ENVIRONMENTAL RESEARCH
LETTERS

LETTER • **OPEN ACCESS**

Detailed height mapping of trees and buildings (HiTAB) in Chicago and its implications to urban climate studies

To cite this article: Peiyuan Li and Ashish Sharma 2024 *Environ. Res. Lett.* **19** 094013

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Detailed height mapping of trees and buildings (HiTAB) in Chicago and its implications to urban climate studies

Peiyuan Li[1,*] and **Ashish Sharma**[1,2,3,*]

[1] Discovery Partners Institute, University of Illinois System, Chicago, IL 60606, United States of America
[2] Department of Climate, Meteorology & Atmospheric Sciences, University of Illinois at Urbana-Champaign, Champaign, IL 61820, United States of America
[3] Environmental Science Division, Argonne National Laboratory, Lemont, IL 60439, United States of America
[*] Authors to whom any correspondence should be addressed.

E-mail: peiyuanl@uillinois.edu and sharmaa@illinois.edu

Keywords: building height, tree height, LiDAR, high-resolution, urban climate models

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
6 January 2024

REVISED
4 July 2024

ACCEPTED FOR PUBLICATION
22 July 2024

PUBLISHED
6 August 2024

Abstract

The vertical dimensions of urban morphology, specifically the heights of trees and buildings, exert significant influence on wind flow fields in urban street canyons and the thermal environment of the urban fabric, subsequently affecting the microclimate, noise levels, and air quality. Despite their importance, these critical attributes are less commonly available and rarely utilized in urban climate models compared to planar land use and land cover data. In this study, we explicitly mapped the height of trees and buildings (HiTAB) across the city of Chicago at 1 m spatial resolution using a data fusion approach. This approach integrates high-precision light detection and ranging (LiDAR) cloud point data, building footprint inventory, and multi-band satellite images. Specifically, the digital terrain and surface models were first created from the LiDAR dataset to calculate the height of surface objects, while the rest of the datasets were used to delineate trees and buildings. We validated the derived height information against the existing building database in downtown Chicago and the Meter-scale Urban Land Cover map from the Environmental Protection Agency, respectively. The co-investigation on trees and building heights offers a valuable initiative in the effort to inform urban land surface parameterizations using real-world data. Given their high spatial resolution, the height maps can be adopted in physical-based and data-driven urban models to achieve higher resolution and accuracy while lowering uncertainties. Moreover, our method can be extended to other urban regions, benefiting from the growing availability of high-resolution urban informatics globally. Collectively, these datasets can substantially contribute to future studies on hyper-local weather dynamics, urban heterogeneity, morphology, and planning, providing a more comprehensive understanding of urban environments.

1. Introduction

Over two-thirds of the global population will be urban dwellers by 2050 (United Nations 2018). In contrast to their rural and natural counterparts, urban landscapes are characterized by a high degree of heterogeneity and compactness. This results in a mosaic of fragmented green (e.g. parks and street trees) and grey spaces (e.g. pavement and buildings), along with a noticeable expansion in vertical dimensions such as skyscrapers. These elements create canyon-like geometry of the urban fabric, where tall buildings reshape wind flow patterns and redirect short- and long-wave radiations through shading, reflection, and trapping mechanisms (Oke 1982). Additionally, the presence of street trees further increases the complexity of the in-canyon flow dynamics, largely influencing the near-surface microclimate at the pedestrian level (Giometto *et al* 2017, Krayenhoff *et al* 2020).

The last decades have witnessed numerous modeling efforts and advances to incorporate precise urban morphology of the built environment to reduce bias and errors in the simulation (Ryu

et al 2015, Krayenhoff *et al* 2020). Among existing approaches, urban computational fluid dynamics (CFD) models offer detailed insights into the flow field within street canyons and reproduce a comprehensive suite of meteorological and environmental variables (Toparlar *et al* 2017, Mirzaei 2021). More recently, machine learning (ML) has become an emerging technique to conduct high-fidelity simulation of complex urban environments (Meyer *et al* 2022, Middel *et al* 2022, Wang *et al* 2023). Many studies have proved the high efficiency of ML methods in resolving urban scenes at an unprecedented granularity and spatial coverage compared to process-based models (Li and Sharma 2024, Yu *et al* 2024). Nevertheless, physical- and data-driven methods have distinct underlying principles and applications, both rely on accurate urban morphological data, which includes measurements of urban fabric in three dimensions (Stewart and Oke 2012). In particular, height information is often less accessible at city or regional scales compared to two-dimensional land use and land cover data. Data scarcity in vertical dimensions not only impedes accurate urban representation but also constantly constrains advancements in high-resolution urban modeling.

It is also worth noting that tree information is even more rare compared to building height data due to their smaller footprints and highly fragmented spatial distributions in the built environment. Yet, urban trees have drawn much attention from urban climate research communities because of their active role as a nature-based solution to multiple urban challenges, such as heat mitigation (Schwaab *et al* 2021), flood prevention (Center for Watershed Protection 2017), carbon reduction (Li *et al* 2023, 2024), pollution uptake (McDonald *et al* 2007, Hirabayashi and Nowak 2016), etc. emphasizing the importance of precise tree information for the understanding of hydrometeorological dynamics in built environments. Besides the need for research purposes, it is critical for cities to have accurate tree inventory for urban design, planning, and management (Woodward *et al* 2023). In particular, tree planting has been widely recommended in city, regional, and national climate action plans (CAPs). For example, the city of Chicago has set an ambitious goal to plant 75 000 trees by 2026, as outlined in its 2022 CAP (City of Chicago 2022). The local agencies, such as the Morton Arboretum, have been instrumental in performing tree censuses since 2010, aiming to monitor the condition of urban green canopy and evaluate the planting goals (The Morton Arboretum 2020). A detailed map of tree will create synergy along with the considerable effort demonstrated in the Chicago tree census and may help the city expand the tree database to include privately owned canopies.

To this end, our study presents maps of buildings and tree heights over the city of Chicago at 1 m spatial resolution. This height dataset height of trees and buildings (HiTAB hereafter) is derived by integrating the archived high-precision light detection and ranging (LiDAR) cloud point data, building footprint inventory, and multi-band satellite images. The co-investigation of tree and building height will benefit not only the advances of urban climate research, but also help a wide range of applications that require the 3D description of the urban fabric, such as the estimation of rooftop solar power production (Tooke *et al* 2012), urban ecology (Casalegno *et al* 2017), property valuation (Kara *et al* 2020), etc. Moreover, we investigate the emerging issues in data fusion processes and unravel the uncertainties that arise when processing geospatial datasets at meter-level in an urban environment. We believe the height dataset will be one of the critical components in urban informatics to foster the research and decision-making of the Chicago region toward a sustainable future.

2. Data and method

Figure 1(a) summarizes the processing workflow used in this study from data sources to the final products. In this section, we describe the details of the datasets and digest their uses in HiTAB.

2.1. LiDAR point cloud

The LiDAR cloud point dataset has been widely used to derive canopy structures for ecological studies (Guo *et al* 2020). The specific LiDAR dataset used in this study is available from the Illinois Height Modernization project (ILHMP—Illinois Height Modernization 2021). This dataset records the spatial coordinates (latitude, longitude, and elevation) of every LiDAR point with a spacing of 0.35 m. Each LiDAR point is assigned with one classification of ground, low vegetation, medium vegetation, high vegetation, buildings, water, noise, bridge deck, and others. Figure 1(b) shows the un-processed LiDAR point cloud at an exemplary street block in Chicago (41.8717°N, 87.6306°W). This LiDAR data serves as the source for height information. Specifically, we investigate the LiDAR points within a 1 m by 1 m sliding window. The points with 'ground' classification are used to derive the digital terrain model (DTM, i.e. bare surface); while the points with maximum elevation (i.e. first return) represent the digital surface model (DSM). The height of surface objects, shown as H_0 in figure 1(a), is then calculated from the difference between the DSM and DTM (figure 1(b)).

