https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

Open Access Earth System Science Data Discussions

High-Resolution Wide-Coverage Urban Canopy Parameters for Urban Simulations in Weather Research and Forecasting Models

Melissa R. Allen-Dumas [1], Pouya Vahmani [2], Bhartendu Pandey [1], Chris Vernon [3], Levi Sweet-Breu [1,4], and Em Rexer [3]

[1]Oak Ridge National Laboratory, Oak Ridge, TN, USA
[2]Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[3]Pacific Northwest National Laboratory, Richland, WA, USA
[4]Baylor University, Waco, TX, USA

Correspondence: Melissa R. Allen-Dumas (allenmr@ornl.gov)

Abstract. Cross-disciplinary researchers focusing on connected urban processes, especially those running numerical weather simulations at microscales, require reliable data on building footprints, heights and locations. For example, quantifying the impact of the thermal radiative properties of buildings on urban heating during a heat wave benefits from representation of the 3-dimensional physical characteristics of buildings within the urban location studied. Additionally, understanding how and where urban pollution travels within a city depends on how the buildings from ground level to building tops distort the air flow throughout the city. Even knowledge about which buildings may be susceptible to flooding given the magnitude of a recent rainstorm can be clarified by knowing where buildings are located with respect to the elevation of the earth beneath and surrounding them. As urban micrometeorological modeling helps answer more local scientific questions, high resolution data with wide coverage are needed. Previous research into these issues has yielded useful products for simulating these effects at resolutions of 1 kilometer and coarser; however, no regional-weather-model-readable data products are available at block level resolution for the full extent of any city, county, or wider region. To address this gap, four data sets are presented here: 1) Chicago, 2) Washington, DC, 3) Los Angeles County, and 4) the Arizona "urban corridor." Parameters include, at 100-meter resolution, frontal area density, plan area density, rooftop area density, plan area fraction, mean building height, standard deviation of building heights, area weighted mean of building heights, building surface area to plan area ratio, height to width ratio, sky view factor and roughness length calculations. These data sets were generated using a new Python tool and validated using statistical and visual methods.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

1 Background and Summary

Urban research provides a critical information for shaping solution-oriented action to address interdependent issues involving humans and the environment (Brelsford et al., 2024). Outcomes of human action, which originates in urban areas, are the result of the connections among urban infrastructure sectors, nested geographic scales of action and influence, local biogeochemistry and micrometeorology. Additionally, the mechanisms and perspectives research brings to system observation and understanding can affect how results are presented and further action is determined. However, to understand the details of coupled human-natural system processes and consequences, research must move beyond simple conceptual coupling of separate models of systems and subsystems (Müller-Hansen et al., 2017) and begin to integrate processes more interactively in empirical and system modeling, and by robustly incorporating human processes into analyses of natural systems (Schlüter et al., 2023).

While the need exists for diverse qualitative and quantitative approaches approaches to research in this space, many numerical weather modeling researchers (e.g., Oke, 1988) have long been studying the contribution of built structures as surface roughness elements to the thermal dynamics of urban areas. Buildings in particular have been shown to increase three-dimensional surface absorption of solar radiation, reduce heat transfer out of the system, and reduce the loss of long-wave radiation from urban neighborhoods (Shahmohamadi et al., 2011; Allen-Dumas et al., 2024). Additionally, exhaust heat from urban building heating and cooling to the urban environment contributes significantly to ambient temperature (Luo et al., 2020). Representing these sources of urban heat within numerical weather models requires high-resolution, urban-scale inputs, as the communication of a neighborhood's urban morphology to the surrounding natural environment is critical to modeling the entire surface energy budget. Urban surfaces absorb, reflect, and radiate heat (Pappaccogli et al., 2020). They also shade different areas throughout a day affecting thermal transport pathways in built up areas (Best, 2006; Ching et al., 2009; Oleson et al., 2010). All of these surface processes influence atmospheric flow patterns above the surfaces that transport heat, mass and momentum (Martilli, 2014; Oke, 1982). The most efficient means for investigating the impact of buildings on the urban atmosphere is to use physics-based numerical weather models that incorporate these built surface features as parameterizations within the model domain (Gabrian et al., 2025; Simpson et al.). While satellite observations and artificial intelligence approaches to developing three-dimensional settlement datasets have been used in the past to represent surface temperature contributions to the urban heat island (Chakraborty et al., 2020; García, 2022; Mohamed and Zahidi, 2024), such measurements are still too coarse to show the details of built structures' contribution to the urban heat island. Representing built morphological characteristics of urban areas as inputs to weather models at sub-kilometer scale allows weather models to calculate complex spatial distributions of heat due to buildings on the local atmosphere so that their amplification of heat waves, for example, can be estimated at the neighborhood level (Allen-Dumas et al., 2020).

Besides modifying the local and regional environments, local built environments can impact geographies elsewhere owing to teleconnections (Seto et al., 2012), as well as influence environmental changes at the planetary scale (Grimm et al., 2008) Understanding these local-to-global impacts, requires understanding urban forms and attendant heterogeneity (Seto and Pandey, 2019). Urban morphology quantification integrating horizontal and vertical dimensions can help advance this science goal by subsequently enabling data-driven characterizations (Fleischmann et al., 2025) to link with proximal as well as distal impacts.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

Currently, UCP datasets based on NUDAPT (Ching et al., 2009) for WRF's Building Effect Parameterization (BEP) urban meteorology scheme are available at 1km resolution. 44 US and 60 Chinese cities' business districts are readily available. Inclusion of this information, when it was generated in 2009, improved the output of the WRF model significantly at urban scale because not only could an impervious surface be incorporated into the model's environment, but the effect of impervious walls at realistic heights forming urban "canyons" could be evaluated for the atmosphere at typical meteorological measurement heights such as 2 meters (temperature, humidity) and 10 meters (wind, turbulent kinetic energy) and validated using land based observations. Subsequently, Liao et al. (2024) developed a global-coverage, 1km resolution, spatially continuous set of urban canopy parameters (GloUCP) dataset for use by the WRF model. Additionally, Cheng et al. (2024) used remote sensing and machine learing to develop a separate 1km resolution global extent urban surface parameter dataset (U-Surf) for use primarily in the Community Earth System global circulation model, although the parameters are adaptable to regional weather models.

As researchers seek more detailed information about meteorological processes at neighborhood level, however, higher resolution parameters pertaining to the built environment that cover entire cities, counties, and larger areas are needed to understand how heat, atmospheric transport and even precipitation affect highly local areas in a city (Li et al., 2024). A major step toward this goal is shown in the World Settlement Footprint 3D (WSF) urban dataset at 90m resolution of Esch et al. (2022), for which which estimates of mean building height, building fraction, total building area, average building height, and total built-up volume were calculated from satellite data and digital elevation data.

Here we present four high-resolution (100m horizontal and 5m vertical) resolution wide-coverage (full city to multiple counties) datasets as urban canopy parameter (UCP) inputs for the Weather Research and Forecasting (WRF) model. These datasets represent the type of information required for detailed characterization of the micrometeorological impacts of buildings in urban neighborhoods across entire cities and in connection with rural areas and larger regions. Based on cm resolution input shapefiles, 132 UCPs are calculated for each grid cell. We provide datasets for four U.S. locations: 1) Chicago, IL 2) Washington, DC, 3) Los Angeles County, CA, and 4) the Arizona "urban corridor." Table 1 lists the UCPs provided.

