UrbanTALES: a large-eddy simulation dataset for urban canopy layer turbulence and parameterization

BAMS Article

UrbanTALES: A Large-Eddy Simulation Dataset for Urban Canopy Layer Turbulence and Parameterization

Negin Nazarian,[a,b] Jiachen Lu,[a] Mathew J. Lipson,[b] Melissa A. Hart,[b,c] Sijie Liu,[b] E. Scott Krayenhoff,[d] Lewis Blunn,[e] and Alberto Martilli[f]

KEYWORDS: Large eddy simulations; Parameterization; Single column models; Urban meteorology; Microscale processes/variability

ABSTRACT: The urban canopy layer (UCL) exhibits complex, heterogeneous flow patterns shaped by urban geometry. Traditionally, research has relied on microscale simulations over limited and often idealized building arrays, leaving a need for more extensive datasets to capture the dynamics across diverse urban neighborhoods. Responding to this gap, we developed an extensive dataset, known hereafter as Urban Turbulence Analyses from Large-Eddy Simulations (UrbanTALES), based on state-of-the-art large-eddy simulations (LESs) over 538 urban layouts (generated using over 3 000 000 CPU hours and 35 TB of storage) with both idealized and realistic configurations. Realistic urban neighborhood configurations were obtained from major cities worldwide, incorporating wide variations in building plan area densities [0.06–0.64] and height distributions [4–50 m]. Idealized urban arrays, on the other hand, include two commonly studied configurations (aligned and staggered building arrays), featuring both uniform and variable height scenarios along with oblique wind directions. UrbanTALES offers canopy-averaged flow data as well as 2D and 3D flow fields tailored for different applications in urban climate research such as the development and testing of urban canopy models. The dataset provides time-averaged wind flow properties, as well as second- and third-order flow moments that are critical for understanding turbulent processes in the UCL. Here, we describe the UrbanTALES dataset and its applications, noting the unique opportunity to use high-fidelity simulated flow in realistic urban neighborhoods to 1) revisit neighborhood-scale urban canopy parameterizations in various climate models and 2) inform in-canopy flow and turbulent analyses in complex urban configurations. UrbanTALES is openly available at https://urbantales.climate-resilientcities.com/ and can be extended to incorporate future LES datasets in the field.

DOI: 10.1175/BAMS-D-25-0061.1

Corresponding author: Negin Nazarian, n.nazarian@unsw.edu.au

Negin Nazarian and Jiachen Lu contributed equally to this work.

Manuscript received 7 March 2025, in final form 18 July 2025, accepted 30 September 2025

SIGNIFICANCE STATEMENT: The urban canopy layer plays a crucial role in shaping urban climate, yet its complexity has often been oversimplified due to limited datasets. To address this, we developed Urban Turbulence Analyses from Large-Eddy Simulations (UrbanTALES), an open-access dataset based on high-resolution large-eddy simulations (LESs) over 538 urban configurations, including both real-world neighborhoods and idealized building layouts. With detailed flow fields and turbulence statistics, UrbanTALES provides a new foundation for improving urban canopy models, turbulence studies, and informing weather/climate simulations over cities. By making it openly available, we aim to foster collaboration and encourage future extensions that enhance our understanding of urban airflow.

AFFILIATIONS: [a] School of Built Environment, University of New South Wales, Sydney, New South Wales, Australia; [b] ARC Centre of Excellence for 21st Century Weather, UNSW Sydney, Sydney, New South Wales, Australia; [c] Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia; [d] School of Environmental Sciences, University of Guelph, Guelph, Ontario, Canada; [e] Met Office, University of Reading, Reading, United Kingdom; [f] Atmospheric Pollution Division, Environmental Department, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Madrid, Spain

1. Introduction

Numerical modeling of urban physics has historically focused on the representation of three key physical processes: 1) airflow and turbulence within and above urban roughness, 2) energy balance of urban surfaces, and 3) mass (water and pollution) transfer in urban canopies (Oke et al. 2017). While these three processes are intricately interdependent, many models typically center around one, with some efforts attributed to coupling their interactions.

Focusing on airflow and turbulent exchanges, computational fluid dynamics (CFD) models, which solve the Navier–Stokes (NS) equations to describe fluid flow, represent a primary method. Microscale CFD simulations have enabled high-resolution (submeter) analyses of three-dimensional flow and turbulent properties within the urban canopy (Toparlar et al. 2017; Nazarian et al. 2018b), while also being used as an effective basis for developing urban canopy parameterizations (Santiago and Martilli 2010; Buccolieri et al. 2021; Nazarian et al. 2018b). These CFD simulations fall into three main categories based on their turbulent parameterizations: 1) Reynolds-averaged NS (RANS), modeling time-averaged flow and turbulence with lower computational costs; 2) large-eddy simulation (LES) models, parameterizing the smallest length scales and resolving larger-scale eddies; and 3) direct numerical simulations (DNSs) explicitly solving NS equations without any turbulence parameterizations. While DNS models enhance our theoretical understanding (Coceal et al. 2007), their high computational cost makes LES and RANS models more practical for assessing urban flow processes. RANS models perform well for time-averaged wind speed and pollution but fail to accurately represent turbulent exchanges, whereas LES is state of the art for accurately capturing turbulent exchanges of momentum and energy in urban environments (Nazarian et al. 2020).

In the last decade, LES models have been extensively utilized to enhance our understanding of urban canopy flow, examining both idealized (Xie et al. 2008) and realistic

(Park et al. 2015) urban layouts. Despite some efforts to account for thermal forcing (Geletič et al. 2021) and trees (Giometto et al. 2017) in realistic street canyons, most of these studies have focused on the impact of built geometry in neutral conditions. This emphasis is not solely due to model limitations or computational constraints in representing realistic urban settings; rather, it stems from two key factors. First, we are still developing our understanding of the impacts of realistic urban form (encompassing both horizontal and vertical heterogeneity) on urban flow and turbulence (Lu et al. 2023a). This has motivated numerous studies that evaluate the contribution of urban form in isolation. Second, there is a recognition that thermal forcing's influence on flow is intricately linked to the ratio between momentum and thermal forcing (Santiago et al. 2014; Simón-Moral et al. 2016; Nazarian et al. 2018a). In other words, there are still numerous scenarios and atmospheric conditions throughout the day when urban canopy processes are predominantly driven by momentum forcing.

In essence, the field is still grappling with fundamental questions about how realistic urban neighborhoods influence airflow within the canopy and flux exchange between the surface and the urban boundary layer, as well as how these processes can be effectively represented in various urban canopy models. Current efforts to address these challenges include evaluating idealized simulations that specifically explore variations in density or height distribution (Lu et al. 2023b), as well as investigations into drag- and length-scale parameterization for urban canopy models (Nazarian et al. 2020; Sützl et al. 2021). In contrast, research focusing on realistic neighborhoods often relies on a handful of urban layouts to examine flow properties such as turbulent and dispersive fluxes (Giometto et al. 2016; Akinlabi et al. 2022) or pollutant dispersion (Lim et al. 2022). A recent approach to characterizing urban configurations also introduced the use of local climate zone (LCZ) classifications (Stewart and Oke 2012) for systematic comparisons of urban neighborhoods (Nagel et al. 2023), i.e., focusing on 10 urban classes as opposed to detailed urban data. However, as urban layouts can vary significantly within each LCZ, and with the increasing availability of high resolution, gridded urban datasets, the reliance on generalized class-based methods such as LCZs may no longer be the most effective approach (Lipson et al. 2022).

Most of the existing literature relies on limited representations of urban areas, and many studies duplicate simulations across various idealized layouts. This approach poses challenges when extrapolating findings to different neighborhoods and introduces uncertainties in parameterizations of surface fluxes. Some notable studies have attempted to address this gap by leveraging more comprehensive datasets. Kanda et al. (2013) used an extensive dataset of more than 100 LESs to refine the parameterization of aerodynamic parameters such as roughness length and displacement height for real urban areas. Their work also provided horizontally averaged turbulent statistics and surface drag corresponding to diverse urban morphologies. More recently, Lu et al. (2023a,b) used more than 100 LES simulations in both idealized and realistic urban layouts to identify key geometrical parameters that capture the effects of horizontal and vertical heterogeneities on urban airflow. Despite these advancements, such datasets are often inaccessible to the broader scientific community or limited in the parameters and setups they release, thereby restricting their broader applicability to urban canopy modeling. To address this gap, we aim to expand, standardize, and openly share the LES cases developed in Lu et al. (2023a,b), enabling their application across a wide range of urban climate research contexts. The outcome is an openly available dataset comprising 538 LES cases for both realistic and idealized urban layouts, hosted on an open, user-friendly web-based platform.

This paper details the workflow for modeling, postprocessing, and openly sharing this LES dataset. The dataset includes canopy-averaged flow properties, as well as two- and three-dimensional fields encompassing time-averaged wind flow properties and

second- and third-order flow moments. We present sample results showcasing canopy- and horizontally averaged properties, illustrating the differences between realistic and idealized cases. Last, we discuss the dataset's applications and introduce the web-based dissemination platform, which facilitates seamless access and usability for the research community.

2. Numerical modeling and computational setup

a. Large-eddy simulation: Model configuration and evaluation. LES is the state-of-the-art modeling approach for accurately and practically capturing turbulent exchanges of momentum and energy in urban environments and is therefore suitable for a range of applications focused on urban canyons as well as surface–atmosphere fluxes over urban surfaces. The LES model used here is the parallelized large-eddy simulation model (PALM), version r4554 (Maronga et al. 2020), a sophisticated computational tool developed over two decades. PALM leverages parallel computing architectures to simulate turbulent flows in complex environments, such as urban areas, with high spatial and temporal resolution (Maronga et al. 2015). Its dynamic regularized gradient subgrid-scale turbulence model enhances accuracy by effectively capturing the interactions between resolved and unresolved turbulence scales. Furthermore, PALM has been extended into PALM-4U, a multiphysics urban climate modeling framework that incorporates surface energy balance processes, radiation, hydrology, vegetation effects, and building energy modeling. These advanced capabilities make PALM an ideal platform for developing the UrbanTALES dataset, which aims to expand beyond the neutral atmospheric conditions explored here to encompass a wider range of urban meteorological scenarios.

