiment. LU and LU + DEM also showed daytime ΔMean UST increases of 0.0088 and 0.0102 m/s, whereas nighttime changes were much weaker.

This behavior is physically consistent with the Monin–Obukhov framework described in Section 2.4, in which friction velocity depends jointly on surface roughness, wind shear, and atmospheric stability. Consequently, LU_INDEX does not determine UST directly; instead, land-use-dependent surface parameters modify the boundary conditions under which near-surface momentum exchange occurs.

Sensible heat flux exhibited an even stronger diurnal response. Full-period regional-mean HFX differences ranged from −2.187 to +0.680 W/m², while nighttime mean differences remained below 0.5 W/m². Most importantly, land-use-related experiments showed daytime HFX RMS values exceeding 20 W/m², indicating pronounced local spatial heterogeneity in sensible heat exchange during daytime.

These HFX responses are consistent with the surface energy-balance framework described in Section 2.4. Different land-use categories can modify several components of the surface energy balance simultaneously through changes in vegetation, radiative properties, soil–vegetation processes, and aerodynamic coupling. HFX is therefore used to diagnose changes in sensible heat exchange associated with land-use perturbations. These diagnostics do not establish that HFX alone determines simulated temperature.

Land-cover replacement was associated with pronounced daytime HFX and TSK responses, supporting a land-cover-related diagnostic pathway relating WPS-mapped LU_INDEX differences to UST, HFX, TSK, and T2 responses. Surface momentum and energy exchange provide possible process interpretations, but the individual mediating contributions are not isolated.

4.4. Surface and Near-Surface Temperature Responses

Surface skin temperature (TSK) and 2 m air temperature (T2) are examined alongside the terrain-related and surface-exchange diagnostics to characterize model responses to static-data replacement. While the 192 h mean fields had larger spatial RMS responses for

TSK than for T2, this comparison alone does not demonstrate spatial smoothing or identify its cause.

For TSK, full-period RMS values for DEM, LU, LU + DEM, and C5m were 0.289, 0.846, 0.838, and 0.984 °C, respectively. The LU + DEM experiment is particularly informative: its regional-mean ΔTSK was nearly zero, whereas its RMS remained 0.838 °C. This combination demonstrates that the small regional mean resulted from strong spatial cancelation between localized warming and cooling rather than an absence of thermal response.

TSK also exhibited pronounced diurnal sensitivity. Daytime RMS values were consistently larger than nighttime values, especially in experiments involving land-use replacement: LU showed 1.597 versus 0.375 °C, LU + DEM showed 1.581 versus 0.384 °C, and C5m showed 1.687 versus 0.530 °C. This daytime amplification is consistent with the stronger daytime HFX heterogeneity and indicates enhanced surface thermal sensitivity during periods of active land–atmosphere exchange.

For the time-mean fields, T2 RMS_{mean} values were 0.164, 0.234, 0.251, and 0.510 K for DEM, LU, LU + DEM, and C5m, respectively, below the corresponding TSK values (Table S11). This indicates smaller T2 response magnitudes but does not by itself demonstrate attenuation or spatial smoothing.

The C5m experiment further illustrates the distinction between regional-mean and spatial responses. Its nighttime mean T2 cooling (−0.245 °C) was stronger than its daytime cooling (−0.189 °C), whereas its daytime RMS remained larger than the nighttime value (0.595 vs. 0.489 °C). Thus, the regional mean alone does not fully characterize the intensity of local temperature responses.

Table S11 shows lower T2 than TSK RMS_{st} responses for all three periods (All, Day, and Night). Figures S6 and S7 show their hourly variations and the corresponding HFX, UST, and PBLH responses. These results are consistent with possible modulation by near-surface exchange and boundary-layer processes, but do not quantify their separate contributions or establish causation. PBLH indicates boundary-layer height, not turbulent mixing intensity.

The results support two diagnostic pathways for organizing the responses to static-data perturbations: the terrain-related pathway links WPS-mapped HGT_M differences with hydrostatically consistent PSFC differences, without establishing circulation-mediated temperature effects, while the land-cover-related pathway relates LU_INDEX differences to diagnosed UST, HFX, TSK, and T2 responses, with physically consistent process interpretations but without isolating individual mediating contributions.