2 Methods

Generation of these data sets was accomplished using the Neighborhood Adaptive Tissues for Urban Resilience Futures (NATURF) tool (Sweet-Breu et al., 2024). This newly-developed tool is a set of Python modules that generate index and binary files readable by the Weather Research and Forecasting (WRF) model, and CSV files for research outside of weather modeling. A flexible adaptation of the National Urban Database and Access Portal Tool (NUDAPT)(Ching et al., 2009), NATURF calculates 132 building dimension parameters (Table 1) at user-defined horizontal resolution from user-provided polygon shapefiles containing building footprint, height and location information. Each parameter's calculation is based on documented, peer-reviewed, previously-validated equations for the parameter. Literature references for each parameter are given in the right-hand column of Table 1. As prescribed by NUDAPT, the vertical component of these parameters is calculated at 15 5m intervals resulting in an upper bound for building height of 75m. Thus, each equation may define up to 60

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

Open Access Earth System Science Data Discussions

Table 1. NATURF Urban Canopy Parameters at 15 Vertical Levels at 5m Intervals.

Urban Canopy Parameter (UCP)	Formula	Units	Values 100m Res	References
Frontal area density (for each 5m height increment and each wind direction) (1-60)	$\frac{FrontalArea}{TotalLotArea}$	$\frac{m^2}{m^2}$	0 - 0.04	Burian et al. (2003)
Plan area density (for each 5m increment above ground level; 61-75)	$\frac{BuildingFootprintArea}{TotalLotArea}$	$\frac{m^2}{m^2}$	0 - 0.5	Burian et al. (2003)
Rooftop area density (76-90)	$\frac{RoofArea}{TotalLotArea}$	$\frac{m^2}{m^2}$	0 - 0.25	Burian et al. (2003)
Plan area fraction (ground level; 91)	$\frac{BuildingFootprintArea}{TotalLotArea}$	$\frac{m^2}{m^2}$	0 - 1	Burian et al. (2003)
Mean building height (mbh; 92)	$\frac{bh_1+bh_2+bh_3+...+bh_n}{n}$	m	0 - 400	Burian et al. (2003)
Standard deviation of building heights (93)	$\sqrt{\frac{\Sigma(bh_i-mbh)^2}{n}}$	m	0 - 40	Burian et al. (2003)
Area weighted mean of building heights (94)	$\frac{\Sigma A_i bh_i}{\Sigma A_i}$	m	0 - 40	Burian et al. (2003)
Building surface area (bsa) to plan area ratio (95)	$\frac{BuildingSurfaceArea}{TotalLotArea}$	$\frac{m^2}{m^2}$	0 - 1.5	Burian et al. (2003)
Frontal area index (96-99)	((*MeanBreadthofBuilding*) * (*BuildingHeight*)) / ((*Mean AlongwindDistanceBetweenBuildingCentroids*) * (*Mean CrosswindDistanceBetweenBuildingCentroids*))	$\frac{m^2}{m^2}$	0 - 0.9	Burian et al. (2003)
Complete Aspect Ratio (100)	$\frac{ExposedSurfaceArea}{TotalLotArea}$	$\frac{m^2}{m^2}$	1 - 2	Burian et al. (2003)
Height to width ratio (101)	$\frac{BuildingHeight}{StreetWidth}$	$\frac{m}{m}$	0 - 2	Burian et al. (2003)
Sky View Factor (102)	$\cos(\arctan(\frac{bh}{0.5W}))$	frac	0 - 1	Dirksen et al. (2019)
Grimmond and Oke (1999) Rule of Thumb (Rt) Roughness Length (103)	$0.1 * bh$	m	0.2 - 1.3	Grimmond and Oke (1999)
Grimmond and Oke (1999) Rule of Thumb (Rt) Displacement Height (104)	$0.67 * bh$	m	2 - 5	Grimmond and Oke (1999)
Raupach (1994) Roughness Length (105, 107, 109, 111)	$bh*(1-RDH)*exp(-\kappa*(C_S+C_R*)^-0.5-\Psi_h))$	m	0 - 3	Raupach (1994)
Raupach (1999) Displacement Height (106, 108, 110, 112)	$bh*(1-(\frac{1-\exp(-\sqrt{(c_{d1}*\Lambda)})}{\sqrt{(c_{d1}*\Lambda)}}))$	m	1 - 13	Raupach (1994)
MacDonald et al. (1998) Roughness Length (113-116)	$bh*(1-RDH)\exp(-(0.5\frac{C_D}{\kappa^2}(1-RDH)\frac{A_f}{A_d})^-0.5)$	m	0 - 4	Macdonald et al. (1998)
MacDonald et al. (1998) Displacement Height (117)	$bh*(1+\frac{1}{A^\lambda}*(\lambda-1))$	m	1 - 12	Macdonald et al. (1998)
Vertical Distribution of Building Heights (118-132)	(Total buildings with Frontal Area Density at the Vertical Level)/(Total buildings in all levels with Frontal Area Density)	frac	0 - 1	

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

parameters (indicated in parentheses in the first column) if it is calculated for all 15 height intervals and for all four cardinal directions.

The data sets highlighted in this paper include UCPs calculated at 100m resolution for the cities of Los Angeles, Chicago, Washington, DC and the six-county "urban corridor" of the state of Arizona. Typical range of the parameter value at 100-meter resolution is shown in the third column of Table 1. Input data to NATURF for these UCP data sets are subsets of Model America, Version 1 (New et al., 2021). This data set includes physical characteristics of the approximately 125 million buildings covering 98% of the United States building stock present through the year 2015. Model America building geometries draw on Microsoft building data (Microsoft, 2020), which include 2D polygons in latitude and longitude coordinates. For Model America, building centroids were identified and unique building identifiers were allocated to each building according to Wang et al. (2019). This identifier allowed for the merging of additional locally-sourced building footprint data to improve accuracy. Building heights were estimated using data gathered from the Advanced Land Observing Satellite (ALOS) World 3D 30m (AW3D30) digital surface model version 2.2 (Tadono et al., 2016) provided by Google Earth Engine. Information from this imagery was extracted using convolutional methods and disaggregated using the slope correction method of Marconcini et al. (2014). Missing height data was imputed with the median height of the twenty closest buildings in proximity.

For the generation of the datasets presented here, the Model America data was downloaded in CSV format and converted to polygon shapefiles using python geopandas and shapely libraries. The polygon shapefiles were read into NATURF and stored as a GeoDataFrame in which each building was identified individually and its UCPs were calculated based on its own geometry and that of its neighbors within each 100m grid cell. While NATURF is capable of reading in any building data set in shapefile format, Model America was chosen as input for these datasets for its reliability, coverage, and availability (New et al., 2022).

More information on NATURF can be found at https://github.com/IMMM-SFA/naturf/ . Output from NATURF for each location includes WRF-readable index and binary files, and CSV files. A conversion of each file to KMZ to display building height patterns on Google Earth is also included for visualization. These files enable users to make qualitative comparisons against Google's 3D building renderings. They also encourage broader validation by the research community.

3 Data Description

Here we describe the locations, the extents, and the horizontal resolution of the four datasets: 1) Chicago, IL 2) Washington, DC, 3) Los Angeles County, CA, and 4) the Arizona "urban corridor." Each dataset has a geopolitically defined boundary, and each has been used in previous research in a specific way. The boundary for each data set is characteristic of a different type of extent. For example, Chicago's boundary is the city limits, Washington, DC's boundary is the District, Los Angeles' boundary is a county and the Arizona Urban Corridor is multiple counties. These data sets can be accessed at the U.S. Department of Energy Office of Science MSD-LIVE data archive. File lists, number of variables, number of grid cells, DOI, and web address for the data sets are included in the description of each.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

3.1 Chicago, IL

Data Title: Urban Parameters for Chicago, Illinois, USA at 100m resolution (Vernon et al., 2026)

Data File List: index, binary file 00001-00499.00001-00455, chicago.csv, chicago_naturf_height_points.kmz

Number of Dataset Variables: 132

Number of Dataset Grid Cells: 485,870

Data DOI: 10.57931/2572297

Data URL: https://data.msdlive.org/records/599zb-k0458

Description: The city of Chicago is situated geographically on the western bank of Lake Michigan in the northern United States about half way between the western Continental Divide and the Atlantic Ocean. The elevation of the city of Chicago is 176 meters above mean sea level (MSL) (https://www.chicago.gov/city/en.html). The boundary used to extract the buildings for city of Chicago from Model America is that available from https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-City-Map/ewy2-6yfk . An earlier study, employing this method for generating the UCPs at 10m horizontal resolution, used a City of Chicago buildings shapefile data set as input (Allen-Dumas et al., 2020).