Note that in this step, we do not distinguish vegetation from buildings. The LiDAR dataset used here was originally designed for hydrological studies

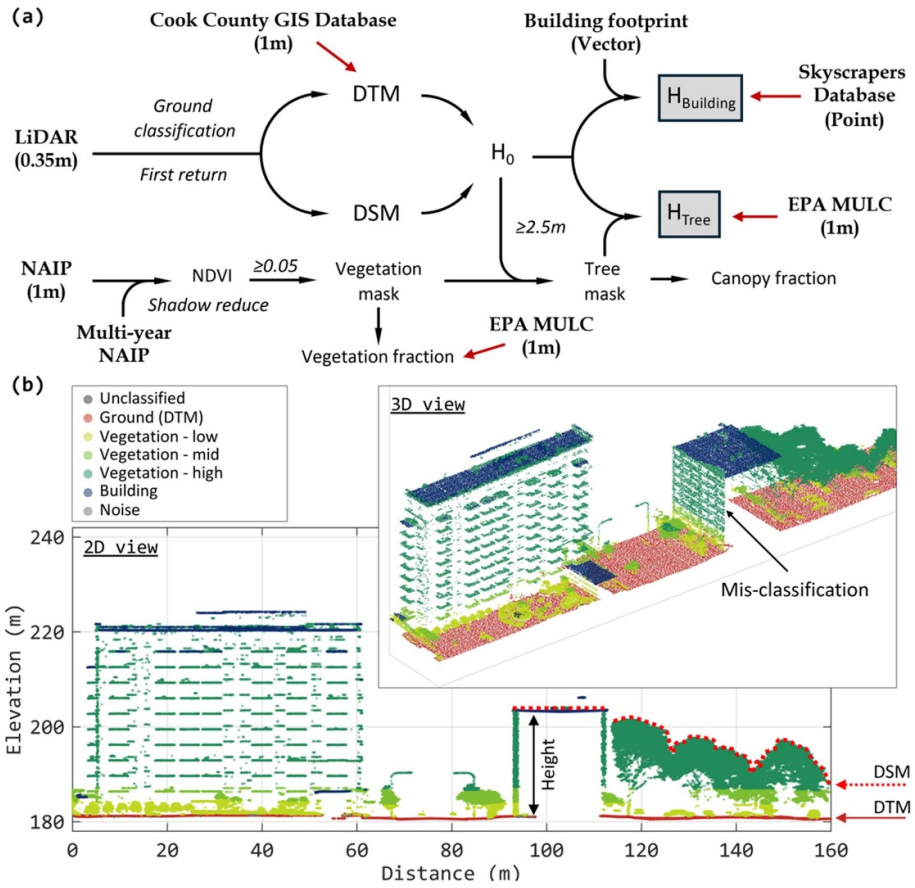

Figure 1. (a) A processing workflow showing data sources and outcomes of this study; and (b) an illustration of LiDAR point cloud data with un-processed classification over a street block in Chicago (41.8717°N, 87.6306°W). Red arrows in (a) indicate the validation data to the intermediate and final variables. DTM—digital terrain model. DSM—digital surface model. A summary list of the datasets used in this study can be found in table S1.

of natural ecosystems. Therefore, the classification attributes can be problematic once applied to urbanized areas. We find that the classes 'ground' and 'buildings' (representing the elevation of roofs) are generally accurate, while most of the building facets are misclassified as vegetation (figure 1(b)). It is necessary to use additional and independent datasets to delineate tree canopies from other high urban structures. In this case, we reconcile the height map with a vectorized building footprint dataset (section 2.2) and normalized difference vegetation index (NDVI) to differentiate trees and buildings. In addition, the LiDAR scan has an 'intensity' band, referring to the signal strength of the LiDAR returns. The distinctive characteristics of LiDAR intensity between plant leaves and pavements have been utilized in many land cover classification studies (Kashani *et al* 2015, Morsy *et al* 2017, Huo *et al* 2018). This characteristic is also used to cross-validate the delineation of tree canopies. More information can be found in Text S1.

The LiDAR scan over the city of Chicago was collected during the late spring of 2017, which is not temporally synchronized with the other datasets used in this study. Because of their high precision, LiDAR surveys over large areas can be labor-intensive and time-consuming. Therefore, they are generally available once every few years. This frequency is much lower than the satellite imageries, which are more commonly available at sub-seasonal, monthly, or even daily intervals. This shortcoming makes LiDAR impractical to reflect objects with strong seasonality, such as tree leaves. However, its high precision can ensure that the tree canopy is clearly distinguished from the surrounding buildings for the purpose of this study.

2.2. Land use and building footprint inventory

We select two sets of land use data over the city of Chicago to delineate the street and building boundaries. The 2018 parcel-level land use inventory (LUI) for the city of Chicago can be found at the Chicago Metropolitan Agency for Planning (CMAP) in vector format (www.cmap.illinois.gov/data/land-use/inventory). The LUI classifies land use into 10 major and 56 minor categories. The street blocks, roads, urban form, and fabric boundaries are precisely outlined by polygons, which can be derived to raster masks at extremely high resolution. We use

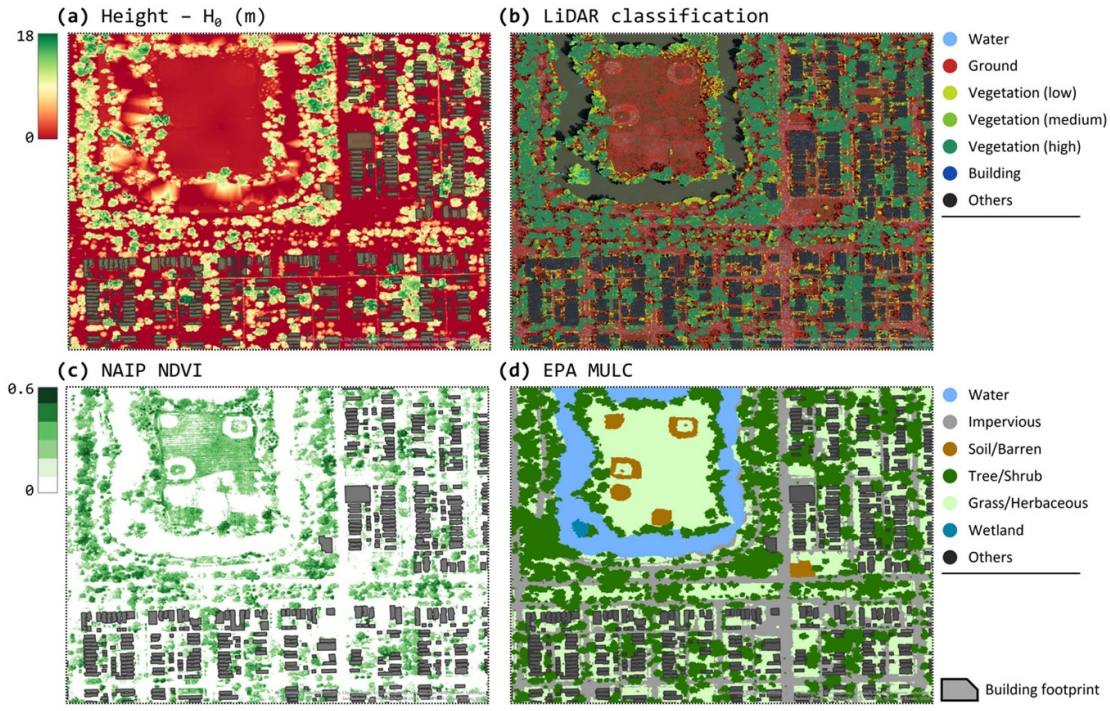

Figure 2. Data layers used in this study over an exemplary street block near Sherman Park in Chicago (41.7952°N, 87.6562°W). (a) Heights of surface objects; (b) LiDAR un-processed classifications; (c) NDVI values derived from NAIP; and (d) land cover classification from US environmental protection agency (EPA) meter-scale urban land cover (MULC) dataset.

this dataset to derive the mask of roads and streets. An intersection of masks associated with the 'ground' classification of LiDAR points can provide the most reliable bare ground elevation for DTM.