PALM has undergone extensive validation against measurements (Yaghoobian et al. 2014; Maronga et al. 2015; Nazarian et al. 2018b; Maronga et al. 2020) and has been widely employed for investigating turbulent flow over urban canopies. For the current study, the simulation setup (as indicated in Table 1) builds on prior research (Nazarian et al. 2020; Lu et al. 2023b,a) and is consistent across all configurations to facilitate systematic comparisons. This setup was further validated against DNS simulations (Coceal et al. 2007) and wind tunnel experiments (Brown et al. 2001) in idealized configurations (Nazarian et al. 2020). Note that both spanwise and streamwise boundaries are treated as periodic (Table 1), so the flow field is effectively tiled in the streamwise direction. This idealization facilitates statistical convergence over an infinitely repeating canopy but precludes upwind fetch-dependent boundary layer evolution and spatial complexity in upwind conditions. Users should therefore be aware that applications requiring realistic inflow development or nonrepeating upwind fetches will need dedicated inlet and downstream boundary formulations.

Table 1. Numerical configurations of PALM in UrbanTALES. For more information on each scheme, see Maronga et al. (2020).

Model configuration	Scheme
Turbulence closure	1.5-order closure
Vertical discretization	Arakawa staggered C grid
Horizontal discretization	Second-order central differences
Time discretization	Minimal storage scheme
Pressure solver	Fastest Fourier Transform in the West (FFTW)
Lateral boundary conditions	Periodic
Surface boundary condition	Neumann following log law (roughness length 0.01 m)
Top boundary condition	Dirichlet (no slip)
Topography	2D building height in ASCII form

Sensitivity analyses from Nazarian et al. (2020) and Lu et al. (2023a) further informed key factors influencing simulation accuracy, including geometrical configurations (domain size and height), grid resolution, and run-time parameters (spinup time, sampling frequency, and time-averaging interval). Domain height and resolution were found to be most critical. The total domain height (H_T) was set to 5–8 times the mean building height (H_{mean}), as smaller heights [e.g., $4H_{\text{mean}}$ used in RANS studies (Santiago and Martilli 2010)] affected Reynolds stress and turbulent kinetic energy (TKE) profiles above the canopy. The H_T was set to ~130 m and kept uniform for all cases to enable more accurate and systematic comparisons. We then restricted the maximum building height to 50 m, equivalent to a 20-story building, to ensure that the simulation domain is adequately large to resolve large-scale eddies generated by buildings, while also facilitating the examination of tall structures within urban contexts.

Grid sensitivity tests demonstrated that resolutions of 1 m and below were necessary for the accurate convergence of 2D fields and profiles. Further comparisons of 1 and 0.5 m resolutions by Lu et al. (2023a) confirmed that vertical profiles of wind speed, momentum fluxes, and TKE profiles remained within a ±5% confidence range. Some underestimation of TKE and maximum wind speed occurred near small building gaps due to smaller eddies in the coarse grid being underresolved. However, 1-m resolution was shown to effectively balance computational efficiency and accuracy and was therefore used for realistic configurations, while 0.5 m was adopted for idealized configurations that are less computationally intensive.

The vertical grid resolution is uniform for the first 50 m and stretched thereafter with a stretching ratio of 1.1 (Nazarian et al. 2020). Finally, sensitivity analyses informed the run-time parameters, informing the 4-h spinup phase (~140 eddy turnover times, H_{domain}/u_τ) for all simulations to achieve quasi-steady-state conditions (section 3). Outputs were then time averaged and stored every half hour over 8 h (~375 eddy turnover times). The simulation setups are consistent with previous applications of PALM for urban studies (Blunn et al. 2022), supporting the setup and essential parameters for achieving high-fidelity LES results.

We studied urban canopy flow with neutral atmospheric conditions; that is, the only external driver is a constant pressure gradient, applied in the whole domain, in the streamwise wind direction (x):

$$\frac{\partial p}{\partial x} = A_a \rho u_\tau^2 / V_{\text{air}}, \tag{1}$$

where A_a is the surface area of air, ρ is the density of air (assumed constant with height), u_τ is the friction velocity, and V_{air} is the total volume of air that equals to total domain volume minus the building volume, which results in a slight variation (within 5%) in u_τ between the densest and sparsest cases. Note that the relative importance of thermal forcing or convection to the dynamic drive was found to be minor if the ratio between roughness scale and stability length scale $H/L_{\text{urb}} \leq 1$ (Simón-Moral et al. 2016), or the ratio between momentum and thermal forcing (Nazarian et al. 2018b), is small. Therefore, our evaluation of the flow is not strictly constrained to neutral conditions; rather, it can be used to inform urban canopy flow with a mild range of thermal forcing. During iterations, the driver is balanced by the force generated through interactions with the roughness elements in the canopy (Sützl et al. 2021). The Reynolds number $\text{Re}_\tau = u_\tau H_T/\nu \sim 2.5 \times 10^6$ is large enough to neglect viscosity effects and for the flow to be Reynolds number independent. In addition to solving the momentum and pressure equations, we introduced a passive scalar with an initial surface concentration of $c_0 = 0.06$ and imposed a uniform vertical gradient of -5×10^{-5} m^{-1} to establish its initial profile. Apart from responding to the urban airflow, the scalar has Neumann boundary conditions

at the top and the bottom of the domain, which is a constant surface flux of 1×10^{-4} and a sink at the top domain of $1 \times 10^{-4} \times (1 - \lambda_0)$, where λ_0 denotes the surface urban density to maintain the total scalar concentration.

b. Simulation cases. A total of 538 LES cases are included in the first release of UrbanTALES. The simulation cases are categorized into two distinct groups: 1) idealized configurations, where urban layouts are determined by varying plan area density (λ_p) and building height distribution (section 1); and 2) realistic configurations, where diverse urban neighborhoods are selected from datasets providing detailed global maps at the building scale (section 2). In idealized arrays, building footprints and street geometry remain uniform, while realistic neighborhoods have varying building footprints and street alignments found in real locations. To ensure comprehensive coverage, we further evaluated a range of building height distributions in both configurations.

Figure 1 displays the geometric parameters for all simulation cases, comparing them with the LES results from Kanda et al. (2013), which was focused on Japanese neighborhoods. In both idealized and realistic categories, our approach ensures the representation of the most typical urban densities observed in global cities, i.e., $\lambda_p \in [0.065, 0.65]$ (Oke et al. 2017). UrbanTALES simulation cases cover the same range of λ_p as Kanda et al. (2013), while representing a larger range of λ_f, particularly more neighborhoods with high λ_p but low λ_f (i.e., compact low and midrise) or low λ_p and high λ_f (i.e., open mid- and high rise). Although the maximum height of the buildings in our simulations was restricted to 50 m (~15 story), this range captures neighborhoods with midrise and some high-rise buildings common in urban areas across most countries in the Americas, Europe, Africa, and Oceania. Some skyscraper-dominated neighborhoods with building heights exceeding 50 m (seen in Asian cities) are not covered in the current dataset, which warrants separate investigation. Nevertheless, the presented dataset provides the largest sample size of urban flow data comprehensive for midrise to high-rise configurations typical of most global cities, providing a robust foundation for studying urban densities and aerodynamic effects.

1) Constructing idealized urban neighborhoods. Conventionally, two designs of idealized building arrays—aligned and staggered configurations—have been used to represent real neighborhoods. The aligned array typically represents scenarios featuring a single,

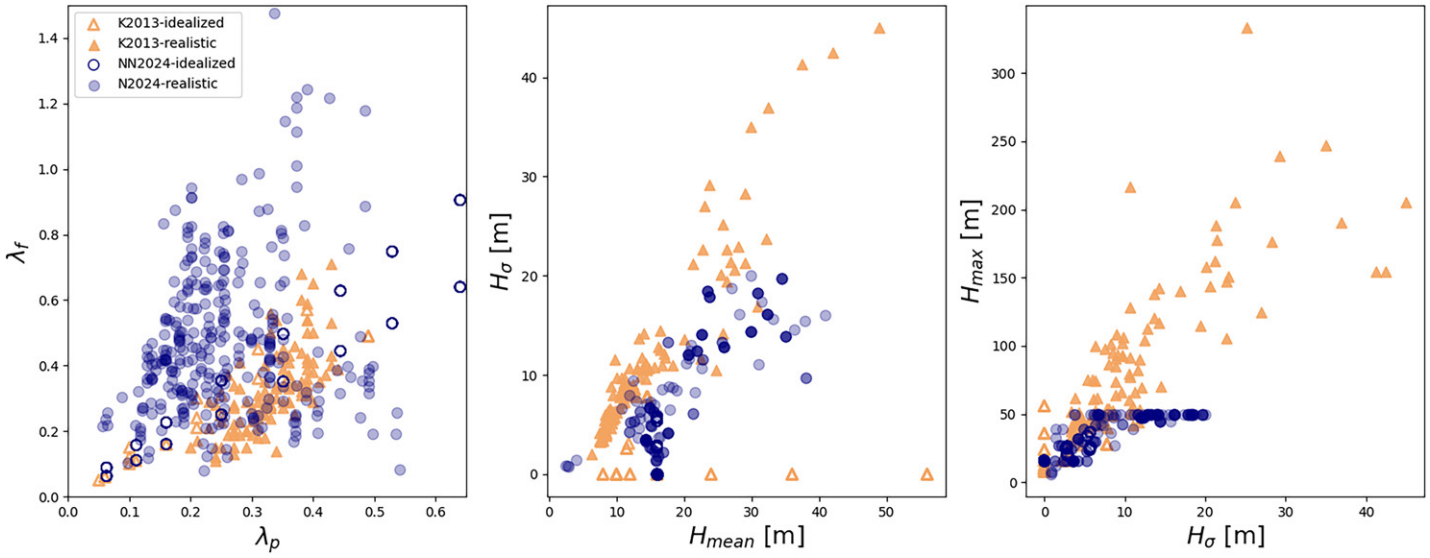
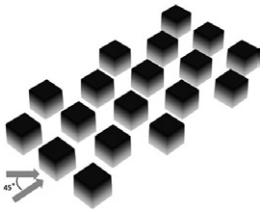
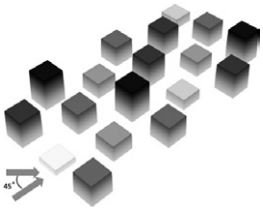
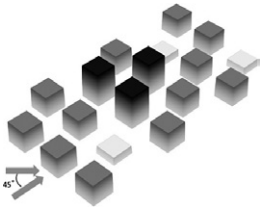
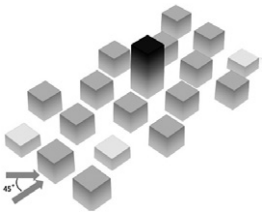
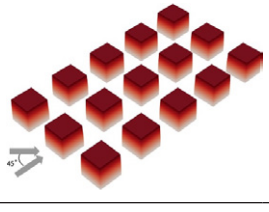
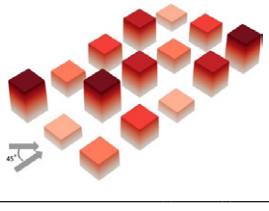
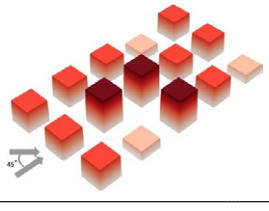
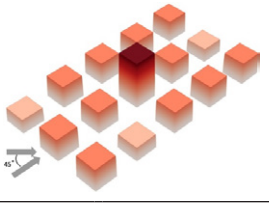