3.2 Washington, DC

Data Title: Urban Parameters Washington, DC 100m (Allen-Dumas et al., 2026b)

Data File List: index, binary file 00001-00248.00001-00222, dc_params.csv, dc_naturf_height_points.kmz

Number of Dataset Variables: 132

Number of Dataset Grid Cells: 57,273

Data DOI: 10.57931/2998536

Data URL: https://data.msdlive.org/records/4qrgx-8ss66

Description: Washington, DC (or the District of Columbia) is near the Mid-Atlantic U.S. coast. The district is surrounded by the state of Maryland to the north, east, and west. Its southwest border is the Potomac River–a border shared by the commonwealth of Virginia. The elevation of Washington, DC ranges from sea level at the Potomac River to 125 meters at Fort Reno Park in the upper northwest (Winegar, 1998). The Anacostia River, a major tributary to the Potomac, runs northeast to southwest through the southeastern part of the district. The district's climate includes generally mild temperatures with hot summers and year-round humidity (Means, 2010). The boundary defining the extent of the buildings processed for the city of Washington, DC was provided by the Tigerline 2018 US counties data set (U.S. Census Bureau, 2018). A recent study used this method to generate UCPs at 10m and 100m horizontal resolution using a shapefile from Open Data DC (Allen-Dumas et al., 2024). This study compared the sensitivity of the WRF model to both the extent and the resolution of these input parameters and to that of NUDAPT at 1km horizontal resolution.

3.3 Los Angeles County

Data Title: Urban Parameters Los Angeles County 100m version 2 (Sweet-Breu et al., 2026)

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

Data File List: index, binary file 00001-01549.00001-01796, LA_Parameters_01232024.csv, LA_naturf_height_points.kmz

Number of Data Variables: 132

Number of Dataset Grid Cells: 2,782,005

Data DOI: 10.57931/3000523

Data URL: https://data.msdlive.org/records/mp5mx-ztp89

Description: Los Angeles County is on the U.S. southwestern coast. It borders Ventura County to the west, Kern County to the north, San Bernardino County to the east, and Orange County to the south. At elevations starting at sea level on the coast up to 3,068 meters in the mountain areas, it spans two of the 344 U.S. National Centers for Environmental Information (NCEI) climate divisions: the south coast drainage and the southeast desert basin (Vose et al., 2014). Climate throughout the county is desert to semi-arid to Mediterranean with dry hot summers and wet mild winters. While precipitation in winter is mostly in the form of rain, in the higher elevation northern mountain areas, precipitation falls as snow in winter (Kesseli, 1942).

Of the four data sets we present here, Los Angeles County is the one that includes the largest area of densely spaced buildings. The boundary around the buildings included for this data set was defined by the 2018 Tigerline US counties (U.S. Census Bureau, 2018). The parameters generated for Los Angeles County were used in two studies to inform a generative model that projects new neighborhood morphologies based additionally on urban scaling laws, empirical fluctuation scaling, and error and bias analyses (Allen-Dumas et al.; Pandey et al.).

3.4 Arizona Urban Corridor

Data Title: Urban Parameters Arizona Urban Corridor 100m (Allen-Dumas et al., 2026a)

Data File List: index, binary file 00001-04003.00001-06665, az_corridor.csv, phoenix_naturf_height_points.kmz

Number of Data Variables: 132

Number of Dataset Grid Cells: 2,495,163

Data DOI: 10.57931/2998499

Data URL: https://data.msdlive.org/records/va0qd-qsn15

Description: UCPs for the Arizona urban corridor were generated at 100m horizontal in the same manner as the previously described data sets. However, the input dataset includes buildings within six Arizona counties (boundaries defined by the Tigerline data (U.S. Census Bureau, 2018)): Coconino, Yavapai, Maricopa, Gila, Pinal, and Pima. Covering elevations from deserts (30 meters), plateaus (2000 meters) to mountains (3500 meters) (Western Regional Climate Center, 2016), these divisions include nine Koppen-Geiger climate classifications from desert to semi-arid to temperate and Mediterranean with dry hot summers and mild winters, mild rainy winters and warm dry summers, or long cold winters, cool summers, and moderate overall precipitation (Peel et al., 2007). This binary data set represents the highest-coverage ($143,468 km^2$ area) UCP dataset of its kind implemented in the WRF modeling system (Deng et al., 2026).