For the June 2022 Henan heatwave under the examined WRF configuration, direct interaction residuals indicate non-additive TSK and T2 responses. Full-period RMS_{mean} values were 0.421 K for TSK and 0.225 K for T2 (Figure S4; Table S10A). The spatial distributions of the temporal RMS of hourly interaction residuals are shown in Figure S5. Daily interaction magnitudes varied (Table S12), and the principal full-period aggregate RMS_{st} comparisons persisted in leave-one-UTC-date-out checks, although magnitudes changed (Table S13). These checks do not establish statistical significance or spatial-pattern stability.

The framework organizes model responses into terrain-related and land-cover-related diagnostic pathways, linking WPS-mapped HGT_M and LU_INDEX differences to responses in PSFC, UST, HFX, PBLH, TSK, and T2. Temperature evaluation against MODIS LST and station observations complements these internal diagnostics by distinguishing internal physical-consistency diagnostics from observation-based performance assessment. Near-surface and boundary-layer processes provide context, but their individual causal contributions are not isolated; improved temperature agreement alone does not validate the proposed mechanisms.

Importantly, larger internal model responses do not necessarily translate into improved agreement with observations. The MODIS and station evaluations in Section 3 show that differences among the static-data configurations are relatively modest compared with the systematic warm biases common to all experiments. Thus, static-data perturbations modify the diagnosed internal variables and local temperature fields, but they do not alone explain the common temperature biases. This distinction between internal physical sensitivity and observation-based temperature performance provides the basis for the discussion in Section 5.

5. Discussion

In this study, we examine terrain- and land-cover-related model responses and their possible process interpretations. For the examined event and model configuration, using static-data products with higher nominal native resolution did not produce consistent improvements in temperature simulation. This section discusses trade-offs in temperature simulation performance, methodological contributions, and study limitations.

5.1. Terrain-Related and Land-Cover-Related Responses

Terrain and surface-pressure differences were strongly correlated (R = −0.9990 to −0.9993), consistent with hydrostatic relationships. Terrain MAE was 7.5 m for ASTER30 and 22.8 m for GMTED5m, with GMTED5m PSFC differences ranging from −35.17 to 43.83 hPa. MCD12Q1 changed LU_INDEX in 22.2% of valid grids, expanded mapped urban coverage, and produced daytime time-mean HFX difference fields with spatial RMS values exceeding 20 W/m². These associations do not establish circulation-mediated temperature effects or identify PBL mixing as the cause of diurnal differences. Direct interaction residuals demonstrate non-additivity, with spatial RMS values of the 192 h mean fields of 0.421 K for TSK and 0.225 K for T2 (Figure S4; Table S10A). Daily diagnostics characterize temporal variability, while leave-one-date-out checks assess the sensitivity of aggregate hourly RMS magnitudes to individual dates (Tables S12 and S13). Neither establishes statistical significance, spatial-pattern stability, or a specific mechanism.

5.2. Counterintuitive Model Performance Trade-Offs

Temperature evaluation shows trade-offs across variables and sampling periods. C5m had the lowest station-based T2 absolute bias, MAE, and RMSE. For TSK, LU had the lowest daytime errors, whereas C5m had the lowest pooled and nighttime errors. Pooled TSK warm biases remained 5.52–5.74 K across configurations on the common Henan sample (N = 183,964), a spread of approximately 0.22 K (Tables S4 and S15). Viewing geometry, emissivity treatment, within-grid heterogeneity, and clear-sky sampling affect MODIS–TSK comparability, while higher pixel coverage reduces sparse-sampling effects but does not eliminate these differences. Their contributions to the shared offset are not isolated. The C5m–CTRL MSE reductions mainly reflect lower squared mean bias despite increased centered error variance (Table S15). Additional model soil moisture and surface energy diagnostics are provided in Figure S10 and Table S16, though these diagnostics do not independently identify the causes of the shared warm bias. Smoothing of mapped static fields may influence performance but was not isolated. Because C5m changes both terrain and land cover and its 5′ source spacing is coarser than the fixed 3 km D02 grid, neither a resolution-only benefit nor a scale-matching advantage is established. These event-specific results do not identify compensation among particular physical processes.