Fig. 1. Distribution of urban geometrical parameters in UrbanTALES compared with LES simulations of Kanda et al. (2013). The λ_p and λ_f represent plan and frontal area density; H_σ, H_{mean}, and H_{max} represent the SD of the building height distribution and the mean and maximum of building height.

uninterrupted street aligned parallel to the prevailing wind direction, simulating an extreme case of urban layouts. In contrast, the staggered array is often employed to calibrate urban canopy models (UCMs) (Santiago and Martilli 2010; McNorton et al. 2021), as it avoids the presence of major wind-aligned corridors, which can artificially reduce building drag. While both configurations simplify real-world complexities, Lu et al. (2023a) demonstrated that staggered arrays offer a closer approximation to average urban flow conditions for low-density neighborhoods. Conversely, aligned arrays better describe high-density urban areas, capturing key features of their airflow dynamics.

UrbanTALES expands idealized configurations used in Lu et al. (2023b). Both staggered and aligned configurations are included with two wind directions (0° and 45°), three height distributions, and four maximum building heights (H_{max}) (Table 2). The same volume of roughness elements is maintained by fixing the mean building height (H_{mean} = 16 m) across all cases. To achieve representative vertical heterogeneity, we then consider three standard deviations (SDs) of building height, $H_\sigma \in [0, 2.8, 5.6\ \text{m}]$, with different maximum building height (H_{max}) to generate four types of urban arrays: 1) uniform urban arrays with constant building height, used to build urban canopy parameterization in Nazarian et al. (2020); 2) continuous configuration representing a continuous change in building height distribution with a relatively low H_{max}; 3) clustered configuration featuring clusters of low and tall buildings, each consisting of three buildings; and 4) high-rise configuration featuring one high-rise and three low-rise building blocks. Each configuration is then constructed with eight urban packing densities, i.e., $\lambda_p \in [0.0625, 0.1111, 0.16, 0.25, 0.35, 0.4444, 0.53, 0.64]$, to achieve horizontal heterogeneity. Further incorporating two varying wind directions, this

Table 2. Dataset details for 224 idealized urban arrays discussed in this study. Two idealized horizontal arrangements are distinguished by red (aligned) and black (staggered) colors. Four building height arrangements (uniform, continuous, clustered, and high rise) with two SD of height configurations are considered. The maximum H_{max} and minimum H_{min} building heights and a 3D view of example (λ_p = 0.25) are shown below each category.

Height config.	Uniform	Continuous		Clustered		High rise	
Staggered							
H_σ (m)	0	2.8	5.6	2.8	5.6	2.8	5.6
H_{mean} (m)	16	16	16	16	16	16	16
H_{min} (m)	16	10	4	11	6	12.5	9
H_{max} (m)	16	20	24	21	26	26.5	37
Topography							
Aligned							
H_σ (m)	0	2.8	5.6	2.8	5.6	2.8	5.6
H_{mean} (m)	16	16	16	16	16	16	16
H_{min} (m)	16	12	8	11.5	7	13	10
H_{max} (m)	16	20	24	20.5	25	25	34
Topography							

yields a total of 224 simulation cases with a domain size from 100 × 60 m^2 (200 × 120 grid points) to 384 × 192 m^2 (768 × 384 grid points).

2) Selecting realistic urban neighborhoods. Globally, building-scale urban datasets are becoming increasingly available in various cities [such as NYC OpenData in the United States and Geoscape (PSMA Australia 2020) in Australia], providing high-quality spatial data. However, differences in data collection methods, formats, resolutions, and metadata standards often limit the creation of uniform and comparable representations of urban neighborhoods. To address this and ensure global coverage of realistic urban neighborhoods (Herfort et al. 2023), we primarily rely on OpenStreetMap (OSM) for building footprints. OSM is a free, editable map of the world, built and maintained by a community of contributors. To supplement OSM data and address gaps in building polygons, we integrated data from Microsoft Open Buildings. This combined approach enabled the development of *OSM2LES* (Lu et al. 2022), a tool designed to efficiently extract building footprints from OSM (OpenStreetMap Contributors 2017) and process them for LES/PALM simulations. Using *OSM2LES*, each realistic urban neighborhood is generated based on the bounding box for the area of interest, the desired resolution, and default building heights (optional) given by users, following three key steps (Fig. 2a):

1) Downloading and rasterization: OSM2LES takes the bounding box specified by the user (blue vertices and lines as indicated in the first column of Fig. 2a) to download raw **.osm* file format from OSM. This OSM output represents the building footprint in vectorial form, which is not suitable for PALM simulations that only recognize a raster of building height for urban topography. The OSM2LES code then rasterizes the building footprint data given a prescribed resolution (1 m here) and fills the building grids with height information or a default value the user prescribes. If height information is not available in OSM, we complement it with secondary sources like World Settlement Footprint (WSF3D; Esch et al. 2022). This additional height data will be released in the next version of OSM2LES. The level of details (LODs; Biljecki et al. 2016) for the resulting realistic layouts is equal to 1.2 for the cases of uniform height and 1.3 for variable heights. In these LODs, the small horizontal building elements (equal or larger than 1 m^2 under 1-m resolution) were identified, but a uniform flat roof shape is applied.
2) Rotation and cropping: The selected domain for urban topography after rasterization may include some artificial empty space at corners (Fig. 2a). Therefore, it is necessary to rotate and crop the resulting domain. The cropping should leave a reasonable lateral space of roughly equal to half (considering the periodic boundary condition in section 2a) of the typical street width (3 m) for better representation of the urban neighborhoods. It is worth noting that rotation of the domain does not break the realism of the simulation in that it is equivalent to applying a specific prevalent wind forcing that exists in the real world. There is a range of domain sizes since the shapes of realistic urban neighborhoods (naturally bounded by surrounding streets) vary inherently. UrbanTALES covers neighborhoods ranging from 300 × 300 m^2 to 1000 × 1000 m in area, which is equivalent to the grid resolution of contemporary mesoscale models.
3) Geometry evaluation: In this step, the statistical information regarding the morphology, such as plan and frontal area density (λ_p, λ_f), alignedness (γ; Lu et al. 2023a), is calculated. The rasterized urban topography (depicted as a 2D array of building heights where zeros indicate nonbuilding grids) is then output in a plain text format, which is required for PALM simulation and is easily readable by users.

The primary objective of this dataset is not to replicate the exact flow conditions of individual neighborhoods but to capture a diverse range of horizontal and vertical heterogeneity