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

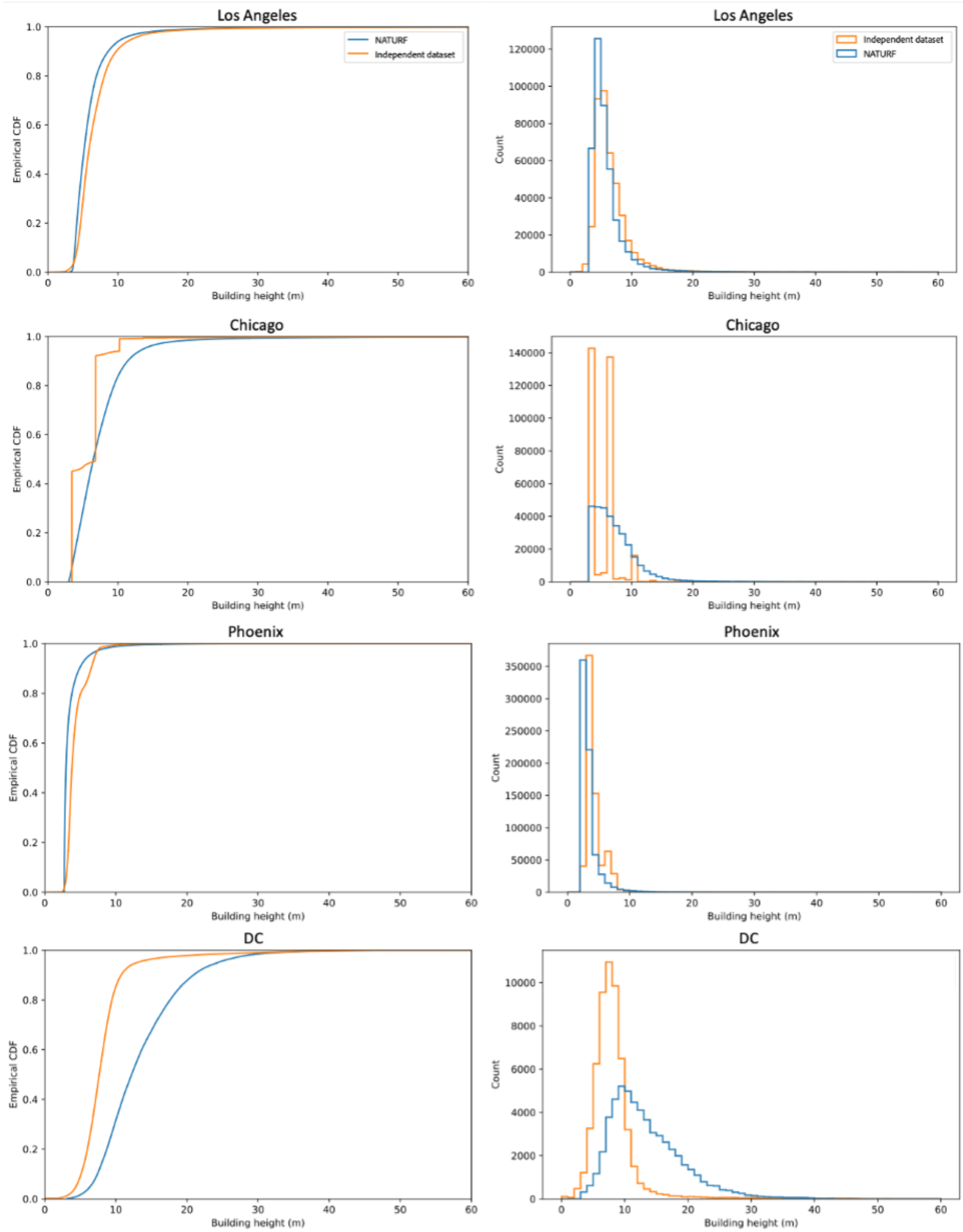

Figure 1. Empirical distributions of building heights across four U.S. cities. Empirical cumulative distribution functions (CDFs) and probability histograms of mean building height (meters) from NATURF and independent reference datasets for Los Angeles, Chicago, Phoenix, and Washington DC.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

Open Access Earth System Science Data Discussions

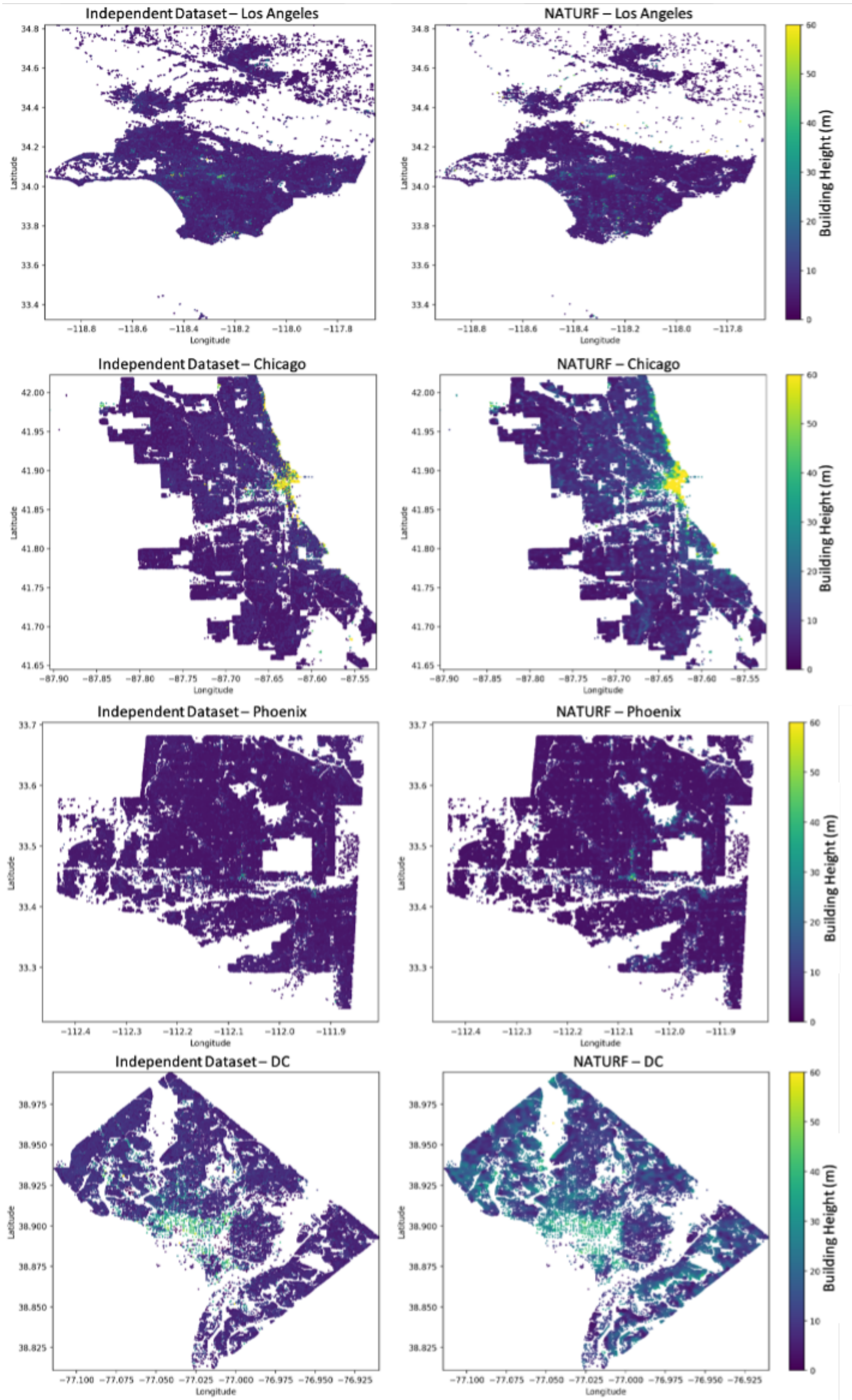

Figure 2. Spatial distribution of building height in NATURF and reference datasets. Maps of mean building height for the four evaluation regions at NATURF's native resolution (70–93 m).