5.3. Contribution in Relation to Previous Studies

Previous studies have examined land-cover sensitivity during heatwaves [12], evaluated combined terrain and land-cover replacements in WRF [13], and investigated land-cover aggregation and physical-parameter representation [14]. Here, WPS-mapped HGT_M and LU_INDEX fields are compared on a fixed 3 km grid, and CTRL, DEM, LU, and LU + DEM quantify separate and non-additive TSK and T2 responses. C5m provides a combined-product comparison outside this factorial analysis. The contribution of this study lies in its integration of documented preprocessing, direct interaction diagnostics, and temperature evaluation against station and satellite observations. No claim is made that any individual component is used for the first time. Internal diagnostics provide process context without isolating intermediate mechanisms. For this event and model configuration, products with higher nominal native resolution did not consistently improve temperature performance, and this product comparison does not isolate the effect of spatial resolution.

5.4. Study Limitations

This study covers one June 2022 heatwave in Henan under one WRF configuration. The land-cover comparison combines differences in source, representative period, native resolution, and classification representation, including water/lake treatment; however, their separate effects are not isolated. D01 and D02 remain fixed at 9 and 3 km, respectively, so no model-grid-resolution effect is tested. In addition, station measurements differ from model diagnostics in measurement height and spatial representativeness. Without independent process observations and targeted sensitivity experiments, the causes of the shared warm bias and the physical explanation for the smaller C5m errors remain unresolved. Transferability requires further events, regions, and configurations.

MODIS thermal-infrared LST is mainly retrieved under clear-sky conditions, so cloud screening and quality control constrain the sampled locations and overpass times. Without independent cloud and precipitation data, the weather conditions associated with missing retrievals and their effects on the evaluation cannot be quantified. Therefore, this evaluation represents TSK–LST agreement for quality-controlled retrievals during the event, rather than all-weather surface-temperature performance across Henan. Common sampling supports comparisons among experiments but does not remove this limitation. Station air-temperature observations provide complementary information but cannot replace missing LST observations. The two variables represent different physical quantities and spatial and temporal scales.

6. Conclusions

In this study, we investigate how remotely sensed topographic and land-cover static boundary conditions modulate mesoscale simulations of extreme heat, using one control run and four sensitivity experiments for the June 2022 heatwave over Henan Province. We quantify how individual and coupled remote-sensing-derived terrain and land-cover datasets modify the model's lower-boundary representation, internal physical responses, and performance in reproducing surface and near-surface thermal conditions. The key physical findings and methodological implications of our study are summarized below.

Simulation differences associated with the remote-sensing-derived surface fields reflect responses to both land-cover and topographic disparities. Land-cover discrepancies modify near-surface momentum and energy exchange through the Noah-MP land-surface scheme. Replacing MODIS15s-lake with MCD12Q1 changed LU_INDEX in 22.2% of valid grid cells and altered the representation of cropland and urban areas. The daytime HFX

difference fields had spatial RMS values exceeding 20 W/m², and these changes were accompanied by pronounced spatial and diurnal responses in simulated surface skin temperature (TSK). By contrast, DEM-related terrain differences were accompanied by hydrostatically consistent surface-pressure differences. ASTER30 produced comparatively small terrain differences (MAE = 7.5 m), whereas GMTED5m produced larger differences (MAE = 22.8 m) and PSFC differences ranging from −35.17 to 43.83 hPa. The strong inverse relationship between ΔHGT_M and ΔPSFC (R = −0.9990 to −0.9993) is consistent with hydrostatic pressure–height relationships but does not demonstrate circulation adjustments or quantify their contribution to temperature responses.