In addition, the building footprint (BF) data from the Chicago data portal provides even more precise boundaries of the architectures in addition to the parcel-level LUI (figure S1). Figure 2(a) shows the height of the surface objects (H_0) at an exemplary street block with a building footprint overlayed. The building footprint differentiates the land as vertical structures (e.g. buildings) and impervious surfaces (e.g. roads or parking lots), allowing us to calculate the building heights (H_B) accurately. We use this dataset to derive the building mask, and reclassify LiDAR points within the mask to 'buildings' (figure 2(b)). This partially corrects the misclassification of the building facets as mentioned in section 2.1. When there is an overlapping between tree canopy and building, we further use NDVI data to detect and differentiate them.

2.3. Multispectral satellite imageries

NDVI is a commonly used vegetation index to reflect biomass density, calculated as

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}}$$

where NIR and Red are reflectances in the near-infrared and red bands from the satellite imagery, respectively. The value of NDVI ranges from $[-1, 1]$, with greater positive values representing denser biomass. We use the 4-band (RGB and NIR) satellite imagery at 1 m resolution from the National Agriculture Imagery Program (NAIP 2023) to calculate NDVI (figure 2(c)). The NAIP dataset has been available on a bi-annual basis during summertime since 2007 over the Chicago region. We primarily use the data from September 2017 to better align with the time frame of the LiDAR data. This 1 m NDVI data provides the ground truth of the presence of vegetation. It is used to derive the high-resolution vegetation mask and resolve the LiDAR misclassification issue.

We further note that tall buildings can cast shadows on street trees, which leads to a significant underestimate of the NDVI in the shaded areas. This shadow phenomenon has been widely seen across NDVI products at various spatial resolutions (Burgess *et al* 1995, Aboutalebi *et al* 2018, Yang *et al* 2022) but becomes more apparent at high spatial resolution and in urban settings. Figure S2 shows an example of this phenomenon by comparing NAIP products in 2017 and 2021. These data were collected with different sun angles. Though NAIP regulates the collection time within 2 h from the local noon with a minimum sun angle greater than 30 degrees (NAIP 2023), shadows are still one of the common issues found in NAIP-based products (Ritz *et al* 2022). In this case, the accurate mapping of NDVI needs to use data from more than one-time frame, preferably from two frames before and after local noon with

high solar angles. Considering the long revisit time (2 years) of the NAIP product, it becomes impractical to create a shadow-free and time-aligned NDVI dataset simultaneously. To mitigate this issue, we combine NDVI maps derived from NAIP 2017 and 2021, which were captured in the afternoon and morning, respectively. Using the combined NDVI map, the tree mask is defined as where NDVI is greater than 0.05 and height is greater than 2.5 m (figure 1(a)). The tree mask represents the planar (2D) coverage of the tree canopy.

2.4. EPA meter-scale urban land cover (MULC) land use classification

The meter-scale urban land cover (MULC) is a dataset developed by the US environmental protection agency (EPA) for EnviroAtlas (Pilant *et al* 2020, figure 2(d)). In each of the 30 cities it covers, the land surface is classified into 10 categories, including impervious surfaces, trees, shrubs, grass, water, crops, etc. As this dataset is specific to cities, it has a good representation of the urban landscapes. We use this data as an independent data source to validate the tree mask that we derived in this study. Compared to MULC, our results indicate good agreement with a true positive rate of 0.86 and true negative rate of 0.96. The precision, F1-score, and Kappa coefficient of our classification are 0.75, 0.80, and 0.77, respectively. The discrepancies, however, are primarily caused by the mismatched data year and the classification regime of MULC. MULC used older source datasets as early as 2006 (Pilant *et al* 2020) and has only two classes of vegetation in the Chicago region: trees and grass/herbaceous. The simple scheme prefers to classify tall and dense vegetation as 'trees' while keeping others as grassland. This potentially underestimates the canopy coverage compared to the real-world situation in 2017. Nevertheless, MULC is the only dataset with a comparable resolution and spatial coverage from a reliable source.

3. Results and discussion

3.1. Statistics of building height in Chicago

The height of the surface objects (H_0) is calculated as the difference between DSM and DTM. Therefore, its accuracy largely depends on the accuracy of ground elevations. We find that satellite-based digital elevation models tend to misestimate the surface elevation as the mean elevation of surface objects, especially over high-rise downtown areas (figure. S3); while LiDAR-based DTMs can provide accurate bare earth elevation. In the comparison, our DTM aligns well with the data published in the Cook County GIS Database (CCGD, figure S3(b), https://hub-cookcountyil.opendata.arcgis.com/) and US Geological Survey 3D Elevation Program (USGS 3DEP, figure S3(d)). Both CCGD and 3DEP are LiDAR-based. The root mean squared errors (RMSE) of this study and 3DEP are 0.98 m and 1.40 m compared to CCGD, respectively (table S2). Nevertheless, defining 'ground elevation' over urban terrain is occasionally difficult, such as downtown Chicago's complex network of rivers, underground parking structures, bridges, and elevated light rails. This uncertainty is primarily responsible for the biases between DTM products.

The height map (H_0) is subsequently delineated to building height (H_B) and tree height (H_T) using building footprint (section 2.2) and tree mask (section 2.3). Figure 3 shows the height maps over two exemplary street blocks with trees and buildings differentiated. Since there is no known spatial gridded building height dataset to validate the height map, we compare our building height to the height of skyscrapers (>500 ft or 150 m) built between 1950 and 2017 in downtown Chicago, which has well-documented height information. Our building height shows a mean bias error of −0.03 m and RMSE of 4.88 m over the tallest 67 skyscrapers (figure S4). The relative error is within 2% of the exact building height. The minor discrepancy is caused by the variance of their roof design. The difference can be notable for the buildings having a spire-shaped roof or an antenna on top. Some skyscrapers reported architectural heights, while others reported the full height with the length of antennas. For residential buildings with sloped roofs, our 1 m raster map can reflect the height variance over each individual structure, thus is considered more accurate compared to the building inventory where only a single value is provided.

According to CMAP LUI in 2018, residential use dominates the land types in Chicago. The distribution of residential building height has noticeable peaks at around 4.4 m, 7.8 m, and 11.7 m, corresponding to the building types of one-story single-family homes, two or three-story homes, and multi-story apartments/flats (figure 4(a)). This characteristic becomes less apparent for non-residential buildings, which exhibit a smoother but skewed distribution with a median of 7.8 m and a mean of 11.3 m ('others' in figure 4(a)). It is noteworthy that the exact height measurements of low-rise buildings are not commonly available. However, it is one of the critical attributes of urban form. Considering the large portion of residential areas in cities, explicit mapping and statistics over these low-rise buildings have profound implications for urban climate modeling studies, which we will discuss in section 3.3.