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

Open Access Earth System Science Data Discussions

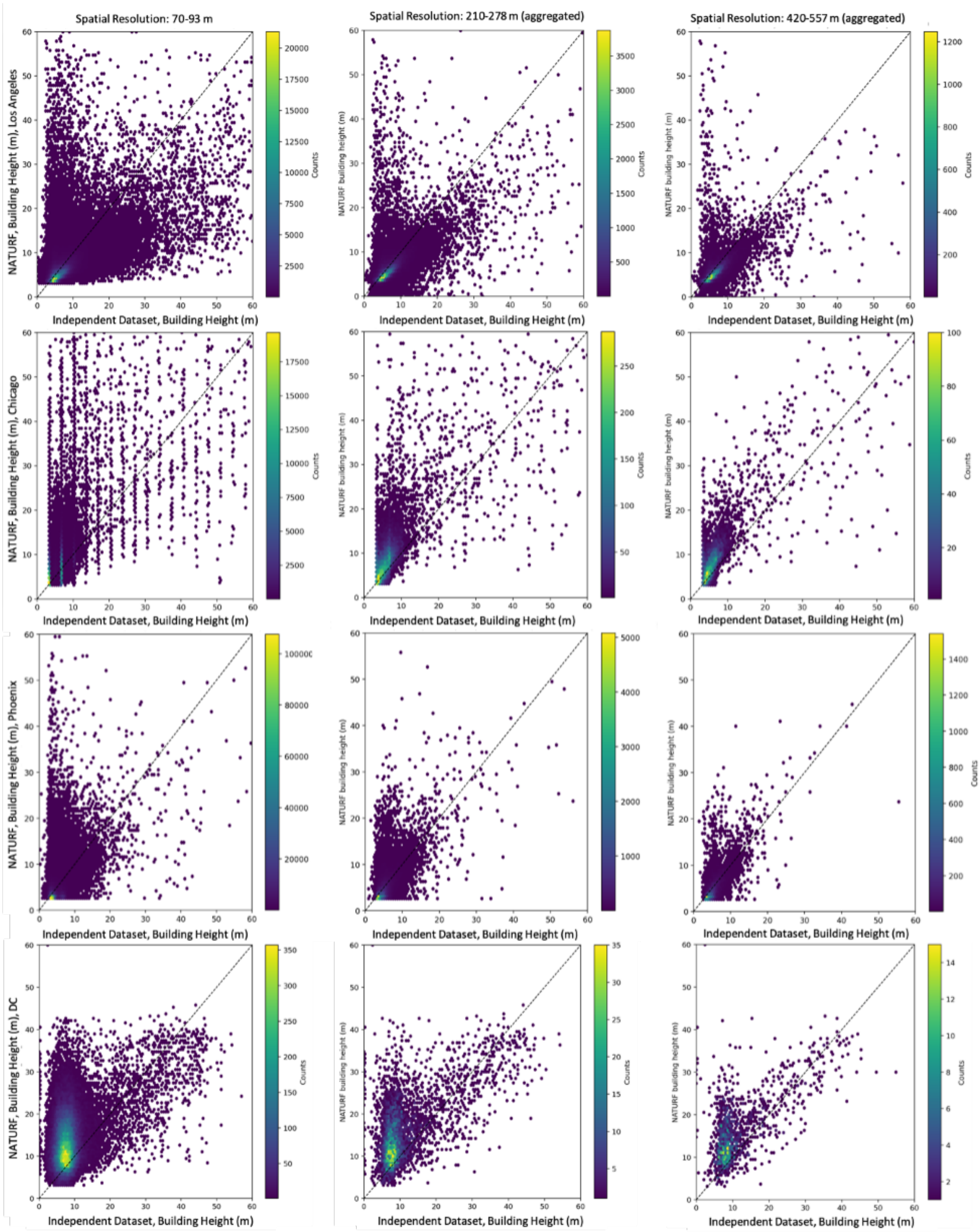

Figure 3. Multi-scale comparison of NATURF and independent building-height data. Scatterplots of NATURF versus independent mean building height at three aggregation scales (70-93 m, 210 m, 420–557 m) for Los Angeles, Chicago, Phoenix, and Washington DC.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

4 Technical Validation

The NATURF dataset was evaluated through a combination of internal consistency checks and external, city-specific comparisons against independent building datasets. Internal validation ensured the logical coherence of key UCPs—e.g., consistency between plan-area fraction, frontal-area indices, and mean building height—while external validation focused on quantitative comparison with authoritative or LiDAR-derived building inventories for four metropolitan regions. These comparisons provide an assessment of how well NATURF reproduces real-world urban structure and variability in built morphology at 100m resolution.

4.1 Comparative Evaluation

The primary focus of this technical validation is building height, a foundational variable within the NATURF framework from which other parameters such as roughness length, sky view factor, and radiative exchange coefficients are derived. Because building height can be independently measured through a range of municipal, remote-sensing, and LiDAR-based datasets, it serves as an effective benchmark for cross-verifying NATURF's physical realism across regions. To begin, however, we present Table 2, which compares building heights across the 2015 Open DC buildings dataset and the 2015 Model America dataset as an example of the differences in the parameter input that can affect the values in the NATURF output.

Table 2. Comparison of building heights in the Model America (MAv1) dataset to those in the Open DC (ODC) dataset for Washington, DC

Statistic	Value
Num Bldgs	339,224
Max Height Diff (Abs)	60m
Min Height Diff (Abs)	0m
Mean Height Diff (Abs)	7.05m
Num Zero Height Diff	18674
Bldgs in MAv1 not in ODC	33,794
Bldgs in ODC not in MAv1	155,559

Visual inspection of the two datasets showed that neither is a perfect representation of all of the buildings in Washington, DC in 2015. For example, the Open DC dataset showed the Lincoln Memorial Reflecting Pool and some parking lots as buildings with 0 height, and the Model America dataset failed to include some of the iconic buildings near the Capitol (among others). Displaying a few statistics across the union of all buildings from both datasets highlights the differences among these two input datasets. Regardless of which dataset is used as input, NATURF makes the same urban canopy parameter calculations using the dimensions provided in the dataset. Therefore, some of the differences calculated in the following comparison will be related to differences between Model America and the local dataset and some will be related to the parameter calculation methods.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

Open Access Earth System Science Data Discussions

The following subsection, Framework for Comparative Evaluation, details the systematic multi-scale comparison between NATURF building-height fields and independent datasets for Los Angeles, Chicago, Phoenix, and Washington, DC, including the methods, datasets, metrics, and results summarized in Figures 1, 2, and 3.

4.2 Framework for Comparative Evaluation

The evaluation followed a consistent, multi-step geospatial workflow designed for cross-city comparability. For each city, we identified an independent, locally curated dataset containing either explicit or proxy estimates of building height. Each dataset was reprojected to a uniform coordinate system (EPSG:3857), and building heights were converted to a consistent metric (meters above ground level) to match NATURF's definition of area-weighted mean building height per grid cell.

Spatial aggregation was performed using centroid-in-polygon joins between the independent datasets' building footprints and the NATURF grid, allowing each grid cell's reference height to reflect all intersecting footprints. This ensured that not only height but also footprint geometry and built fraction were incorporated into the validation. For datasets with per-building attributes (e.g., heights or number of stories), we computed per-cell area-weighted statistics—mean, median, 90th percentile (p90), and 99th percentile (p99)—to capture the vertical distribution of structures in the independent datasets for best comparison to those in the NATURF datasets.

To assess scale dependence, we repeated the evaluation at three aggregation levels: the original NATURF resolution (70–93 m), and aggregated grids of approximately 210 m and 420–557 m. This multi-scale approach mitigates local positional mismatches and allows for regional-scale statistical comparison. All height fields were converted to meters and harmonized with NATURF's above-ground height definition. For example, the City of Chicago dataset provides building number of stories, which we converted using a representative floor height of 3.3 m per story. In contrast, Los Angeles County GIS provides direct HEIGHT attributes (often in feet), Phoenix LiDAR data include HGT_AGL (m) explicitly, and Washington DC LiDAR products provide $height_m = MAX_Z - MIN_Z(m)$, that is the maximum elevation of the roof structure minus the ground elevation. Table 3 lists the datasets used for validation of the NATURF data along with their description and provenance, key attributes, year of publication, and citation.

For each city, NATURF and the independent building datasets were co-registered, and their overlap regions were extracted for comparison. Area-weighted mean building height values were computed per NATURF grid cell using building polygon intersections. Cells without building coverage or containing zero or invalid heights were excluded. Statistical comparisons included bias, mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2).

Figures 1, 2, and 3 summarize the comparative evaluation. Figure 1 shows empirical cumulative distribution functions (CDFs) and height histograms for both NATURF and reference datasets across the four cities. Figure 2 presents spatial maps of mean building height (NATURF vs reference) at the original grid resolution, enabling visual inspection of urban structure representation across diverse morphologies. Figure 3 provides scatter plots at three spatial resolutions (native, 3×, and 6×), illustrating the effect of aggregation on model–data agreement.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

Table 3. Datasets Used for Validation

City	Independent Dataset	Description and Provenance	Key Attribute(s)	Year	Citation
Chicago (IL)	*City of Chicago Building Footprints*	City-maintained (data.cityofchicago.org) building footprints with "number of stories" attribute, converted to meters (3.3 m/story).	stories × 3.3 m	2023	City of Chicago Open Data Portal (2023). Building Footprints Dataset. (data.cityofchicago .org/syp8-uezg)
Washington DC	*DC GIS Building Footprints with LiDAR 3D Height Statistics*	Official DC Open Data product combining building polygons with LiDAR-derived surface elevation map metrics.	$height_m = MAX_Z - MIN_Z(m)$	2024	Government of the District of Columbia (2024). *Building Footprints with LiDAR 3D Heights.*
Los Angeles (CA)	*Los Angeles County GIS – Building Outlines (2020)*	Countywide vector dataset of building footprints with per-building height (feet) (citybes.lbl.gov/). Converted to meters and cleaned for outliers.	$HEIGHT \rightarrow meters$	2020	Los Angeles County GIS Data Portal (2020). *Building Outlines Dataset.*
Phoenix (AZ)	*LiDAR-derived 3D Building Footprints – Metro Phoenix (2014)*	LiDAR-based footprints from USGS/Woolpert (flights: Sep–Oct 2014) hosted by ASU GeoData; includes height above ground.	HGT_AGL (m)	2014	USGS/Woolpert (2014). *LiDAR-derived 3D Building Footprints for Metro Phoenix (2014).*