T2 RMS_{st} responses were smaller than the corresponding TSK responses for All, Day, and Night, with a larger contrast during daytime (Table S11). This pattern is consistent with possible modulation by near-surface exchange and boundary-layer processes, but the diagnostics do not isolate their individual contributions. Direct interaction residuals indicate non-additive TSK and T2 responses for this event and model configuration, with full-period RMS_{mean} values of 0.421 and 0.225 K, respectively (Figure S4; Table S10A).

No configuration consistently minimized errors across variables and periods. The common-sample pooled TSK bias remained 5.52–5.74 K. C5m had the smallest station-based T2 errors; LU had the smallest daytime TSK errors; and C5m had the smallest pooled and nighttime TSK errors. Pooled and Night MAE/RMSE rankings persisted across coverage thresholds, whereas some close Day rankings changed. Thus, stronger internal responses did not ensure better observational agreement.

We propose a framework that organizes terrain-related and land-cover-related diagnostic pathways. This framework compares WPS-mapped HGT_M and LU_INDEX differences, internal model responses in PSFC, UST, HFX, PBLH, TSK, and T2, and temperature evaluation against MODIS LST and station observations, distinguishing internal physical-consistency diagnostics from observation-based performance without establishing causality at every link.

For the examined event and model configuration, using static-data products with higher nominal native resolution did not produce consistent improvements in temperature simulation. The compared products differed in data source, representative period, and nominal native resolution; the land-cover products also differed in classification representation, including water/lake treatment. Therefore, these experiments do not isolate the independent effect of spatial resolution. Static-data selection should consider these product characteristics together with WPS grid mapping and interactions with model physics. While the workflow combines WPS-mapped field checks, model-process diagnostics, and station and MODIS LST evaluation as a reference for assessing static datasets in regional heatwave simulations, its transferability to other regions and events remains to be tested.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18213760/s1, Figure S1: Spatial coverage of common valid D02 grid-level MODIS LST–WRF TSK paired samples for daytime and nighttime overpasses during 12–19 June 2022; Figure S2: Spatial distribution of valid sampling frequency for (a) daytime and (b) nighttime MODIS LST–WRF TSK evaluation during 12–19 June 2022; Figure S3: Daily differences in MODIS LST–WRF TSK errors relative to CTRL for common valid D02 grid-cell pairs from 12–19 June 2022; Figure S4: Spatial distributions of the time-mean terrain–land-cover interaction residual for TSK and T2 over Henan for the full period, daytime, and nighttime; Figure S5: Spatial distributions of the temporal RMS of the hourly terrain–land-cover interaction residual for TSK and T2 over Henan for the full period, daytime, and nighttime; Figure S6: Beijing-time hourly composites of TSK and T2 responses to static-data changes over Henan; Figure S7: Beijing-time hourly composites of surface-exchange and boundary-layer-height responses to static-data changes over Henan; Figure

S8: Hourly observed 1.5 m air temperature and simulated native 2 m temperature at four illustrative stations during the 192-hour evaluation window; Figure S9: Mean hourly station-temperature bias by Beijing hour for common valid observations and all five configurations; Figure S10: Hourly model diagnostics averaged with equal weights over 17,871 fixed Henan grid cells: first-layer soil moisture, downward shortwave and longwave radiation, sensible and latent heat fluxes, and boundary-layer height; Table S1: D02 input grids and interpolation/aggregation settings; Table S2: D02 smoothing and missing-value settings; Table S3: Generation of fractional land-use coverage and dominant categories; Table S4: Sensitivity of TSK–MODIS LST errors and experiment rankings to the minimum paired-pixel coverage threshold; Table S5: Daily numbers of retained D02 grid-cell paired samples at each minimum paired-pixel coverage threshold; Table S6: Daily numbers and grid-cell coverage of common valid D02 grid-level MODIS LST–WRF TSK paired samples during 12–19 June 2022; Table S7: Distribution and representativeness of common valid paired samples across elevation groups and land-cover categories during 12–19 June 2022; Table S8: Daily errors and error differences relative to CTRL for common valid D02 grid-level MODIS LST–WRF TSK paired samples during 12–19 June 2022; Table S9: Leave-one-date-out sensitivity of accumulated errors and experiment rankings; Table S10: Regional statistics of terrain–land-cover interaction residuals and single-factor and combined temperature responses over Henan; Table S11: Regional-mean responses and RMS measures of TSK, T2, HFX, UST, and PBLH over Henan for the full period, day-time, and nighttime; Table S12: Daily interaction magnitudes and magnitudes relative to single-factor responses; Table S13: Leave-one-day-out interaction statistics and date influence; Table S14: Selection and common-hour coverage of illustrative stations; Table S15: Common-sample temperature errors and C5m–CTRL mean-squared-error decomposition; Table S16: Model soil-moisture and surface-energy diagnostics and descriptive associations; Table S17: D02 land-cover categories and land–water masks before and after WRF initialization; Table S18: D02 mask and available initialization surface-field changes relative to CTRL; Table S19: MODIS LST–WRF TSK errors stratified by elevation group; Table S20: MODIS LST–WRF TSK errors stratified by land-cover category.