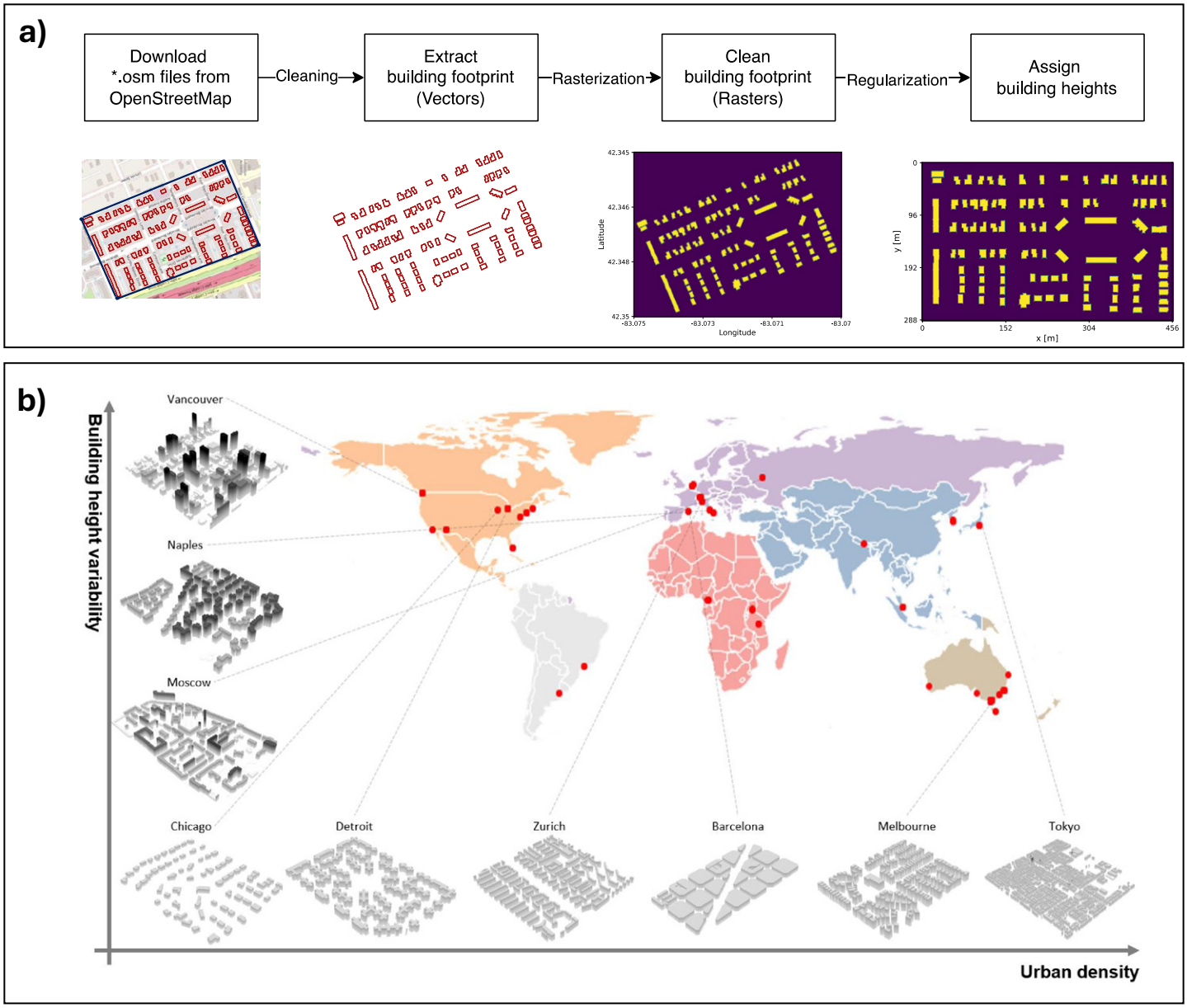


Fig. 2. (a) Schematic diagram of downloading, rasterization, and regularization of building footprints using OSM2LES to prepare urban topography for LES simulation. (b) Geographic representation of realistic neighborhoods selected for LES simulations. We consider both uniform and variable height scenarios to ensure that we cover both horizontal and vertical heterogeneities that enable comparison with idealized cases (Lu et al. 2023a).

(Fig. 1) to assess the impact of urban morphology on airflow and turbulence fields, further informing the development of urban canopy models. As such, urban roughness elements other than buildings, such as trees, vehicles, and temporary structures, were not included in the flow simulations and left for future extensions of the dataset. Similarly, while surface elevation can fundamentally influence urban canopy flow, it is challenging to generalize and was therefore excluded from the urban topography in this study.

A total of 314 realistic simulation cases are considered in UrbanTALES, of which 160 feature uniform building heights primarily used to study the impact of horizontal heterogeneity (Lu et al. 2023a), and 154 incorporate variable height data. These cases are taken from a diverse range of cities spanning 22 countries (Fig. 2b) ensuring a representative global dataset. The simulations span a wide range of packing and frontal area densities shown in Fig. 1, providing a robust foundation for investigating urban aerodynamics in realistic urban configurations.

3. UrbanTALES data processing and dissemination

In this section, we describe the postprocessing of PALM simulation outputs. This step is critical due to the substantial data volume, exceeding 35 TB in its original format. To manage this large dataset and provide quality-controlled, relevant information, we employ standard time- and spatial-averaging procedures. Time-averaging is performed over intervals longer than the scale of the slowest eddies, while spatial averaging is conducted over lengths greater than the typical spatial deviations in the flow based on the desired applications (Raupach and Shaw 1982; Finnigan 1985).

All cases use a consistent time-averaging period, where outputs are averaged over 7 h following a 4-h spinup phase (detailed in section 2a). The adequacy of the spinup and averaging intervals is validated by analyzing the time series of total kinetic energy (KE) across all 538 cases in UrbanTALES (Fig. 3). A quasi-steady state is achieved in all simulations, as indicated by the stabilization of volume-averaged KE after the spinup period. Realistic urban geometries, which generate less KE than idealized building arrays, reach a quasi-steady state more quickly. This is likely due to realistic morphologies having higher roughness, slowing down and stabilizing the flow faster, while idealized configurations have a uniform repeating unit that creates persistent streamwise motion (Coceal et al. 2007). For less than 1% of cases—three cases with the highest density of 0.64 in idealized configurations—a quasi-steady state is achieved slightly later, but still within the 4-h spinup period. These results confirm the appropriateness of the chosen time intervals and eddy turnover time for ensuring robust and reliable simulation outputs.

Figure 4 provides an overview of the postprocessed data in UrbanTALES and illustrates how the released outputs are categorized to support urban airflow analyses and the development or refinement of urban canopy models. We share a metadata file for the entire UrbanTALES dataset that captures statistical morphology information for each simulation domain. This encompasses configurations and neighborhood characteristics (for idealized cases), urban densities (λ_p, λ_p, and γ), building height statistics (H_{min}, H_{max}, and H_σ), and prevailing wind directions. For realistic urban neighborhoods, additional location information is provided, including the city, country, and neighborhood coordinates (latitude and

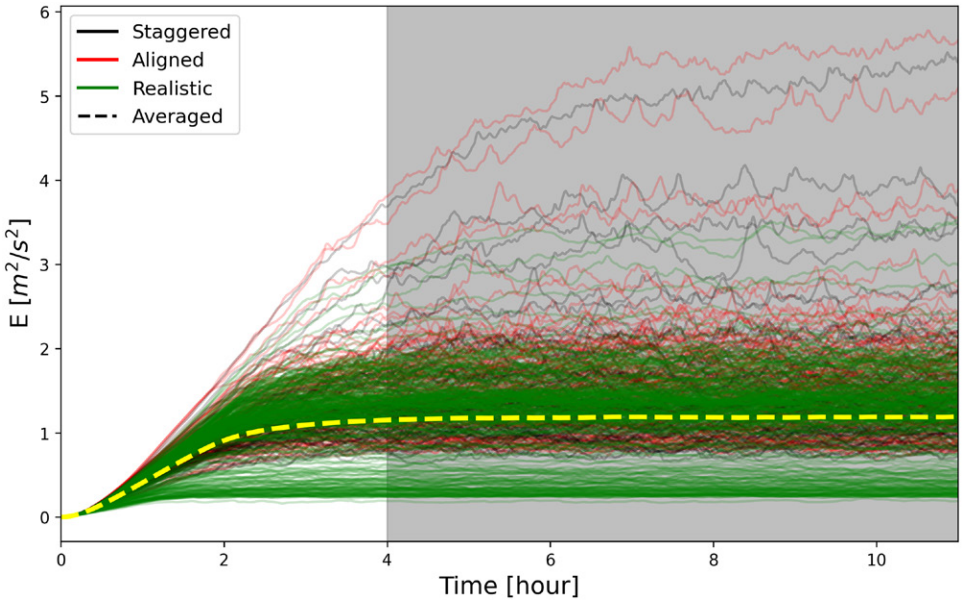


Fig. 3. Time series of total KE indicating the quasi-steady state and the time-averaging window from the 4 to 11 h. The color of the lines reflects the urban geometry configuration, where black, red, and green lines (with transparency to distinguish different cases) represent staggered, aligned, and realistic, respectively. The yellow dotted line represents the 538-cases-averaged time series.

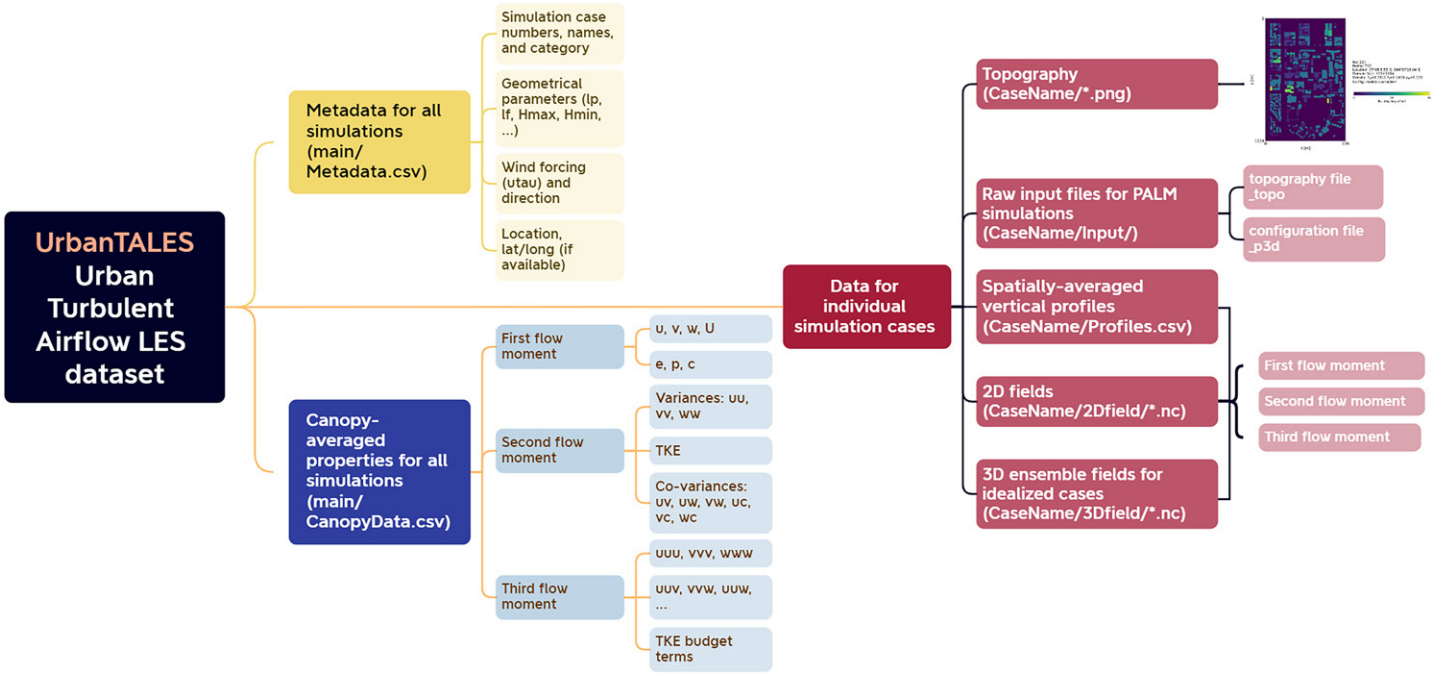


FIG. 4. Overview of the UrbanTALES dataset file structure. The metadata datasheet and canopy-averaged flow statistics are provided at the whole dataset level. The topography information (including visualized urban geometry along with some geometrical details), PALM input, vertical profiles, pedestrian-level flow field, and ensemble 3D field are provided at the individual case level.