235 ## 4.3 Evaluation Results and Discussion

Table 4 summarizes for the four evaluated regions the comparative metrics for building height. In the case of this evaluation, NATURF-generated average building height per grid cell (or aggregated grid cell) are compared to best available local data aggregated to the same scale. Comparisons are made using the metrics of bias, mean absolute error (MAE), root mean squared error (RMSE), and R^2 value. Bias is the overall difference between model output and observations, while mean absolute error and root mean squared error are measures of the average of the absolute differences between the two. The R^2 value, which

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

ranges between 0 and 1, shows how representative the generated 100m building heights are of the buildings in the datasets used to validate them.

NATURF captures the overall spatial variability and magnitude of building heights, though discrepancies remain in certain cities, reflecting both temporal and methodological differences in the validation datasets.

Table 4. Comparative Metrics for Building Height in the Evaluated Regions. Resolution refers to the approximate length of a side of a grid cell in the UCP data sets. For each city, three resolutions are shown: native output resolution, aggregated 3x3 cells, and aggregated 6x6 cells. N refers to the number of cells included at each resolution. Metrics are applied to average building height across each grid cell.

City	Resolution	N	Bias (m)	MAE (m)	RMSE (m)	R^2
Chicago (IL)	$\sim$ 70 m	314,142	1.75	2.77	4.94	0.28
	$\sim$ 210 m	17,677	2.41	4.03	8.91	0.50
	$\sim$ 420 m	5,106	2.36	3.86	7.88	0.60
Washington DC	$\sim$ 70 m	56,413	5.16	5.96	7.82	0.11
	$\sim$ 210 m	4,863	4.59	6.13	8.51	0.33
	$\sim$ 420 m	1,452	4.27	5.51	7.59	0.41
Los Angeles (CA)	$\sim$ 93 m	417,904	-0.81	1.83	3.80	0.34
	$\sim$ 278 m	71,932	-0.93	1.77	3.70	0.31
	$\sim$ 557 m	21,867	-0.89	1.82	4.08	0.20
Phoenix (AZ)	$\sim$ 70 m	704,519	-0.76	—	2.05	0.10
	$\sim$ 210 m	44,695	-0.51	—	2.34	0.40
	$\sim$ 420 m	12,891	-0.57	—	2.20	0.49

At fine resolution (70–93 m), cell-level correlations (R^2) are moderate (0.1–0.3), increasing systematically with spatial aggregation (up to 0.6 at 400 m). This reflects both the spatial heterogeneity of building footprints and differences in acquisition year, input sources, and pre-processing methods among the reference datasets. For instance, Chicago's independent dataset estimates height from the number of stories, which introduces discretization artifacts (integer-step distributions) visible in the CDFs and histograms (Figure 1). Phoenix's LiDAR data, though highly accurate, represent 2014 conditions and thus omit more recent development. These examples underscore that no single dataset constitutes an absolute ground truth—each carries its own methodological and temporal uncertainties. Nonetheless, across all four cities, NATURF reproduces the overall structure and scale of urban morphology and building height distributions, particularly when evaluated at neighborhood (200–400 m) resolution. This consistency demonstrates that NATURF's parameterization calculation provides a reasonable and spatially coherent representation of average building height, and by extension, other dependent parameters across diverse climatic and morphological contexts.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

4.4 Initial Validation of Additional Parameters for the Arizona Urban Corridor via Implementation in WRF

In preparation for a study on urban air conditioning use (Deng et al., 2026), we compared the Arizona Corridor building density variables (parameters 1-90, Table 1) and the mean building height parameter (parameter 90) to those of the WRF lookup table for land classes defined by the 2016 National Land Cover Data set (Yang et al., 2018). For the building density values, differences were between -0.3 and 0.3 m^2/m^2 with NATURF values generally higher than those in the lookup table, especially in Tucson. Building height differences ranges from -5 to 5 m with NATURF values mostly lower than those in the lookup tables.

Two preliminary experiments were run with WRF for the month of July, 2023 one in which the lookup morphological values were used, and the other in which the NATURF UCPs mean building height, building plan area fraction, and building surface area fraction were read in for for each urban grid cell across the six-county Arizona urban corridor.

The average diurnal cycle for the month showed that while air conditioning consumption was slightly less for the simulation using the NATURF UCPs from 11pm to 9am local time, consumption was much higher during the hours of 9am to 11pm. Further similar experiments showed that for June - September diurnal averages, root mean squared error compared to observed air conditioning use during those hours was significantly less for the simulation results produced using the NATURF UCP inputs. Full details for the experiments can be found in Deng et al. (2026) and the supplementary material.

5 Conclusions and Future Work

We have presented and validated four urban canopy parameter datasets at 100m resolution for use primarily in the WRF model, but also for other studies requiring building footprints, heights and locations at this resolution. These four examples show how the NATURF model can generate these parameters at the extent of a city, a district, a county and a multi-county region. While presenting only four such datasets is not as comprehensive as some of the previously released global datasets are, they do provide readily-available products in the binary file type required by WRF for highly studied areas at a higher resolution and extent than have been previously available.

Future work will target urban canopy parameter generation for all buildings in the U.S. using updated building datasets as well as urban canopy parameter generation for buildings in neighborhoods associated with urban expansion projections (e.g., (Pandey et al.)).

6 Code availability

The code used to produce these data sets is the Neighborhood Adaptive Tissues for Urban Resilience Futures tool (NATURF) (Sweet-Breu et al., 2024). This Python workflow and its documentation can be found at https://immm-sfa.github.io/naturf/index.html . The code sits atop geopandas and the Hamilton open source framework and it calculates 132 building parameters from shapefiles containing building footprint and height information. Files are output in both binary and CSV format. The binary files are readable by the Weather Research and Forecasting (WRF) model for use with the Building Environment Parame-

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

terization (BEP) scheme. Input files used for these data files are from the open source data set, Model America Version 1 (https://doi.ccs.ornl.gov/dataset/9e968129-1ca9-5a71-93fa-24f009132a3a) featuring every U.S. building known in 2015 (New et al., 2021).

7 Data availability

Data are available by name or data DOI through the Department of Energy's MSD-LIVE (https://data.msdlive.org/) data repository. Specific links to data are included in the Data Description section, and are shown below to facilitate access.

– Title: Urban Parameters for Chicago, , Illinois, USA at 100m resolution
  DOI: 10.57931/2572297 URL: https://data.msdlive.org/records/599zb-k0458
  Citation: Vernon, C., Allen-Dumas, M., and Vahmani, P.: Urban Parameters for Chicago, Illinois, USA at 100m resolution, https://data. msdlive.org/records/599zb-k0458, 2026.

– Title: Urban Parameters Washington, DC 100m
  DOI: 10.57931/2998536 URL: https://data.msdlive.org/records/4qrgx-8ss66
  Citation: Allen-Dumas, M., Vernon, C., and Vahmani, P.: Urban Parameters Washington, DC 100m, https://data.msdlive.org/records/4qrgx-8ss66, 2026.