Author Contributions: Conceptualization and methodology, C.X., W.D. and A.S.; MODIS data processing, H.T. and Y.Z.; DEM data processing, Y.Q., B.L. and J.W.; model simulations and data processing, W.D., H.C., M.Z. and H.K.; writing, W.D., C.X., H.C. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (Grant No. 2024YFD2301301), which provided primary financial support for this study, the Henan Province Key R&D and Promotion Special Project (Science and Technology Research) (Grant No. 232102320031), the Applied Technology Research Fund of the China Meteorological Administration (CMA)/Henan Key Laboratory of Agrometeorological Support and Applied Technique (Grant No. KP202306), the Henan Provincial Key Scientific and Technological Projects (Grant No. 241111220900), and the Key Project of the Open Research Fund of Anyang National Climate Observatory (Grant No. AYNC0F202419).

Data Availability Statement: The data used in this study comprise publicly available source datasets, restricted-access station observations, and model inputs and outputs. NCEP FNL analyses are available from the NCAR Research Data Archive. MODIS MOD11A1 V061 LST, ASTER GDEM Version 3, and MCD12Q1 V061 land-cover products are available from NASA Earthdata. Observations from 111 national meteorological stations were obtained from the Henan Meteorological Big Data Cloud Platform (Tianqing) and are not publicly available because of access restrictions. Requests for these observations may be directed to the corresponding author, but access is subject to approval by the data provider. The supporting dataset package, Henan_ExtremeHeat, is available from the corresponding author upon reasonable request. It includes D02 static geographical fields (HGT_M and LU_INDEX) for CTRL, DEM, LU, LU + DEM, and C5m, hourly D02 model outputs for 16 June 2022, preprocessing and model configuration files, and README documentation. The

complete model output archive is also available from the corresponding author upon reasonable request but is not included in the supporting package owing to its large volume.

Acknowledgments: The authors acknowledge the Henan Meteorological Big Data Cloud Platform (Tianqing) for providing access to the meteorological station observations used in this study. During the preparation of the major revision of this manuscript, the authors used OpenAI's ChatGPT with the GPT-5.6 Sol model for language polishing, translation, and assistance in refining the wording of the responses to the reviewers. The authors have reviewed and edited all AI-generated outputs and take full responsibility for the content of this publication.