3.2. Statistics of tree height in Chicago

The high accuracy of building height demonstrates the reliability of our methodology. The derivation of tree height follows a similar procedure but with the building mask replaced by a tree mask reconciling

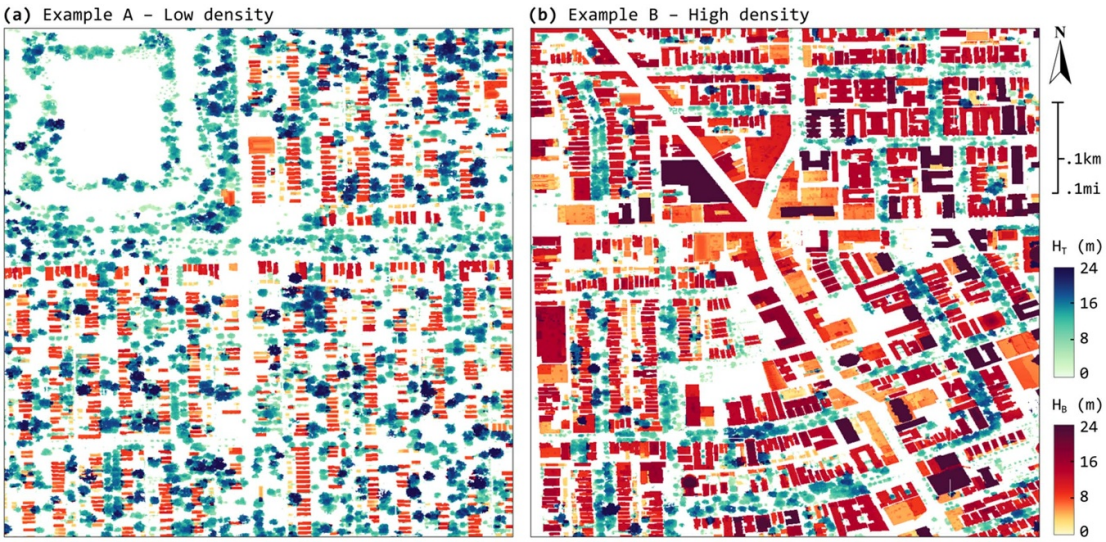

Figure 3. Height maps over two exemplary street blocks with (a) low development intensity (41.7952°N, 87.6562°W), and (b) high development intensity (41.9329°N, 87.6449°W).

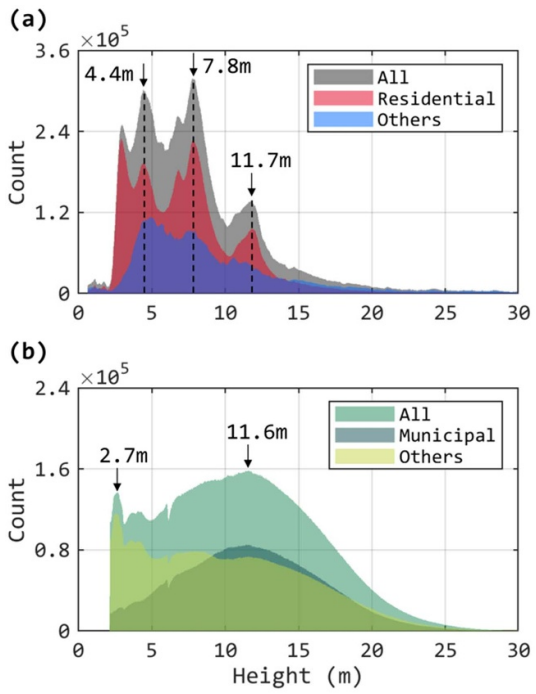

Figure 4. The height distributions of (a) buildings and (b) trees in the city of Chicago. 'Others' in (a) represents buildings that are not located in the residential land cover in CMAP LUI. 'Municipal' in (b) refers to trees that are located along roads/streets and in city parks. 'Others' in (b) refers to trees that are not 'Municipal'.

with NDVI and H_0. The tree mask indicates that the city of Chicago (excluding O'Hare International Airport) has a canopy coverage of 16.7% in 2017 (figure 5(a)). This number is lower than the estimates from University of Vermont Spatial Analytics Lab (UVM, 19.3%, O'Neil-Dunne *et al* 2014, Darling 2023), while being significantly higher than the commonly used tree canopy coverage (TCC) map from national land cover database (NLCD) 2017 at 30 m resolution (Housman *et al* 2021, ∼7%, figure 5(b)). From our investigation (Text S2), we find three major reasons for the discrepancies. (i) Dataset with coarser resolution (i.e. the, NLCD) tends to underestimate the coverage due to the small footprint of trees in urban environments (Pourpeikari Heris *et al* 2022); their signals are usually diluted when averaging into a relatively large pixel. (ii) At a hyper-local scale (∼1 m), uncertainties will emerge from the classification scheme due to the similarity of spectral signatures between tree leaves and some building materials. In this case, the canopy coverage is sensitive to the classification criteria and the availability of information (Text S2 and figure ST2). The presence of green roofs and canopies over low roofs will further complicate the situation (Text S3 and figure ST4). (iii) In addition, the phenology of deciduous trees affects the canopy size, density, and NDVI, thus affecting quantification as well. Nevertheless, the high-resolution dataset will certainly be more accurate. These phenomena are worth in-depth investigations.

The focus of HiTAB is the tree height (H_T) rather than the canopy coverage. The statistical distribution of trees in Chicago shows a very different shape compared to the distribution of buildings, exhibiting one major peak around 11.6 m and a local peak at 2.7 m (figure 4(b)). The number of tall trees (>11.6 m) gradually declines with the increase in tree height. Height distributions notably differ between municipal trees (i.e. public parks and street trees) and the rest. The former follows a normal distribution $N \sim (11.9, 4.8)$, yielding the major peak. In contrast, the rest of the trees have an even height distribution under 13 m and naturally decline thereafter (figure 4(b)).

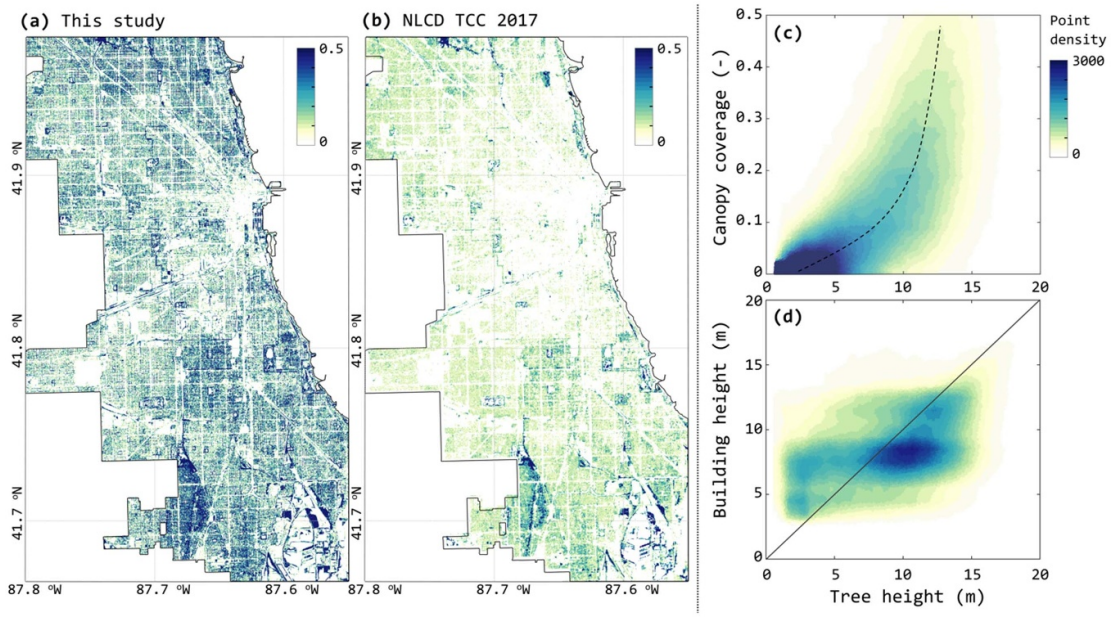

Figure 5. Tree canopy coverage (TCC) at 1 m spatial resolution from (a) this study; and 30 m spatial resolution from (b) NLCD 2017. Relationships between tree height and (c) tree canopy coverage; and (d) building height over the city of Chicago. The dashed line indicates the superlinear trend between the averaged tree height and canopy coverage. The solid line in (d) shows the 1:1 ratio.