– Title: Urban Parameters Los Angeles County 100m version 2
  DOI: 10.57931/3000523 URL: https://data.msdlive.org/records/mp5mx-ztp89
  Citation: Sweet-Breu, L., Allen-Dumas, M., Vernon, C., and Vahmani, P.: Urban Parameters Los Angeles County 100m version 2, https://data.msdlive. org/records/mp5mx-ztp89, 2026.

– Title: Urban Parameters Arizona Urban Corridor 100m
  DOI: 10.57931/2998499 URL: ttps://data.msdlive.org/records/va0qd-qsn15
  Citation: Allen-Dumas, M., Vernon, C., and Vahmani, P.: Urban Parameters Arizona Urban Corridor 100m, https://data.msdlive.org/records/ va0qd-qsn15, 2026.

8 Usage Notes

Usage of the binary data with associated index in WRF Preprocessing System WPS follows the same procedure as that for using the NUDAPT data Glotfelty et al. (2013). First, the index and the binary file should be placed in their own directory (named something like "Chicago" or "AZ_Corridor") within the WPS_GEOG directory. Next, the URB_PARAM section of the GEOGRID.TBL must be edited to priority = 1 and optional = no. Additionally, the rel_path=NUDAPT_44 line must point instead to the NATURF-generated data set as, e.g., rel_path=Chicago. Additionally, if these parameters/variables are not already set in the BEP code portion of WRF (module_sf_bep.F), they must be set to these values:

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

- Set the number of vertical layers in each urban grid cell (nz_um) to 15 or larger in the BEP code before compiling WRF.
- Set the num_urban_layers in the WRF namelist.input to 15 to accommodate the urban parameters.
- A variable called num_urban_hi must be added to the namelist.input for the dimension of the distribution of building height. This value must be set to 15.

To use the urban parameters in CSV format for morphological characterizations, e.g., morphology parameters-based neighborhood delineation, the "building_geometry" column containing WKT text strings can be used to spatialize the urban parameters by creating geometry objects from the text strings for use in a GIS software or with geospatial packages (e.g., geopandas in Python). Once spatialized, users can leverage spatially constrained clustering algorithms Guo (2008) to cluster buildings based on all or select urban parameters employing dimensionality reduction methods, if necessary.

Finally, if these data sets are used for further research of any type, this paper should be cited.

Author contributions. MRAD conceived the paper and the data generation methods, generated some of the data sets, and wrote several sections of the paper. PV designed and executed the technical validation, generated the corresponding figures and tables, and wrote the corresponding text. BP contributed to the background section and usage notes. CV generated some of the data sets, developed the accompanying meta-repository, troubleshot the data generation code, and reviewed the document. LSB generated one of the datasets and contributed to the data generation code. ER contributed to the data generation code.

Competing interests. The authors declare no competing interests.

Acknowledgements. This research was sponsored by the DOE Office of Science as a part of the research in Multi-Sector Dynamics within the Earth and Environmental System Modeling Program. It used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the DOE Office of Science. There is no grant number associated with the project, but more information can be found here: https://im3.pnnl.gov/ . The authors especially appreciate the encouragement of Integrated Multiscale Multisector Modeling (IM3) Principal Investigator, Jennie Rice, of Pacific Northwest National Laboratory, to generate the data sets described here.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

References

Allen-Dumas, M., Vernon, C., and Vahmani, P.: Urban Parameters Arizona Urban Corridor 100m, https://data.msdlive.org/records/va0qd-qsn15, 2026a.

Allen-Dumas, M., Vernon, C., and Vahmani, P.: Urban Parameters Washington, DC 100m, https://data.msdlive.org/records/4qrgx-8ss66, 2026b.

Allen-Dumas, M. R., McManamay, R. A., Pandey, B., Brelsford, C. M., New, J. R., Li, H., Li, F., and Singh, V.: Multi-Model Methodology for Urban Development Scenarios and Evaluation of Future City-Scale Building Energy Efficiency, in Prep.

Allen-Dumas, M. R., Rose, A. N., New, J. R., Omitaomu, O. A., Yuan, J., Branstetter, M. L., Sylvester, L. M., Seals, M. B., Carvalhaes, T. M., Adams, M. B., et al.: Impacts of the morphology of new neighborhoods on microclimate and building energy, Renew. Sustain. Energy Rev., 133, 110 030, 2020.

Allen-Dumas, M. R., Sweet-Breu, L. T., Brelsford, C. M., Wang, L., New, J. R., and Bass, B. C.: Sensitivity of mesoscale modeling to urban morphological feature inputs and implications for characterizing urban sustainability, npj Urban Sustainability, 4, 49, 2024.

Best, M.: Progress towards better weather forecasts for city dwellers: from short range to climate change, Theor. Appl. Climatol., 84, 47–55, 2006.

Brelsford, C., Jones, A., Pandey, B., Vahmani, P., Allen-Dumas, M., Rastogi, D., Sparks, K., Bukovsky, M., Dronova, I., Hong, T., et al.: Cities are concentrators of complex, multisectoral interactions within the human-Earth system, Earth's Future, 12, e2024EF004 481, 2024.

Burian, S., Han, W., and Brown, M.: Morphological analyses using 3D building databases: Houston, Texas, Department of Civil and Environmental Engineering, University of Utah, 2003.

Chakraborty, T., Hsu, A., Manya, D., and Sheriff, G.: A spatially explicit surface urban heat island database for the United States: Characterization, uncertainties, and possible applications, ISPRS Journal of Photogrammetry and Remote Sensing, 168, 74–88, 2020.

Cheng, Y., Zhao, L., Chakraborty, T., Oleson, K., Demuzere, M., Liu, X., Che, Y., Liao, W., Zhou, Y., and Li, X.: U-Surf: a global 1 km spatially continuous urban surface property dataset for kilometer-scale urban-resolving Earth system modeling, Earth System Science Data Discussions, 2024, 1–38, 2024.

Ching, J., Brown, M., Burian, S., Chen, F., Cionco, R., Hanna, A., Hultgren, T., McPherson, T., Sailor, D., Taha, H., et al.: National Urban Database and Access Portal Tool, Bull. Am. Meteorol. Soc., 90, 1157–1168, 2009.

Deng, X., Cohen, J. J., Rodriguez, J., Svoma, B. M., Gonzales, C., and Georgescu, M.: A hybrid physics-machine learning framework for improving simulations of urban air-conditioning electricity consumption, Energy and Buildings, p. 117982, 2026.

Dirksen, M., Ronda, R., Theeuwes, N., and Pagani, G.: Sky view factor calculations and its application in urban heat island studies, Urban Climate, 30, 100 498, 2019.

Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., Marconcini, M., Roth, A., and Zeidler, J.: World Settlement Footprint 3D-A first three-dimensional survey of the global building stock, Remote sensing of environment, 270, 112 877, 2022.

Fleischmann, M., Samardzhiev, K., Brázdová, A., Dančejová, D., and Winkler, L.: The Hierarchical Morphotope Classification: A Theory-Driven Framework for Large-Scale Analysis of Built Form, https://doi.org/10.48550/arXiv.2509.10083, 2025.

Gabrian, I. S., Dinicila, S., Dumitrescu, A., Velea, L., and Cheval, S.: Simulating the Urban Heat Island during heat wave events using WRF urban parameterizations: a case study for Bucharest (Romania), Geomatics, Natural Hazards and Risk, 16, 2549 490, 2025.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

García, D. H.: Analysis of urban heat island and heat waves using Sentinel-3 images: a study of Andalusian Cities in Spain, Earth Systems and Environment, 6, 199–219, 2022.

Glotfelty, T., Tewari, M., Sampson, K., Duda, M., Chen, F., and Ching, J.: NUDAPT 44 documentation, National Center for Atmospheric Research Research Applications Laboratory Doc, 9, 2013.

Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., and Briggs, J. M.: Global change and the ecology of cities, Science, 319, 756–760, 2008.

Grimmond, C. and Oke, T. R.: Aerodynamic properties of urban areas derived from analysis of surface form, Journal of Applied Meteorology and Climatology, 38, 1262–1292, 1999.

Guo, D.: Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP), International Journal of Geographical Information Science, 22, 801–823, https://doi.org/10.1080/13658810701674970, 2008.

Kesseli, J. E.: The climates of California according to the Köppen classification, Geographical Review, pp. 476–480, 1942.

Li, D., Sun, T., Yang, J., Zhang, N., Vahmani, P., and Jones, A.: Structural uncertainty in the sensitivity of urban temperatures to anthropogenic heat flux, Journal of Advances in Modeling Earth Systems, 16, e2024MS004 431, 2024.

Liao, W., Li, Y., Liu, X., Wang, Y., Che, Y., Shao, L., Chen, G., Yuan, H., Zhang, N., and Chen, F.: GloUCP: a global 1 km spatially continuous urban canopy parameters for the WRF model, Earth System Science Data Discussions, 2024, 1–21, 2024.

Luo, X., Vahmani, P., Hong, T., and Jones, A.: City-scale building anthropogenic heating during heat waves, Atmosphere, 11, 1206, 2020.

Macdonald, R., Griffiths, R., and Hall, D.: An improved method for the estimation of surface roughness of obstacle arrays, Atmospheric Environment, 32, 1857–1864, 1998.

Marconcini, M., Marmanis, D., Esch, T., and Felbier, A.: A novel method for building height estmation using TanDEM-X data, in: 2014 IEEE Geoscience and Remote Sensing Symposium, pp. 4804–4807, IEEE, 2014.

Martilli, A.: An idealized study of city structure, urban climate, energy consumption, and air quality, Urban Clim., 10, 430–446, 2014.

Means, J.: Roadside Geology of Maryland, Delaware, and Washington, DC, Mountain Press Publishing Company, 2010.

Microsoft: Microsoft building footprints, https://www.microsoft.com/en-us/maps/building-footprints, 2020.

Mohamed, O. Y. and Zahidi, I.: Artificial intelligence for predicting urban heat island effect and optimising land use/land cover for mitigation: Prospects and recent advancements, Urban Climate, 55, 101 976, 2024.

Müller-Hansen, F., Schlüter, M., Mäs, M., Donges, J. F., Kolb, J. J., Thonicke, K., and Heitzig, J.: Towards representing human behavior and decision making in Earth system models–an overview of techniques and approaches, Earth System Dynamics, 8, 977–1007, 2017.

New, J., Bass, B., Berres, A., Clinton, N., Adams, M., Li, F., Stubbings, A., and Chowdhury, S.: Model America: Data and models for every US building, Tech. rep., Environmental System Science Data Infrastructure for a Virtual Ecosystem . . . , 2021.

New, J., Adams, M., Bass, B., Clinton, N., Berres, A., Garrison, E., Guo, T., and Allen-Dumas, M.: Model America: A crude energy model and data for nearly every US building, Available at SSRN 4220628, 2022.

Oke, T. R.: The energetic basis of the urban heat island, Q. J. R. Meteorol. Soc., 108, 1–24, https://doi.org/10.1002/qj.49710845502, 1982.

Oke, T. R.: Street design and urban canopy layer climate, Energ. Buildings, 11, 103–113, 1988.

Oleson, K. W., Bonan, G. B., Feddema, J., Vertenstein, M., and Kluzek, E.: Technical description of an urban parameterization for the Community Land Model (CLMU), Tech. rep., NCAR, Boulder, 2010.

Pandey, B., Allen-Dumas, M. R., Singh, V. K., New, J. R., and Vernon, C.: Evaluating Generative Adversarial Networks for Projecting Urban Morphologies, in Prep.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

Pappaccogli, G., Giovannini, L., Zardi, D., and Martilli, A.: Sensitivity analysis of urban microclimatic conditions and building energy consumption on urban parameters by means of idealized numerical simulations, Urban Clim., 34, 1–17, 2020.

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrology and earth system sciences, 11, 1633–1644, 2007.

Raupach, M.: Simplified expressions for vegetation roughness length and zero-plane displacement as functions of canopy height and area index, Boundary-layer Meteorology, 71, 211–216, 1994.

Schlüter, M., Brelsford, C., Ferraro, P. J., Orach, K., Qiu, M., and Smith, M. D.: Unraveling complex causal processes that affect sustainability requires more integration between empirical and modeling approaches, Proceedings of the National Academy of Sciences, 121, e2215676 120, 2023.

Seto, K. C. and Pandey, B.: Urban Land Use: Central to Building a Sustainable Future, One Earth, 1, 168–170, 2019.

Seto, K. C., Reenberg, A., Boone, C. G., Fragkias, M., Haase, D., Langanke, T., Marcotullio, P., Munroe, D. K., Olah, B., and Simon, D.: Urban land teleconnections and sustainability, Proceedings of the National Academy of Sciences, 109, 7687–7692, 2012.

Shahmohamadi, P., Che-Ani, A., Maulud, K., Tawil, N., and Abdullah, N.: The impact of anthropogenic heat on formation of urban heat island and energy consumption balance, Urban Stud. Res., 2011, 2011.

Simpson, A., Ellmauer-Klambauer, A., Dengler, S., Staudinger, M., Zuvela-Aloise, M., and de Wit, R.: Analysis of Heat Waves and Urban Heat Island Effects in Central European Cities and Implications for Urban Planning.

Sweet-Breu, L., Allen-Dumas, M., Vernon, C., and Vahmani, P.: Urban Parameters Los Angeles County 100m version 2, https://data.msdlive.org/records/mp5mx-ztp89, 2026.

Sweet-Breu, L. T., Rexer, E., Vernon, C. R., Allen-Dumas, M. R., and Krawczyk, S.: naturf: a package for generating urban parameters for numerical weather modeling, Journal of Open Source Software, 9, 2024.

Tadono, T., Nagai, H., Ishida, H., Oda, F., Naito, S., Minakawa, K., and Iwamoto, H.: Generation of the 30 m-mesh global digital surface model by ALOS PRISM, The international archives of the photogrammetry, remote sensing and spatial information sciences, 41, 157–162, 2016.

U.S. Census Bureau: tl$_2018_u s_c ounty$”, *TIGER/Line Shapefiles*, , 2018.

Vernon, C., Allen-Dumas, M., and Vahmani, P.: Urban Parameters for Chicago, Illinois, USA at 100m resolution, https://data.msdlive.org/records/599zb-k0458, 2026.

Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, M. J., Williams Jr, C. N., Fenimore, C., Gleason, K., and Arndt, D.: Improved historical temperature and precipitation time series for US climate divisions, Journal of Applied Meteorology and Climatology, 53, 1232–1251, 2014.

Wang, N., Vlachokostas, A., Borkum, M., Bergmann, H., and Zaleski, S.: Unique Building Identifier: A natural key for building data matching and its energy applications, Energy and Buildings, 184, 230–241, 2019.

Western Regional Climate Center: Climate of Arizona, https://wrcc.dri.edu/climate/narrative_az.php#:~:text=The%20desert%20valleys%20of%20southwestern,the%20Lower%20Colorado%20River%20Valley., 2016.

Winegar, D.: Highroad Guide to the Chesapeake Bay, Carolina Wren Press, 1998.

https://doi.org/10.5194/essd-2026-710
Preprint. Discussion started: 9 September 2026

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., Case, A., Costello, C., Dewitz, J., Fry, J., et al.: A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies, ISPRS journal of photogrammetry and remote sensing, 146, 108–123, 2018.