Conflicts of Interest: Author Xiefei Zhi was employed by the company Stellerus Technology Co., Ltd., Hong Kong, China. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Seneviratne, S.I.; Zhang, X.; Adnan, M.; Badi, W.; Dereczynski, C.; Di Luca, A.; Ghosh, S.; Iskandar, I.; Kossin, J.; Lewis, S.; et al. Weather and Climate Extreme Events in a Changing Climate. In *Climate Change 2021: The Physical Science Basis*; Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., et al., Eds.; Cambridge University Press: Cambridge, UK, 2021; pp. 1513–1766. https://doi.org/10.1017/9781009157896.013.
2. Intergovernmental Panel on Climate Change. Summary for Policymakers. In *Climate Change 2021: The Physical Science Basis*; Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., et al., Eds.; Cambridge University Press: Cambridge, UK, 2021; pp. 3–32. https://doi.org/10.1017/9781009157896.001.
3. National Climate Center; China Meteorological Administration. *China Climate Bulletin 2022*; National Climate Center: Beijing, China, 2023.
4. Sun, B.; Wang, H.J.; Huang, Y.Y.; Yin, Z.C.; Zhou, B.T.; Duan, M.K. Characteristics and Causes of the Hot-Dry Climate Anomalies in China during Summer of 2022. *Trans. Atmos. Sci.* **2023**, *46*, 1–8. https://doi.org/10.13878/j.cnki.dqkxxb.20220916003. (In Chinese)
5. Zhang, D.; Huang, Y.; Zhou, B.; Wang, H.; Sun, B. Who Is the Major Player for 2022 China Extreme Heat Wave? Western Pacific Subtropical High or South Asian High? *Weather Clim. Extrem.* **2024**, *43*, 100640. https://doi.org/10.1016/j.wace.2024.100640.
6. Giannaros, C.; Melas, D.; Giannaros, T.M. On the Short-Term Simulation of Heat Waves in the Southeast Mediterranean: Sensitivity of the WRF Model to Various Physics Schemes. *Atmos. Res.* **2019**, *218*, 99–116. https://doi.org/10.1016/j.atmosres.2018.11.015.
7. Yoon, D.; Kang, T.; Cha, D.-H.; Min, K.-H.; Kim, J.; Chun, J.A.; Seo, E. Role of Land–Atmosphere Interaction in the 2016 Northeast Asia Heat Wave: Impact of Soil Moisture Initialization. *J. Geophys. Res. Atmos.* **2023**, *128*, e2022JD037718. https://doi.org/10.1029/2022JD037718.
8. Li, S.; Liu, Y.; Pan, Y.; Li, Z.; Lyu, S. Integrating Remote-Sensing and Assimilation Data to Improve Air Temperature on Hot Weather in East China. *Remote Sens.* **2021**, *13*, 3409. https://doi.org/10.3390/rs13173409.
9. Ma, B.; Shao, Y.; Yang, H.; Lu, Y.; Gao, Y.; Wang, X.; Xie, Y.; Wang, X. Land Cover Mapping in East China for Enhancing High-Resolution Weather Simulation Models. *Remote Sens.* **2024**, *16*, 3759. https://doi.org/10.3390/rs16203759.
10. Siewert, J.; Kroszczyński, K. Evaluation of High-Resolution Land Cover Geographical Data for the WRF Model Simulations. *Remote Sens.* **2023**, *15*, 2389. https://doi.org/10.3390/rs15092389.
11. Niu, G.-Y.; Yang, Z.-L.; Mitchell, K.E.; Chen, F.; Ek, M.B.; Barlage, M.; Kumar, A.; Manning, K.; Niyogi, D.; Rosero, E.; et al. The Community Noah Land Surface Model with Multiparameterization Options (Noah-MP): 1. Model Description and Evaluation with Local-Scale Measurements. *J. Geophys. Res. Atmos.* **2011**, *116*, D12109. https://doi.org/10.1029/2010JD015139.
12. Thisara Sathsara, K.L.; Kusaka, H. Impact of High-Resolution Land Use Data on Numerical Simulations of Colombo's Thermal Environment during Heatwaves. *Theor. Appl. Climatol.* **2025**, *156*, 106. https://doi.org/10.1007/s00704-024-05242-9.
13. De Meij, A.; Vinuesa, J.F. Impact of SRTM and Corine Land Cover Data on Meteorological Parameters Using WRF. *Atmos. Res.* **2014**, *143*, 351–370. https://doi.org/10.1016/j.atmosres.2014.03.004.