longitude). An example metadata file is available in the appendix, with a detailed definition of each parameter included in the *readme.md* file within the folder of each simulation case. In each individual simulation folder, we also include the input files required to reproduce the simulations in PALM. These consist of the **p3d* namelist file and the 2D urban topography file (**topo*) that are in PALM input formats. This ensures the reproducibility of the flow field simulations.

Simulation outputs are offered with varying levels of spatial averaging to support a range of urban flow modeling applications. The temporally averaged 3D flow field is horizontally averaged to derive vertical profiles, canopy averaged to evaluate the cumulative effects of urban morphology, and sampled at the pedestrian level for urban wind studies (Fig. 4). Canopy-averaged flow statistics are provided across the dataset. The definition of urban canopy varies in the literature (Castro 2017). Here, we use a simple urban canopy definition that spans vertical levels from the ground to H_{mean}, and volume-averaging is applied to the 3D flow field. The outputs include first-order flow moments such as velocity components (u, v, w), wind speed (U), pressure (p), and scalar (c); second-order flow moments such as variances (TKE components) and covariances (e.g., Reynolds stresses); and third-order flow moments required for calculating TKE budget terms. These metrics collectively represent the aggregate effects of buildings on the flow field (Martilli et al. 2002) and are essential for urban canopy parameterizations (Lu et al. 2024).

For vertical profiles, we used intrinsic spatial averaging (Mignot et al. 2008) over airflow where averaging is performed over the part of the domain occupied by air. This approach may cause discontinuities in vertical profiles (e.g., TKE profiles in Fig. 8) at the heights where building density changes but can be corrected by introducing additional terms to account for density variations (Blunn et al. 2022). Similar to variables considered in the canopy-averaged process, we include the first, second, and third flow moments but extend the dataset with other related flow statistics such as turbulence length scales and eddy diffusivity for momentum (Nazarian et al. 2020; Lu et al. 2024).

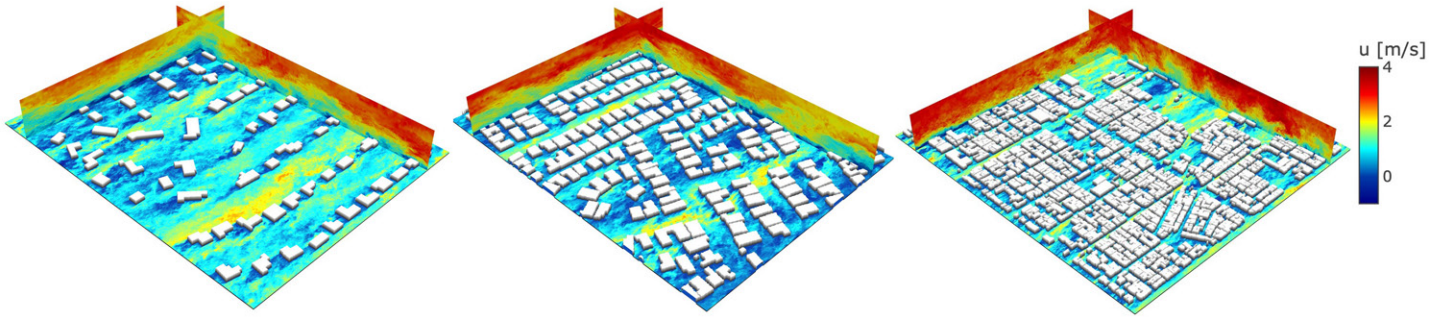


Fig. 5. Three examples of instantaneous wind fields over realistic neighborhoods from Northfield Township, United States (λ_p = 0.131); Melbourne, Australia (λ_p = 0.342); and Tokyo, Japan (λ_p = 0.590), respectively. The horizontal plane is sampled at the pedestrian height (z = 1.75 m). Two vertical cross sections at positions x = y = 55 m are also shown. Full animations are available online (see the data availability statement).

For pedestrian-level analyses, the horizontal flow field was sampled at a grid level of z = 1.75 m. Figure 5 illustrates wind speed patterns over sparse, medium, and dense urban configurations as examples of these outputs. For rigorous fluid dynamic evaluation of the dataset, UrbanTALES also includes the ensemble-average 3D field over repeating units for idealized configurations with uniform building height (Nazarian et al. 2020). Additional 3D datasets will be available upon request as they are large datasets that cannot be transferred for all simulation cases.

Web-based data platform. The extensive scope (Fig. 4) and case count of UrbanTALES could make it challenging for users to locate and get access to a subset of the dataset that fits their analyses and applications. To address this, we developed a web-based data platform for UrbanTALES (accessible through https://urbantales.climate-resilientcities.com/), which allows users to filter the data based on options such as horizontal and vertical urban configurations, prevailing wind directions, and density ranges.

4. The UrbanTALES dataset

In this section, we show three examples of commonly produced analyses enabled by UrbanTALES. Figure 6 reproduced Fig. 11 from Lu et al. (2023a), focused on canopy-averaged flow properties (average over the air volume between the ground and mean building height) such as normalized mean wind speed ($\langle\overline{U}\rangle/u_\tau$), TKE ($\langle\overline{k}\rangle/u_\tau^2$), Reynolds stress ($-\sqrt{\langle\overline{u'w'}\rangle+\langle\overline{v'w'}\rangle}/u_\tau^2$), and dispersive fluxes ($-\sqrt{\langle\tilde{u}\tilde{w}\rangle+\langle\tilde{v}\tilde{w}\rangle}/u_\tau^2$). These flow parameters are shown against common density parameters such as λ_p and λ_f, as well as the newly introduced street connectivity parameter (γ) that was shown to better depict the canopy-averaged flow statistics in Lu et al. (2023a). Specifically, γ quantifies the connectivity of airflow as it goes through the building gaps along the prevailing wind direction.

A clear distinction emerges between the behavior of realistic urban neighborhoods and idealized urban layouts commonly used to inform urban canopy parameterizations. Realistic configurations exhibit weaker and less consistent relationships with the conventional density metrics (λ_p and λ_f) and frequently deviate from regression lines fitted to the idealized datasets. In contrast, the street connectivity parameter (γ), which quantifies the uninterrupted distance urban airflow can travel without being blocked or diverted, demonstrates a stronger correlation across the entire dataset. While a detailed quantitative assessment of these relationships lies beyond the scope of this manuscript, the scatterplots underscore the utility of canopy-averaged properties for urban airflow analyses facilitated by the UrbanTALES dataset.