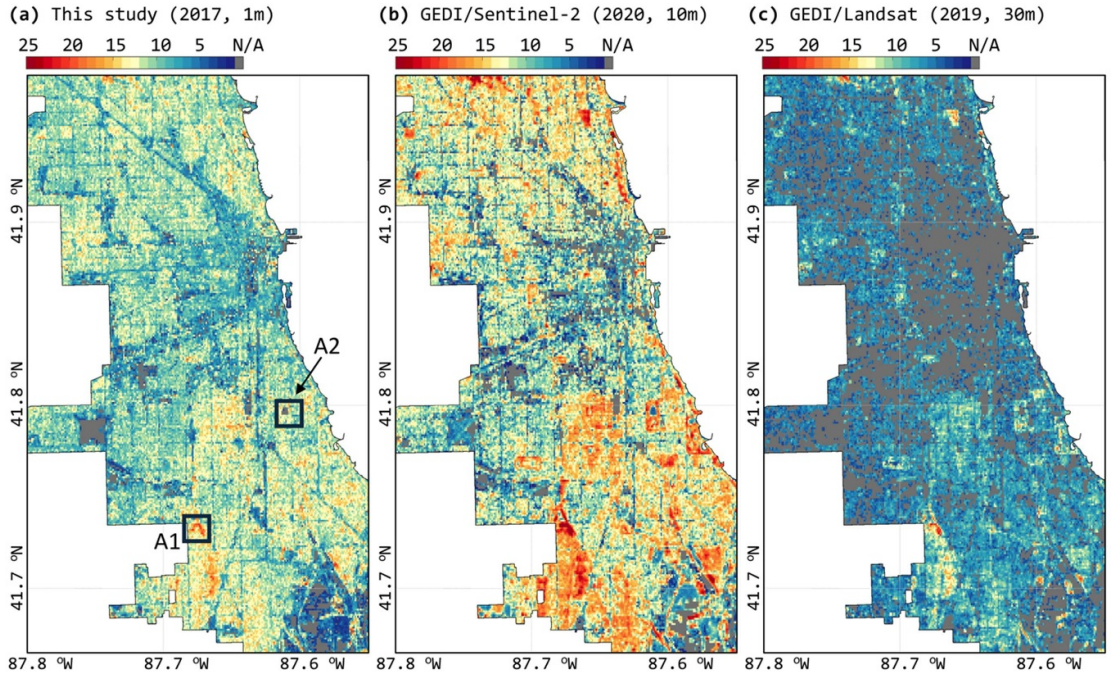

Figure 6. Maps of canopy height from (a) this study; (b) Lang *et al* (2022) using GEDI and Sentinel-2 at 10 m resolution; and (c) Potapov *et al* (2021) using GEDI and Landsat at 30 m resolution. Figure S5 shows the details in the black boxes marked in (a).

Figure 6 shows the spatial distribution of averaged tree height over Chicago for a 100 m sliding window with comparisons to existing canopy height products. Our estimated tree height lies between two other available estimates; it is 1.7 m lower than Lang *et al* (2022) and 4.1 m higher than Potapov *et al* (2021) on average. When scrutinizing the areas with large discrepancies, we find that Lang *et al* (2022) tends to misrecognize low vegetation as trees and overestimate their heights (figure S5(b)). This can be attributed to inaccurate surface elevation or to uncertainties in the classification scheme. The height estimation from Potapov *et al* (2021) is limited by its coarser resolution (figure S5(c)). As mentioned previously, signals from trees can be easily diluted due to their small footprints. Nevertheless, these products depict different

data years from ours. They also aimed to provide canopy height at a global scale, thus inevitably having spatial performance variances. Our high-resolution height map can be used for local calibration of these global products to ensure their accuracy with extended data coverage.

Additional findings can be derived when linking tree height to other geospatial properties. For example, averaged tree coverage has a positive correlation with tree height (figure 5(c)). The upper bound of this trend is superliner, meaning tall trees expand canopy coverage more effectively. Despite trees taking time to reach their mature height, planting tall-growing tree species will be more beneficial for a long-term investment. Another critical aspect is the relative height between trees and buildings. Intriguingly, in 67% of Chicago's land, trees are taller than the buildings (figure 5(d)). In 24% of those areas, the trees are twice the height of the buildings. This measurement is crucial for estimating the sky view factor, the roughness of urban terrain, and many urban studies on outdoor thermal comfort, street walkability, building energy consumption, etc. Yet, due to the abiding absence of accurate height information at high resolution, this measurement has been rarely considered or precisely quantified in the characterization of urban land, leaving large uncertainties in the current generation of urban models. In the next section, we will elaborate on how our dataset contributes to urban climate modeling.

3.3. Implications for urban environmental modeling

Given its unprecedented resolution, the dataset can be directly used in urban CFD simulations or aggregated to a desired granularity for coupled urban climate models. For example, urban canopy models, represented by the single-layer scheme PUCM/ASLUM (Wang *et al* 2013, 2021) and multi-layer scheme BEP-Tree (Martilli *et al* 2002, Krayenhoff *et al* 2020), require building and tree heights as model inputs. Additionally, when coupled with the regional Weather Research and Forecast/urban modeling framework (WRF-UCM, Chen *et al* 2011), these models need to incorporate the local climate zone (LCZ) concept for efficient parameterization at the regional scale (Stewart and Oke 2012). HiTAB can contribute from two aspects: (a) by providing localized values for the urban morphological parameters used in WRF, such as canyon aspect ratio; (b) by refining the existing LCZ classification. In Text S4, we analyze the height information with the 100 m LCZ Level 0 dataset in Chicago. Our analysis indicates that despite its relatively high resolution, the categorical implementation of LCZ vastly limits the land cover diversity and oversimplifies the urban terrain. Theoretically, LCZ determines the land use category based on the roughness and compactness of urban landscapes (Stewart and Oke 2012). However, the current derivation of LCZ relies on satellite data and does not consider the exact height of buildings or trees (Bechtel *et al* 2019), which potentially fails to represent the roughness accurately.

To better quantify this effect, we calculate the height ratio between buildings and trees (H_B/H_T) over a 100 m sliding window (figure 7(a)), corresponding to the resolution of Level 0 LCZ dataset (figure 7(b)). We find that trees are taller than buildings in most areas of Chicago. The variances of H_B/H_T lead to a transition of the 'roughness element' from buildings in city core, to the mix of trees and buildings in city outskirt, and eventually to trees in the suburban regions. The transition will subsequently affect the opening width of the streets (W) and canyon aspect ratio (H/W). For example, the value of W will be overestimated if only using building footprint (figure 7(c)). Tree canopies tend to close the street canyon when the coverage is high (figure 7(d)). Figure 7(e) shows the distribution of W when considering both buildings and trees as 'roughness elements'. Similarly, H/W varies with the consideration of buildings (figure 7(f)), trees (figure 7(g)), or both (figure 7(h)). More practically, one can determine the overall roughness as a linear or non-linear combinations of building and trees depending on the values of H_B/H_T to achieve a smoother transition. We anticipate in-depth quantitative investigations on surface roughness based on our high-resolution height maps and the measurement effort of the urban boundary layer.