14. de Bode, M.; Hedde, T.; Roubin, P.; Durand, P. A Method to Improve Land Use Representation for Weather Simulations Based on High-Resolution Data Sets—Application to Corine Land Cover Data in the WRF Model. *Earth Space Sci.* **2023**, *10*, e2021EA002123. https://doi.org/10.1029/2021EA002123.
15. Gallo, K.; Hale, R.; Tarpley, D.; Yu, Y. Analysis of MODIS LST Compared with WRF Model and in Situ Data over the Waimakariri River Basin, Canterbury, New Zealand. *Remote Sens.* **2012**, *4*, 3501–3527. https://doi.org/10.3390/rs4113501.
16. Duan, S.-B.; Li, Z.-L.; Li, H.; Göttsche, F.-M.; Wu, H.; Zhao, W.; Leng, P.; Zhang, X.; Coll, C. Validation of Collection 6 MODIS Land Surface Temperature Product Using In Situ Measurements. *Remote Sens. Environ.* **2019**, *225*, 16–29. https://doi.org/10.1016/j.rse.2019.02.020.
17. National Centers for Environmental Prediction; National Weather Service; NOAA; U.S. Department of Commerce. *NCEP FNL Operational Model Global Tropospheric Analyses, Continuing from July 1999*; NCAR Research Data Archive: Boulder, CO, USA, 2000. https://doi.org/10.5065/D6M043C6.
18. Danielson, J.J.; Gesch, D.B. *Global Multi-Resolution Terrain Elevation Data 2010 (GMTED2010)*; U.S. Geological Survey Open-File Report 2011-1073; U.S. Geological Survey: Reston, VA, USA, 2011; 26p. https://doi.org/10.3133/ofr20111073.
19. Abrams, M.; Crippen, R.; Fujisada, H. ASTER Global Digital Elevation Model (GDEM) and ASTER Global Water Body Dataset (ASTWBD). *Remote Sens.* **2020**, *12*, 1156. https://doi.org/10.3390/rs12071156.
20. Broxton, P.D.; Zeng, X.; Sulla-Menashe, D.; Troch, P.A. A Global Land Cover Climatology Using MODIS Data. *J. Appl. Meteorol. Climatol.* **2014**, *53*, 1593–1605. https://doi.org/10.1175/JAMC-D-13-0270.1.
21. Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Liu, Z.; Berner, J.; Wang, W.; Powers, J.G.; Duda, M.G.; Barker, D.M.; et al. *A Description of the Advanced Research WRF Model Version 4*; NCAR Technical Note NCAR/TN-556+STR; National Center for Atmospheric Research: Boulder, CO, USA, 2019; 145p. https://doi.org/10.5065/1dfh-6p97.
22. Friedl, M.A.; McIver, D.K.; Hodges, J.C.F.; Zhang, X.Y.; Muchoney, D.; Strahler, A.H.; Woodcock, C.E.; Gopal, S.; Schneider, A.; Cooper, A.; et al. Global Land Cover Mapping from MODIS: Algorithms and Early Results. *Remote Sens. Environ.* **2002**, *83*, 287–302. https://doi.org/10.1016/S0034-4257(02)00078-0.
23. Friedl, M.; Sulla-Menashe, D. *MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500 m SIN Grid V061*; NASA EOSDIS Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA, 2022. https://doi.org/10.5067/MODIS/MCD12Q1.061.
24. Wang, W.; Bruyère, C.; Duda, M.G.; Dudhia, J.; Gill, D.O.; Kavulich, M.; Keene, K.; Lin, H.-C.; Michalakes, J.; Rizvi, S.; et al. *ARW Version 3 Modeling System User's Guide*; National Center for Atmospheric Research: Boulder, CO, USA, 2014.
25. Ge, J.; Pitman, A.J.; Guo, W.; Zan, B.; Fu, C. Do Uncertainties in the Reconstruction of Land Cover Affect the Simulation of Air Temperature and Rainfall in the CORDEX Region of East Asia? *J. Geophys. Res. Atmos.* **2019**, *124*, 3647–3670. https://doi.org/10.1029/2018JD029945.
26. Delucchi, L. *pyModis: Python Library to Download and Process MODIS Data from NASA Server*, Version 2.4.1; 2024. Available online: https://pypi.org/project/pyModis/2.4.1/ (accessed on 19 August 2026).
27. GDAL/OGR Contributors. *GDAL/OGR Geospatial Data Abstraction Software Library*, Version 3.4.3; Open Source Geospatial Foundation: Beaverton, OR, USA, 2022. Available online: https://gdal.org (accessed on 19 August 2026).
28. PROJ Contributors. *PROJ: Cartographic Projections and Coordinate Transformations Library*, Version 8.2.1; Open Source Geospatial Foundation: Beaverton, OR, USA, 2022. Available online: https://proj.org (accessed on 19 August 2026).
29. Wan, Z.; Hook, S.; Hulley, G. *MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1 km SIN Grid V061*; NASA EOSDIS Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA, 2021. https://doi.org/10.5067/MODIS/MOD11A1.061.
30. Hong, S.-Y.; Dudhia, J.; Chen, S.-H. A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation. *Mon. Weather Rev.* **2004**, *132*, 103–120. https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2.
31. Iacono, M.J.; Delamere, J.S.; Mlawer, E.J.; Shephard, M.W.; Clough, S.A.; Collins, W.D. Radiative Forcing by Long-Lived Greenhouse Gases: Calculations with the AER Radiative Transfer Models. *J. Geophys. Res. Atmos.* **2008**, *113*, D13103. https://doi.org/10.1029/2008JD009944.
32. Yang, Z.-L.; Niu, G.-Y.; Mitchell, K.E.; Chen, F.; Ek, M.B.; Barlage, M.; Longuevergne, L.; Manning, K.; Niyogi, D.; Tewari, M.; et al. The Community Noah Land Surface Model with Multiparameterization Options (Noah-MP): 2. Evaluation over Global River Basins. *J. Geophys. Res. Atmos.* **2011**, *116*, D12110. https://doi.org/10.1029/2010JD015140.
33. Hong, S.-Y.; Noh, Y.; Dudhia, J. A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. *Mon. Weather Rev.* **2006**, *134*, 2318–2341. https://doi.org/10.1175/MWR3199.1.
34. Jiménez, P.A.; Dudhia, J.; González-Rouco, J.F.; Navarro, J.; Montávez, J.P.; García-Bustamante, E. A Revised Scheme for the WRF Surface Layer Formulation. *Mon. Weather Rev.* **2012**, *140*, 898–918. https://doi.org/10.1175/MWR-D-11-00056.1.