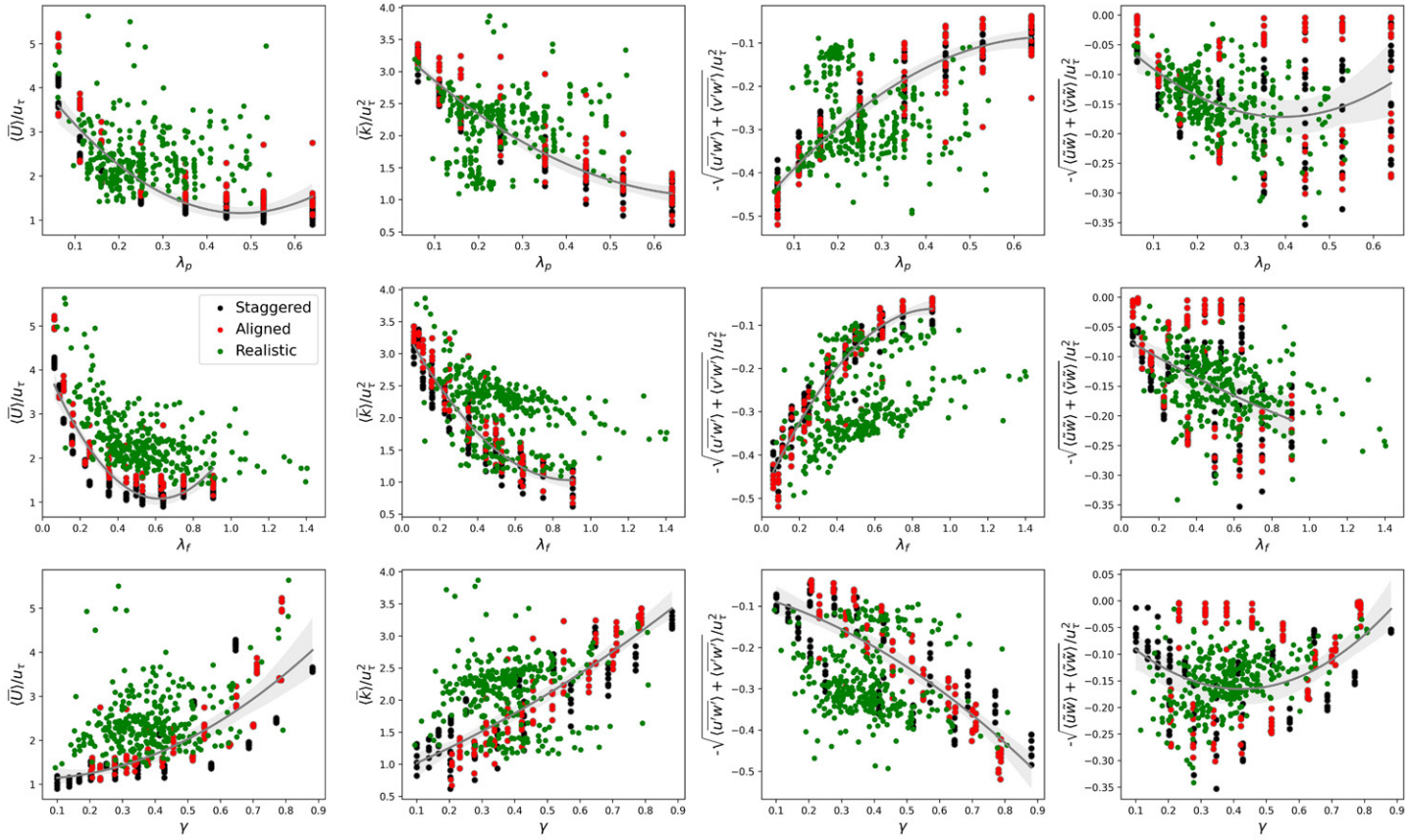


FIG. 6. Scatterplots of canopy-averaged flow statistics in UrbanTALES against plan and frontal area density and street connectivity (Lu et al. 2023a). The *y* axis in four columns represents normalized mean wind speed ($\langle\overline{U}\rangle/u_\tau$), TKE ($\langle\overline{k}\rangle/u_\tau^2$), Reynolds stress ($-\sqrt{\overline{\langle u'w'\rangle}+\overline{\langle v'w'\rangle}}/u_\tau^2$), and dispersive fluxes ($-\sqrt{\langle\widetilde{u}\widetilde{w}\rangle+\langle\widetilde{v}\widetilde{w}\rangle}/u_\tau^2$), respectively. We used colors to distinguish aligned (red), staggered (black), and realistic (green) urban geometrical configurations. In each subplot, a gray line shows the fitting for a semi-logarithmic regression over only idealized urban configurations.

Figure 7 further compares aerodynamic parameters roughness length (z_0) and displacement height (d) by reproducing Fig. 10 from Kanda et al. (2013), which refined the urban aerodynamic parameterization introduced by Macdonald et al. (1998) using data from Tokyo neighborhoods. Despite differences in geometric conditions, as illustrated in Fig. 1, UrbanTALES demonstrates strong agreement with the formulations proposed by Kanda et al. (2013). For realistic urban configurations, UrbanTALES aligns more closely with Kanda's formulations than the cases originally assessed in Kanda et al. (2013). This improved fit underscores the importance of integrating more information on vertical heterogeneities in the urban environment. Furthermore, it signals the reliability of the geometry extraction process from OSM2LES and the numerical configurations described in section 2.

Finally, Fig. 8 shows an example of the vertical profiles of flow parameters (as shown in Fig. 6). The entire dataset exhibits significant variability in the values and shapes of these profiles, as represented by the faded lines, reflecting differences in configurations, urban densities (λ_p), building height statistics (mean, maximum, and standard deviation), and wind direction. Data are clustered based on density (λ_p), highlighting how denser urban neighborhoods tend to slow the flow within the canopy, reduce TKE and Reynolds stress, and increase dispersive fluxes. This clustering approach provides insights into how urban morphology influences vertical transport behavior across various configurations. Understanding the vertical distribution of flow characteristics within and above the urban canopy layer is critical to deriving parameterization formulas for urban canopy flow. The shape of the flow profile is particularly significant, as it directly influences the predictions of the wind speeds

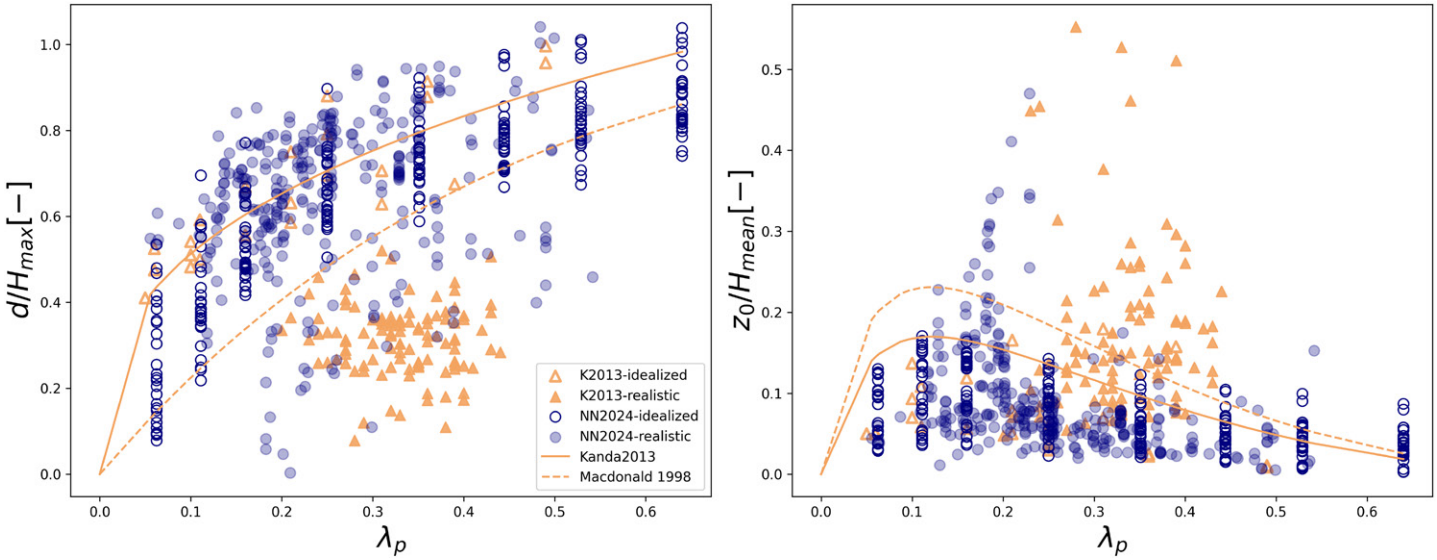


Fig. 7. Scatterplots of urban roughness flow parameters [roughness length (z_0) and displacement height (d)] in UrbanTALES against plan area density λ_p. We used colors and markers to distinguish data points from Kanda et al. (2013) (sandy brown with hollow and solid triangles for idealized and realistic configurations) and from UrbanTALES (navy blue with hollow and solid circles for idealized and realistic configurations). In each subplot, the solid and dashed lines are showing the regression formulas from Kanda et al. (2013) and Macdonald et al. (1998), respectively.

in the canopy, which are vital to understanding ventilation and the dispersion of pollutants in urban areas (Castro 2017). Furthermore, wind speed and TKE profiles play a key role in the parameterization of turbulence length scales, a crucial aspect to accurately modeling the exchange of momentum and energy between the urban surface and the atmosphere (Coceal and Belcher 2004; Nazarian et al. 2020).

5. Conclusions: UrbanTALES application

This study presents UrbanTALES, a comprehensive and open-access large-eddy simulation (LES) dataset available to the urban climate research community. UrbanTALES provides spatially averaged, 2D, and 3D data reflecting urban flow statistics, offering unprecedented opportunities for assessing the impact of form on urban canopy processes. The dataset required significant computational resources and storage capacity, over 3 000 000 CPU hour

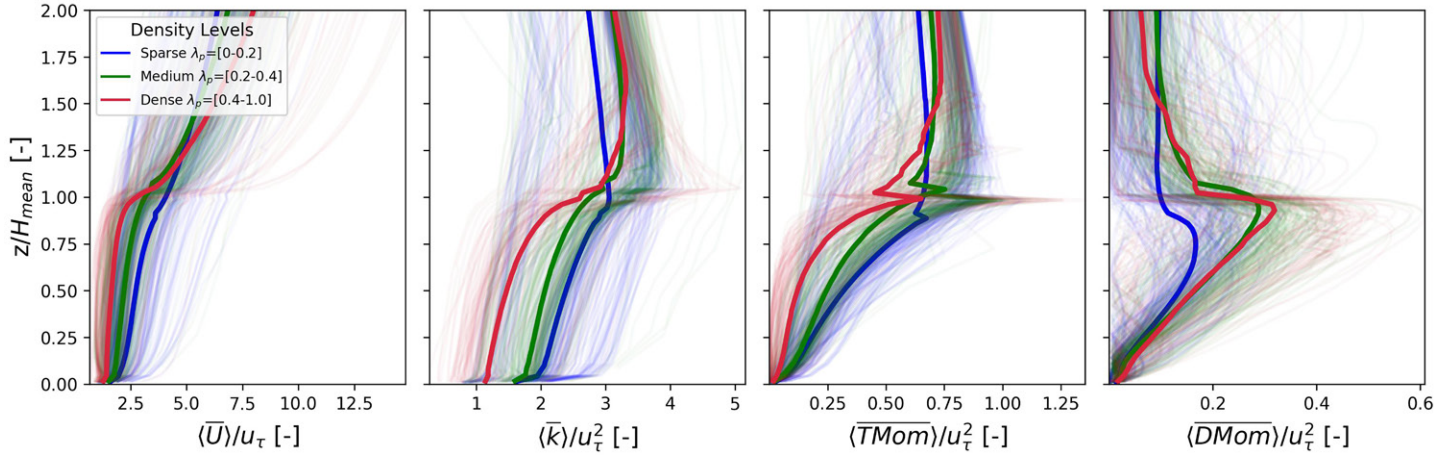


Fig. 8. Vertical profiles up to $2H_{mean}$ of wind speed $U=\sqrt{u^2+v^2}$, TKE $\overline{k}=(1/2)\left(\overline{u'u'}+\overline{v'v'}+\overline{w'w'}\right)$, and total turbulent and dispersive momentum flux $\mathbf{TMom}=\sqrt{\overline{u'w'}^2+\overline{v'w'}^2}$ $\mathbf{DMom}=\sqrt{\tilde{u}\tilde{w}^2+\tilde{v}\tilde{w}^2}$ normalized by friction velocity, respectively. We separated UrbanTALES into three categories—sparse (cyan), medium (orange), and dense (red) urban configurations based on the plane area density λ_p, which is shown for individual cases in the background with high transparency. The category-averaged vertical profiles are shown in solid lines with the same coloring.

and 35 TB, respectively, reflecting the meticulous effort involved in creating a high-resolution, large-scale LES dataset. To ensure that high computational costs are utilized effectively, UrbanTALES outputs are released with close attention to key gaps in urban climate research, facilitating future advancements.