Nevertheless, the oversimplification is a compromise due to the dearth high-resolution geospatial information and the constraints in computational resources. There are a couple of recent attempts to use spatial-distributed urban canopy parameters (UCPs) to mitigate the limitations of the categorical methods (Otte *et al* 2004, Salamanca *et al* 2011, Sun *et al* 2021). The derivation of UCP will benefit from the spatial gridded height dataset to fully account for the variances of urban features and the presence of vegetation. This is especially important for those advanced schemes that aim to quantify the ecohydrological (Meili *et al* 2020) and physiological functionality of trees (Li and Wang 2020, 2021, Li *et al* 2023) under urban forcings. Detailed information will also foster the data-driven models to resolve more complicated multi-objective optimization tasks (Li *et al* 2022). With this ever-growing interest in detailed parameterization and the emerging focus on urban greenery, the 'tree-resolving' capability is becoming one of the most essential components for the next generation 'urban-resolving' climate models (Sharma *et al* 2021). This evolution underscores the need for high-resolution urban datasets at a city or regional scale,

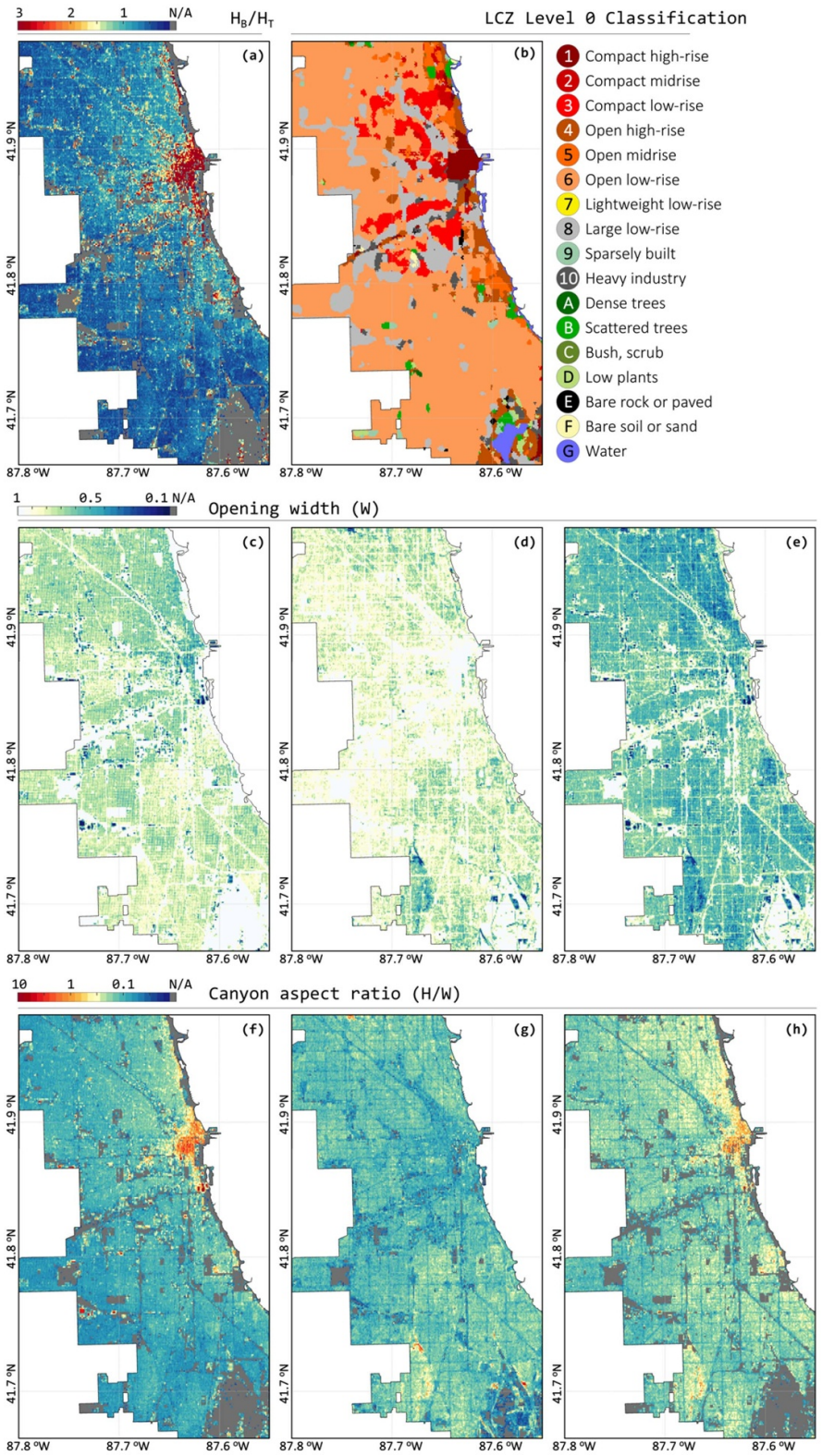

Figure 7. Maps of (a) the height ratio of buildings and trees (H_B/H_T); (b) current LCZ Level 0 classifications in Chicago. Subplots (c)–(e) show the estimation of street opening widths when considering buildings (c), trees (d), and both (e) as 'roughness elements'. Subplots (f)–(h) show the estimation of canyon aspect ratios (H/W) when considering buildings (f), trees (g), and both (h) as 'roughness elements'.

as exemplified by the pioneering efforts showcased in this study.

Apart from the applications on physical-based models, the high-resolution nature of this dataset will also help with data-driven urban studies to achieve hyper-local predictions (Venter *et al* 2020, Zumwald *et al* 2021). The precision and accuracy of such ML models largely depend on the richness of information (Wang *et al* 2023). Nevertheless, most of the studies still use the 'buffer zone' concept to average urban

parameters to a certain coarser resolution (Alonso and Renard 2020). The averaging method, however, may not fully reflect the impacts of urban features on the hyper-local environment. For instance, in the northern hemisphere, tall trees located south of a house may provide a better cooling effect in summer than trees in the north, and *vice versa* in the southern hemisphere. To address this problem, a new iteration of spatial embedding technique using neural networks has emerged (Fan *et al* 2021, Yu *et al* 2024). This novel technique considers the relative spatial layout of the urban features to the point of interest, thus resolving the dynamics at an extremely high resolution or even point scale. The fine granularity offered by our dataset will empower the implementation of such novel methods to capture the dynamics that cannot be reflected with a coarser resolution.

3.4. Caveats and future studies

Applying non-urban-oriented datasets (i.e. LiDAR, NAIP) to urban areas at extremely high resolution leads to notable uncertainties, such as the misclassification issue of LiDAR (section 2.2) and the shadow issue of NDVI (section 2.3). In addition, we find it intricate to distinguish the green roofs on tall buildings and the over-roof canopies from tall trees. The number of green roofs in Chicago has increased from 360 in 2010 to over 500 in 2023, accounting for 0.4% of the total roof area (Chicago Data Portal 2023). Even though we used multiple independent datasets to delineate green and gray spaces, further investigations by incorporating the exact mapping of green roofs are needed to enhance their accuracy.

The tree canopy in the Chicago region has been dominated by deciduous species (The Morton Arboretum 2020), with a strong seasonal variation of leaf density. Though LiDAR and NAIP NDVI datasets can accurately quantify the leaf density in a single time frame, their long revisit period makes it difficult to reflect the plant phenology. However, modeling results may be sensitive to biomass density change due to plant seasonality, especially in data-driven and/or long-term simulations (Yu *et al* 2024). We anticipate reconciling existing datasets on NDVI, leaf area index, or leaf area density across the spatial and temporal scales to derive the temporal evolution of biomass density maps at high resolution in a future study.