35. Kain, J.S. The Kain–Fritsch Convective Parameterization: An Update. *J. Appl. Meteorol.* **2004**, *43*, 170–181. https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.
36. National Center for Atmospheric Research. WRF Model Users' Guide: Dynamics. Available online: https://www2.mmm.ucar.edu/wrf/site/documentation/users_guide/dynamics.html (accessed on 19 August 2026).
37. Efron, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.* **1979**, *7*, 1–26. https://doi.org/10.1214/aos/1176344552.
38. Gilleland, E. Bootstrap Methods for Statistical Inference. Part I: Comparative Forecast Verification for Continuous Variables. *J. Atmos. Ocean. Technol.* **2020**, *37*, 2117–2134. https://doi.org/10.1175/JTECH-D-20-0069.1.
39. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 3rd ed.; Academic Press: Oxford, UK, 2011.
40. Holton, J.R.; Hakim, G.J. *An Introduction to Dynamic Meteorology*, 5th ed.; Academic Press: Amsterdam, The Netherlands, 2013.
41. Paulson, C.A. The Mathematical Representation of Wind Speed and Temperature Profiles in the Unstable Atmospheric Surface Layer. *J. Appl. Meteorol.* **1970**, *9*, 857–861.
42. He, C.; Valayamkunnath, P.; Barlage, M.; Chen, F.; Gochis, D.; Cabell, R.; Schneider, T.; Rasmussen, R.; Niu, G.-Y.; Yang, Z.-L.; et al. *The Community Noah-MP Land Surface Modeling System Technical Description Version 5.0*; NCAR Technical Note NCAR/TN-575+STR; National Center for Atmospheric Research: Boulder, CO, USA, 2023. https://doi.org/10.5065/ew8g-yr95.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.