The primary objective of UrbanTALES is manifold. First, we aim to facilitate systematic assessments of the impact of urban form on urban canopy processes. Second, this dataset provides valuable insights for various urban canopy models, including simplified tiled, single-layer, and multilayer urban canopy models currently developed in the field (Lipson et al. 2024). These models are instrumental in offline assessments of urban climate as well as in representing land–atmosphere exchanges over cities in regional model integrations. In particular, UrbanTALES introduces a higher-order database of urban turbulence, recording third flow moments to enable evaluation of TKE and momentum flux budgets (Akinlabi et al. 2022). These budget terms provide a critical foundation for developing higher-order turbulence closure schemes in urban canopy parameterizations (Wilson 1988).

Finally, the dataset's applications extend to machine learning, supporting 2D pedestrian-level wind environment predictions (Lu et al. 2025) and 3D urban-atmosphere flux predictions (Y. Lu et al. 2023).

Looking ahead, the UrbanTALES dataset is set to expand its scope. Future extensions will include additional urban processes such as solar heating (Nazarian and Kleissl 2015), the presence of street trees (Lindberg et al. 2018), and the transport of moisture and passive scalars. Integrating indoor urban climate dynamics is also planned, paving the way for broader applications of UrbanTALES. These enhancements will further its applicability beyond urban canopy parameterization to integrated urban services (IUS) (Esau et al. 2024) and multiphysics urban models (Krayenhoff et al. 2020), addressing a wider array of urban climate phenomena.

Furthermore, we are exploring ways to transform UrbanTALES into a collaborative platform, enabling researchers worldwide to contribute their LES datasets. This initiative aims to increase the dataset's reach and impact, offering global access to high-quality, computationally intensive model outputs. While automating the integration process will require substantial effort, we welcome inquiries and partnerships with research groups that have existing LES outputs and are interested in incorporating their data into the platform or future UrbanTALES releases. By enhancing accessibility and fostering collaboration, UrbanTALES has the potential to become a foundational resource for the urban climate research community.

Acknowledgments. This work was supported by the Australian Research Council as part of the Centre of Excellence for Climate Extremes (CE170100023) and the Centre of Excellence for the Weather of the 21st Century (CE230100012), as well as the Natural Sciences and Engineering Research Council of Canada Discovery Grant. Simulations were undertaken with the assistance of resources and services from the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS), with access to computational resources provided by the National Computational Infrastructure (NCI), which is supported by the Australian Government through the National Computational Merit Allocation Scheme.

Data availability statement. All data files described in this article are shared publicly at https://urbantales.climate-resilientcities.com/. This dataset is over 2 terabytes and includes all input files for model configuration as well as output data files and can be downloaded for individual cases. PALM, an open-source numerical model, is used here (Maronga et al. 2020). Version r4554 of the model is used with minor changes for outputs and is provided at https://unsw-my.sharepoint.com/:u:/g/personal/z3524567_ad_unsw_edu_au/EWmp3oHh6vVAnsLHAcME5O4B1Uy2RjP23-AdlTuCHDGfIA?e=f3tVPW as *PALM-r4554.tar*. For further transparency, the input files (for both model configuration and

topography) were also provided in the dataset folder of each use case. The OSM2LES python package for obtaining the topography of realistic neighborhoods is available at https://pypi.org/project/OSM2LES/.

APPENDIX

Metadata for UrbanTALES

Table A1 shows the sample metadata file for both realistic and idealized building arrays in UrbanTALES.

TABLE A1. Metadata for 314 realistic urban neighborhoods and 224 idealized building arrays in UrbanTALES. For realistic cases, the site name first reflects the country, and the city name by starting from a two-letter country name that follows ISO Country Codes and then followed by the first three letters of the city names. The height configuration is reflected as "*U*" for uniform height and "*V*" for variable cases, which is followed by a number counting the cases that fall into the category. The wind direction configuration is reflected as a two-digit number after "_*d*". For idealized configurations, the height configuration is reflected at the beginning of the names in the same manner as for the realistic cases. The horizontal arrangement can be "staggered – S" or "aligned – A" in the following and followed by the plane area density, in two digits (25 for 0.25, for example). Then, the name reflects building height distribution in abbreviation as described in section 1, which is finished by the prevailing wind direction indicator.

Site name	City	Country	λ_p	λ_f	γ_m	WD	H_{mean}	H_σ	H_{max}	H_{min}	SAR	Lon	Lat
US25_d00	—	—	0.25	0.25	0.25	0	16.0	0.0	16.0	16.0	S	—	—
VS25-HR-28_d00	—	—	0.25	0.25	0.25	0	16.0	2.8	26.5	8.0	S	—	—
VA25-SM-56_d45	—	—	0.25	0.35	0.52	45	16.0	5.6	24.0	8.0	A	—	—
..													
AU-Syd-U1_d00	Sydney	AU	0.25	0.88	0.41	0	16.0	0.0	16	16.0	R	151.25	–33.93
AU-Syd-V1_d00	Sydney	AU	0.33	0.38	0.36	0	17.8	8.8	50	3.0	R	151.22	–33.92
AU-Mel-V1_d00	Melbourne	AU	0.32	0.26	0.33	0	13.4	5.7	36.0	3.0	R	145.14	–37.91
U.S.-Det-U1_d00	Detroit	U.S.	0.09	0.32	0.64	0	16.0	0.0	16.0	16.0	R	–83.04	42.36
CH-Zur-U1_d00	Zurich	CH	0.25	0.59	0.41	0	16.0	0.0	16.0	16.0	R	8.56	47.41

References

Akinlabi, E., B. Maronga, M. G. Giometto, and D. Li, 2022: Dispersive fluxes within and over a real urban canopy: A large-eddy simulation study. *Bound.-Layer Meteor.*, **185**, 93–128, https://doi.org/10.1007/s10546-022-00725-6.

Biljecki, F., H. Ledoux, and J. Stoter, 2016: An improved LOD specification for 3D building models. *Comput. Environ. Urban Syst.*, **59**, 25–37, https://doi.org/10.1016/j.compenvurbsys.2016.04.005.

Blunn, L. P., O. Coceal, N. Nazarian, J. F. Barlow, R. S. Plant, S. I. Bohnenstengel, and H. W. Lean, 2022: Turbulence characteristics across a range of idealized urban canopy geometries. *Bound.-Layer Meteor.*, **182**, 275–307, https://doi.org/10.1007/s10546-021-00658-6.

Brown, M., R. Lawson, D. DeCroix, and R. Lee, 2001: Comparison of centerline velocity measurements obtained around 2D and 3D building arrays in a wind tunnel. *Int. Society of Environmental Hydraulics Conf.*, Tempe, AZ, International Society of Environmental Hydraulics, https://www.osti.gov/servlets/purl/783425.

Buccolieri, R., J. L. Santiago, and A. Martilli, 2021: CFD modelling: The most useful tool for developing mesoscale urban canopy parameterizations. *Build. Simul.*, **14**, 407–419, https://doi.org/10.1007/s12273-020-0689-z.

Castro, I. P., 2017: Are urban-canopy velocity profiles exponential? *Bound.-Layer Meteor.*, **164**, 337–351, https://doi.org/10.1007/s10546-017-0258-x.

Coceal, O., and S. E. Belcher, 2004: A canopy model of mean winds through urban areas. *Quart. J. Roy. Meteor. Soc.*, **130**, 1349–1372, https://doi.org/10.1256/qj.03.40.

——, A. Dobre, T. G. Thomas, and S. E. Belcher, 2007: Structure of turbulent flow over regular arrays of cubical roughness. *J. Fluid Mech.*, **589**, 375–409, https://doi.org/10.1017/S002211200700794X.

Esau, I., and Coauthors, 2024: A city-scale turbulence-resolving model as an essential element of integrated urban services. *Urban Climate*, **56**, 102059, https://doi.org/10.1016/j.uclim.2024.102059.

Esch, T., and Coauthors, 2022: World settlement footprint 3D - A first three-dimensional survey of the global building stock. *Remote Sens. Environ.*, **270**, 112877, https://doi.org/10.1016/j.rse.2021.112877.

Finnigan, J. J., 1985: Turbulent transport in flexible plant canopies. *The Forest-Atmosphere Interaction*, Springer, 443–480, https://doi.org/10.1007/978-94-009-5305-5_28.

Geletič, J., M. Lehnert, P. Krč, J. Resler, and E. S. Krayenhoff, 2021: High-resolution modelling of thermal exposure during a hot spell: A case study using PALM-4U in Prague, Czech Republic. *Atmosphere*, **12**, 175, https://doi.org/10.3390/atmos12020175.

Giometto, M. G., A. Christen, P. Egli, M. Schmid, R. Tooke, N. Coops, and M. Parlange, 2017: Effects of trees on mean wind, turbulence and momentum exchange within and above a real urban environment. *Adv. Water Resour.*, **106**, 154–168, https://doi.org/10.1016/j.advwatres.2017.06.018.