Though high precision spatial datasets are not commonly available across the globe for the current moment, following the derivation process of HiTAB, it is possible to generate nationwide building and tree height maps. For instance, USGS 3DEP offers LiDAR cloud point data over the United States, covering more than 400 cities and towns; while the vectorized building footprint can be found in an advanced ML study over North and South Americas, Australia, and some countries in Africa (Microsoft Maps 2018). For European cities, building footprints are publicly available from OSM Buildings (https://osmbuildings.org/data/). More cities, metropolises, and countries have planned to conduct high-precision LiDAR scans in recent years, such as the National LiDAR Programme in the United Kingdom and the National LiDAR High Density project in France. It is also possible to leverage ML techniques to use the data from US cities as the training sample for global cities to improve the data coverage in a fast and efficient way (Demuzere *et al* 2021). This will subsequently enhance the overall capability of Earth system models to resolve urban dynamics with high resolution over larger scales (Sharma *et al* 2021).

4. Concluding remarks

In this study, we quantified the height of buildings and trees at 1 m resolution over the city of Chicago using multiple sources of high-precision geospatial information. To our knowledge, HiTAB is one of the first few publicly available spatially gridded dataset on both buildings and tree heights at such high resolution. We are also working to expand the dataset coverage to the Chicago Metropolitan Area using newer data sources.

From the analysis in this study, we find that buildings in Chicago can be characterized by three different heights, corresponding to the building types. In most neighborhoods, trees are taller than buildings, resulting in canopies over the roof. We also note the significant differences in TCC between NLCD and our datasets, highlighting the importance of high-resolution datasets in urban environments. Moreover, we identified several challenges when integrating non-urban-oriented datasets for urban studies due to the intricate urban landscapes. Identifying these uncertainties presents opportunities for further investigation and improvement on related topics. Collectively, our dataset and methodology offered convenience and insights for urban climate studies. Practically, it will also help the communities to monitor and identify regions lacking canopy covers and guide strategic afforestation in the Chicago region.

Data availability statement

Datasets used in this study are publicly available and can be accessed via the following links as of January 2024. Illinois Height Modernization project: https://clearinghouse.isgs.illinois.edu/data/elevation/illinois-height-modernization-ilhmp; Chicago parcel-level land use inventory: www.cmap.illinois.gov/data/land-use/inventory; Chicago building footprint: https://data.cityofchicago.org/Buildings/Building-Footprints-current-/hz9b-7nh8; National Agriculture Imagery Program: https://naip-usdaonline.hub.arcgis.com/; and EPA

Meter-scale Urban Land Cover: https://doi.org/10.3390/rs12121909.

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5281/zenodo.10463648.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research's Urban Integrated Field Laboratories CROCUS project research activity, under Contract Number DE-AC02-06CH11357. This research is supported by NSF Awards # 2139316, 2230772, and 2330565 and NASA Award #80NSSC22K1683, and generous support from the Walder Foundation and Commonwealth Edison Company. We acknowledge the City of Chicago and the Chicago Metropolitan Agency for Planning for providing the data used in this study. We also thank Lindsay Darling from Purdue University for directing us to the UVM-SAL dataset used in this study. We also thank the Metropolitan Mayors Caucus, Chicago Region Tree Initiative, and Morton Arboretum for useful conversations.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID iDs

Peiyuan Li https://orcid.org/0000-0002-0057-2637
Ashish Sharma https://orcid.org/0000-0001-9332-8256

References

Aboutalebi M, Torres-Rua A F, Kustas W P, Nieto H, Coopmans C and McKee M 2018 Assessment of different methods for shadow detection in high-resolution optical imagery and evaluation of shadow impact on calculation of NDVI, and evapotranspiration *Irrig. Sci.* **1** 1–23

Alonso L and Renard F 2020 A new approach for understanding urban microclimate by integrating complementary predictors at different scales in regression and machine learning models *Remote Sens.* **12** 152434

Bechtel B, Alexander P J, Beck C, Böhner J, Brousse O, Ching J and Xu Y 2019 Generating WUDAPT Level 0 data—Current status of production and evaluation *Urban Clim.* **27** 24–45

Burgess D W, Lewis P and Muller J P A L 1995 Topographic effects in AVHRR NDVI data *Remote Sens. Environ.* **54** 543223–232

Casalegno S, Anderson K, Cox D T C, Hancock S and Gaston K J 2017 Ecological connectivity in the three-dimensional urban green volume using waveform airborne lidar *Sci. Rep.* **7** 45571

Center for Watershed Protection 2017 *Review of the available literature and data on the runoff and pollutant removal capabilities of urban trees* (available at: https://owl.cwp.org/mdocs-posts/review-of-the-available-literature-and-data-on-the-runoff-and-pollutant-removal-capabilities-of-urban-trees/)

Chen F, Kusaka H, Bornstein R, Ching J, Grimmond C S B, Grossman-Clarke S and Zhang C 2011 The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems *Int. J. Climatol.* **31** 273–88

Chicago Data Portal 2023 Green roofs map (available at: https://data.cityofchicago.org/Environment-Sustainable-Development/Green-Roofs-Map/u23m-pa73)

Ching J *et al* 2018 WUDAPT: an urban weather, climate, and environmental modeling infrastructure for the anthropocene *Bull. Am. Meteorol. Soc.* **99** 1907–24

City of Chicago 2022 2022 Chicago climate action plan (available at: www.chicago.gov/city/en/sites/climate-action-plan/home.html)

Darling L 2023 *Chicago region land cover* (https://doi.org/10.17605/OSF.IO/62NVZ)

Demuzere M, Kittner J and Bechtel B 2021 LCZ generator: a web application to create local climate zone maps *Front. Environ. Sci.* **9**

Fan C, Yang Y and Mostafavi A 2021 Neural embeddings of urban big data reveal emergent structures in cities (arXiv:2110.12371)

Giometto M G, Christen A, Egli P E, Schmid M F, Tooke R T, Coops N C and Parlange M B 2017 Effects of trees on mean wind, turbulence and momentum exchange within and above a real urban environment *Adv. Water Resour.* **106** 154–68

Guo Q, Su Y, Hu T, Guan H, Jin S, Zhang X and Coops N C 2020 Lidar boosts 3D ecological observations and modelings: a review and perspective *IEEE Geosci. Remote Sens. Mag.* **9** 232–57

Hirabayashi S and Nowak D J 2016 Comprehensive national database of tree effects on air quality and human health in the United States *Environ. Pollut.* **215** 48–57

Housman I W *et al* 2021 *National land cover database tree canopy cover methods v2021* (U.S. Department of Agriculture Forest Service Geospatial Technology and Applications Center 24) p 9 (available at: https://data.fs.usda.gov/geodata/rastergateway/treecanopycover/docs/TCC_v2021-4_Methods.pdf)

Huo L-Z, Silva C A, Klauberg C, Mohan M, Zhao L-J, Tang P and Hudak A T 2018 Supervised spatial classification of multispectral LiDAR data in urban areas *PLoS One* **13** 10e0206185

ILHMP—Illinois Height Modernization 2021 LiDAR data (available at: https://clearinghouse.isgs.illinois.edu/data/elevation/illinois-height-modernization-ilhmp)

Kara A, van Oosterom P, Çağdaş V, Işıkdağ Ü and Lemmen C 2020 3 Dimensional data research for property valuation in the context of the LADM Valuation Information Model *Land Use Policy* **98** 104179

Kashani A G, Olsen M J, Parrish C E and Wilson N 2015 A review of LiDAR radiometric processing: from ad hoc intensity correction to rigorous radiometric calibration *Sensors* **15** 28099–128

Krayenhoff E S, Jiang T, Christen A, Martilli A, Oke T R, Bailey B N and Crawford B R 2020 A multi-layer urban canopy meteorological model with trees (BEP-Tree): street tree impacts on pedestrian-level climate *Urban Clim.* **32** 100590

Lang N, Jetz W, Scjindler K and Wegner J D 2022 A high-resolution canopy height model of the Earth (arXiv:2204.08322)

Li P and Sharma A 2024 Hyper-local temperature prediction using detailed urban climate informatics *J. Adv. Model. Earth Syst.* **16** e2023MS003943

Li P, Sharma A, Wang Z-H and Wuebbles D 2023 Assessing impacts of environmental perturbations on urban biogenic

carbon exchange in the Chicago region *J. Adv. Model. Earth Syst.* **15** 10e2023MS003867
Li P and Wang Z-H 2020 Modeling carbon dioxide exchange in a single-layer urban canopy model *Build. Environ.* **184** 107243
Li P and Wang Z-H 2021 Uncertainty and sensitivity analysis of modeling plant CO_2 exchange in the built environment *Build. Environ.* **189** 107539
Li P, Wang Z-H and Wang C 2024 The potential of urban irrigation for counteracting carbon-climate feedback *Nat. Commun.* **15** 2437
Li P, Xu T, Wei S and Wang Z-H 2022 Multi-objective optimization of urban environmental system design using machine learning *Comput. Environ. Urban Syst.* **94** 101796
Martilli A, Clappier A and Rotach M W 2002 An urban surface exchange parameterisation for mesoscale models *Bound.-Layer Meteorol.* **104** 261–304
McDonald A G, Bealey W J, Fowler D, Dragosits U, Skiba U, Smith R I and Nemitz E 2007 Quantifying the effect of urban tree planting on concentrations and depositions of PM10 in two UK conurbations *Atmos. Environ.* **41** 388455–8467
Meili N, Manoli G, Burlando P, Bou-Zeid E, Chow W T L, Coutts A M and Fatichi S 2020 An urban ecohydrological model to quantify the effect of vegetation on urban climate and hydrology (UT&C v1.0) *Geosci. Model. Dev.* **13** 335–62
Meyer D, Grimmond S, Dueben P, Hogan R and van Reeuwijk M 2022 Machine learning emulation of urban land surface processes *J. Adv. Model. Earth Syst.* **14** e2021MS002744
Microsoft Maps 2018 *Country wide open building footprints datasets in United States* (available at: https://github.com/microsoft/USBuildingFootprints)
Middel A, Nazarian N, Demuzere M and Bechtel B 2022 Urban climate informatics: an emerging research field *Front. Environ. Sci.* **10** 867434
Mirzaei P A 2021 CFD modeling of micro and urban climates: problems to be solved in the new decade *Sustain. Cities Soc.* **69** 102839
Morsy S, Shaker A and El-Rabbany A 2017 Multispectral LiDAR data for land cover classification of urban areas *Sensors* **17** 958
NAIP—National Agriculture Imagery Program 2023 Data (https://doi.org/10.5066/F7QN651G)
O'Neil-Dunne J, MacFaden S and Royar A 2014 A versatile, production-oriented approach to high-resolution tree-canopy mapping in urban and suburban landscapes using GEOBIA and data fusion *Remote Sens.* **6** 12837–65
Oke T R 1982 The energetic basis of the urban heat island *Q. J. R. Meteorol. Soc.* **108** 1–24
Oke T R 2006 *Initial guidance to obtain representative meteorological observations at urban sites* WMO/TD-No. 1250 (World Meteorological Organization) (available at: https://library.wmo.int/idurl/4/35333)
Otte T L, Lacser A, Dupont S and Ching J K S 2004 Implementation of an urban canopy parameterization in a mesoscale meteorological model *J. Appl. Meteorol.* **43** 111648–1665
Pilant A, Endres K, Rosenbaum D and Gundersen G 2020 US EPA enviroatlas meter-scale urban land cover (MULC): 1-m pixel land cover class definitions and guidance *Remote Sens.* **12** 1909
Potapov P, Li X, Hernandez-Serna A, Tyukavina A, Hansen M C, Kommareddy A and Hofton M 2021 Mapping global forest canopy height through integration of GEDI and Landsat data *Remote Sens. Environ.* **253** 112165
Pourpeikari Heris M, Bagstad K J, Troy A R and O'Neil-Dunne J P M 2022 Assessing the accuracy and potential for improvement of the national land cover database's tree canopy cover dataset in urban areas of the conterminous United States *Remote Sens.* **14** 1219
Ritz A L *et al* 2022 Assessing the utility of naip digital aerial photogrammetric point clouds for estimating canopy height of managed loblolly pine plantations in the southeastern united states *Int. J. Appl. Earth Observ. Geoinf.* **113** 103012
Ryu Y-H, Bou-Zeid E, Wang Z-H and Smith J A 2015 Realistic representation of trees in an urban canopy model *Bound.-Layer Meteorol.* **159** 193–220
Salamanca F, Martilli A, Tewari M and Chen F 2011 A study of the urban boundary layer using different urban parameterizations and high-resolution urban canopy parameters with WRF *J. Appl. Meteorol. Climatol.* **50** 1107–28
Schwaab J, Meier R, Mussetti G, Seneviratne S, Bürgi C and Davin E L 2021 The role of urban trees in reducing land surface temperatures in European cities *Nat. Commun.* **12** 6763
Sharma A, Wuebbles D J and Kotamarthi R 2021 The need for urban-resolving climate modeling across scales *AGU Adv.* **2** e2020AV000271
Stewart I D and Oke T R 2012 Local climate zones for urban temperature studies *Bull. Am. Meteorol. Soc.* **93** 121879–1900
Sun Y, Zhang N, Miao S, Kong F, Zhang Y and Li N 2021 Urban morphological parameters of the main cities in china and their application in the WRF model *J. Adv. Model. Earth Syst.* **13** e2020MS002382
The Morton Arboretum 2020 *Chicago region tree census* (available at: https://mortonarb.org/science/tree-census/)
Tooke T R, Coops N C, Christen A, Gurtuna O and Prévot A 2012 Integrated irradiance modelling in the urban environment based on remotely sensed data *Sol. Energy* **86** 102923–2934
Toparlar Y, Blocken B, Maiheu B and van Heijst G J F 2017 A review on the CFD analysis of urban microclimate *Renew. Sustain. Energy Rev.* **80** 1613–40
United Nations 2018 World urbanization prospects: The 2018 revision (available at: https://population.un.org/wup/Publications/)
Venter Z S, Brousse O, Esau I and Meier F 2020 Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data *Remote Sens. Environ.* **242** 111791
Wang C, Wang Z-H and Ryu Y-H 2021 A single-layer urban canopy model with transmissive radiation exchange between trees and street canyons *Build. Environ.* **191** 107593
Wang H, Yang J, Chen G, Ren C and Zhang J 2023 Machine learning applications on air temperature prediction in the urban canopy layer: a critical review of 2011–2022 *Urban Clim.* **49** 101499
Wang Z-H, Bou-Zeid E and Smith J A 2013 A coupled energy transport and hydrological model for urban canopies evaluated using a wireless sensor network *Q. J. R. Meteorol. Soc.* **139** 6751643–1657
Woodward A, Hinwood A, Bennett D, Grear B, Vardoulakis S, Lalchandani N and Williams C 2023 Trees, climate change, and health: an urban planning, greening and implementation perspective *Int. J. Environ. Res. Public Health* **20** 18
Yang X, Zuo X, Xie W, Li Y, Guo S and Zhang H 2022 ACorrection method of NDVI topographic shadow effect for rugged terrain *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **15** 8456–72
Yu Y, Li P, Huang D and Sharma A 2024 Street-level temperature estimation using graph neural networks: performance, feature embedding and interpretability *Urban Clim.* **56** 102003
Zumwald M, Knüsel B, Bresch D N and Knutti R 2021 Mapping urban temperature using crowd-sensing data and machine learning *Urban Clim.* **35** 100739