——, ——, C. Meneveau, J. Fang, M. Krafczyk, and M. B. Parlange, 2016: Spatial characteristics of roughness sublayer mean flow and turbulence over a realistic urban surface. *Bound.-Layer Meteor.*, **160**, 425–452, https://doi.org/10.1007/s10546-016-0157-6.

Herfort, B., S. Lautenbach, J. Porto De Albuquerque, J. Anderson, and A. Zipf, 2023: A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. *Nat. Commun.*, **14**, 3985, https://doi.org/10.1038/s41467-023-39698-6.

Kanda, M., A. Inagaki, T. Miyamoto, M. Gryschka, and S. Raasch, 2013: A new aerodynamic parametrization for real urban surfaces. *Bound.-Layer Meteor.*, **148**, 357–377, https://doi.org/10.1007/s10546-013-9818-x.

Krayenhoff, E. S., and Coauthors, 2020: A multi-layer urban canopy meteorological model with trees (BEP-Tree): Street tree impacts on pedestrian-level climate. *Urban Climate*, **32**, 100590, https://doi.org/10.1016/j.uclim.2020.100590.

Lim, H. D., D. Hertwig, T. Grylls, H. Gough, M. V. Reeuwijk, S. Grimmond, and C. Vanderwel, 2022: Pollutant dispersion by tall buildings: Laboratory experiments and large-eddy simulation. *Exp. Fluids*, **63**, 92, https://doi.org/10.1007/s00348-022-03439-0.

Lindberg, F., and Coauthors, 2018: Urban Multi-scale Environmental Predictor (UMEP): An integrated tool for city-based climate services. *Environ. Modell. Software*, **99**, 70–87, https://doi.org/10.1016/j.envsoft.2017.09.020.

Lipson, M. J., N. Nazarian, M. A. Hart, K. A. Nice, and B. Conroy, 2022: A transformation in city-descriptive input data for urban climate models. *Front. Environ. Sci.*, **10**, 866398, https://doi.org/10.3389/fenvs.2022.866398.

——, and Coauthors, 2024: Evaluation of 30 urban land surface models in the urban-PLUMBER project: Phase 1 results. *Quart. J. Roy. Meteor. Soc.*, **150**, 126–169, https://doi.org/10.1002/qj.4589.

Lu, J., W. Li, S. Hobeichi, S. A. Azad, and N. Nazarian, 2025: Machine learning predicts pedestrian wind flow from urban morphology and prevailing wind direction. *Environ. Res. Lett.*, **20**, 054006, https://doi.org/10.1088/1748-9326/adc148.

——, N. Nazarian, and M. A. Hart, 2022: OSM2LES - A Python-based tool to prepare realistic urban geometry for LES simulation from OpenStreetMap Version 0.1.0. Zenodo, https://doi.org/10.5281/zenodo.6566346.

——, ——, ——, E. S. Krayenhoff, and A. Martilli, 2023a: Novel geometric parameters for assessing flow over realistic versus idealized urban arrays. *J. Adv. Model. Earth Syst.*, **15**, e2022MS003287, https://doi.org/10.1029/2022MS003287.

——, ——, ——, ——, and ——, 2023b: Representing the effects of building height variability on urban canopy flow. *Quart. J. Roy. Meteor. Soc.*, **150**, 46–67, https://doi.org/10.1002/qj.4584.

——, ——, ——, ——, and ——, 2024: A one-dimensional urban flow model with an eddy-diffusivity mass-flux (EDMF) scheme and refined turbulent transport (MLUCM v3.0). *Geosci. Model Dev.*, **17**, 2525–2545, https://doi.org/10.5194/gmd-17-2525-2024.

Lu, Y., X.-H. Zhou, H. Xiao, and Q. Li, 2023: Using machine learning to predict urban canopy flows for land surface modeling. *Geophys. Res. Lett.*, **50**, e2022GL102313, https://doi.org/10.1029/2022GL102313.

Macdonald, R., R. Griffiths, and D. Hall, 1998: An improved method for the estimation of surface roughness of obstacle arrays. *Atmos. Environ.*, **32**, 1857–1864, https://doi.org/10.1016/S1352-2310(97)00403-2.

Maronga, B., and Coauthors, 2015: The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and oceanic flows: Model formulation, recent developments, and future perspectives. *Geosci. Model Dev.*, **8**, 2515–2551, https://doi.org/10.5194/gmd-8-2515-2015.

——, and Coauthors, 2020: Overview of the PALM model system 6.0. *Geosci. Model Dev.*, **13**, 1335–1372, https://doi.org/10.5194/gmd-13-1335-2020.

Martilli, A., A. Clappier, and M. W. Rotach, 2002: An urban surface exchange parameterisation for mesoscale models. *Bound.-Layer Meteor.*, **104**, 261–304, https://doi.org/10.1023/A:1016099921195.

McNorton, J. R., and Coauthors, 2021: An urban scheme for the ECMWF Integrated Forecasting System: Single-column and global offline application. *J. Adv. Model. Earth Syst.*, **13**, e2020MS002375, https://doi.org/10.1029/2020MS002375.

Mignot, E., E. Barthélemy, and D. Hurther, 2008: Turbulent kinetic energy budget in a gravel-bed channel flow. *Acta Geophys.*, **56**, 601–613, https://doi.org/10.2478/s11600-008-0020-3.

Nagel, T., R. Schoetter, V. Bourgin, V. Masson, and E. Onofri, 2023: Drag coefficient and turbulence mixing length of local climate zone-based urban morphologies derived using obstacle-resolving modelling. *Bound.-Layer Meteor.*, **186**, 737–769, https://doi.org/10.1007/s10546-022-00780-z.

Nazarian, N., and J. Kleissl, 2015: CFD simulation of an idealized urban environment: Thermal effects of geometrical characteristics and surface materials. *Urban Climate*, **12**, 141–159, https://doi.org/10.1016/j.uclim.2015.03.002.

——, A. Martilli, and J. Kleissl, 2018a: Impacts of realistic urban heating, Part I: Spatial variability of mean flow, turbulent exchange and pollutant dispersion. *Bound.-Layer Meteor.*, **166**, 367–393, https://doi.org/10.1007/s10546-017-0311-9.

——, ——, L. Norford, and J. Kleissl, 2018b: Impacts of realistic urban heating. Part II: Air quality and city breathability. *Bound.-Layer Meteor.*, **168**, 321–341, https://doi.org/10.1007/s10546-018-0346-6.

——, E. S. Krayenhoff, and A. Martilli, 2020: A one-dimensional model of turbulent flow through "urban" canopies (MLUCM v2.0): Updates based on large-eddy simulation. *Geosci. Model Dev.*, **13**, 937–953, https://doi.org/10.5194/gmd-13-937-2020.

Oke, T. R., G. Mills, A. Christen, and J. A. Voogt, 2017: *Urban Climates.* Cambridge University Press, 525 pp., https://doi.org/10.1017/9781139016476.

OpenStreetMap Contributors, 2017: Planet dump. https://planet.osm.org.

Park, S.-B., J.-J. Baik, and B.-S. Han, 2015: Large-eddy simulation of turbulent flow in a densely built-up urban area. *Environ. Fluid Mech.*, **15**, 235–250, https://doi.org/10.1007/s10652-013-9306-3.

PSMA Australia, 2020: Geoscape Buildings v2.0 (2020). https://geoscape.com.au/.

Raupach, M. R., and R. H. Shaw, 1982: Averaging procedures for flow within vegetation canopies. *Bound.-Layer Meteor.*, **22**, 79–90, https://doi.org/10.1007/BF00128057.

Santiago, J. L., and A. Martilli, 2010: A dynamic urban canopy parameterization for mesoscale models based on computational fluid dynamics Reynolds-averaged Navier–Stokes microscale simulations. *Bound.-Layer Meteor.*, **137**, 417–439, https://doi.org/10.1007/s10546-010-9538-4.

——, E. S. Krayenhoff, and A. Martilli, 2014: Flow simulations for simplified urban configurations with microscale distributions of surface thermal forcing. *Urban Climate*, **9**, 115–133, https://doi.org/10.1016/j.uclim.2014.07.008.

Simón-Moral, A., J. L. Santiago, and A. Martilli, 2016: Effects of unstable thermal stratification on vertical fluxes of heat and momentum in urban areas. *Bound.-Layer Meteor.*, **163**, 103–121, https://doi.org/10.1007/s10546-016-0211-4.

Stewart, I. D., and T. R. Oke, 2012: Local climate zones for urban temperature studies. *Bull. Amer. Meteor. Soc.*, **93**, 1879–1900, https://doi.org/10.1175/BAMS-D-11-00019.1.

Sützl, B. S., G. G. Rooney, and M. van Reeuwijk, 2021: Drag distribution in idealized heterogeneous urban environments. *Bound.-Layer Meteor.*, **178**, 225–248, https://doi.org/10.1007/s10546-020-00567-0.

Toparlar, Y., B. Blocken, B. Maiheu, and G. Van Heijst, 2017: A review on the CFD analysis of urban microclimate. *Renewable Sustainable Energy Rev.*, **80**, 1613–1640, https://doi.org/10.1016/j.rser.2017.05.248.

Wilson, J. D., 1988: A second-order closure model for flow through vegetation. *Bound.-Layer Meteor.*, **42**, 371–392, https://doi.org/10.1007/BF00121591.

Xie, Z.-T., O. Coceal, and I. P. Castro, 2008: Large-eddy simulation of flows over random urban-like obstacles. *Bound.-Layer Meteor.*, **129**, 1–23, https://doi.org/10.1007/s10546-008-9290-1.

Yaghoobian, N., J. Kleissl, and K. Tha, 2014: An improved three-dimensional simulation of the diurnally varying street-canyon flow. *Bound.-Layer Meteor.*, **153**, 251–276, https://doi.org/10.1007/s10546-014-9940-4.