



# Multi-scale Analysis of Flow over Heterogeneous Urban Environments

Maarten van Reeuwijk<sup>1</sup> · Jingzi Huang<sup>1</sup>

Received: 5 April 2025 / Accepted: 16 September 2025 / Published online: 10 October 2025  
© The Author(s) 2025

## Abstract

A computationally efficient multi-scale planar-averaging framework for urban areas is developed, which enables efficient computation of coarse-grained velocity and scalar fields. We apply the multi-scale framework to a large-eddy simulation of an idealised heterogeneous urban environment of 512 buildings based on a typical London height distribution. We observe that for this geometry, the characteristic urban lengthscale  $\ell \approx 50$  m, which is the averaging lengthscale  $L$  at which as much variance in the mean flow is resolved as is unresolved. For  $L > 400$  m, the statistics become approximately homogeneous, suggesting that non-building-resolving numerical weather prediction (NWP) models can be applied without modification at resolutions of 400 m and above for the case under consideration. We derive the multi-scale planar- and Reynolds-averaged momentum equation and show that for neutral cases, NWP models require parameterisation of the distributed drag and the unresolved turbulence and dispersive stress. An a priori analysis reveals that the drag parameterisation from Sützl et al. (Bound-Layer Meteorol 178:225–248, 2020) holds reasonably well for resolutions  $L$  above 200 m. Below this value, the problem becomes inhomogeneous and the parameterisation works less well. The unresolved stresses are well represented by a  $k - \omega$  closure with a value of  $\omega = 0.4 \text{ s}^{-1}$ . However, an even more accurate closure can be derived from the Sützl drag parameterisation that does not require further turbulence information.

**Keywords** Drag parameterisation · Heterogeneity · Multi-scale analysis · Urban canopy · Large-eddy simulation

---

✉ Maarten van Reeuwijk  
m.vanreeuwijk@imperial.ac.uk

Jingzi Huang  
jingzi.huang17@imperial.ac.uk

<sup>1</sup> Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK

## 1 Introduction

The resolution of Numerical Weather Prediction (NWP) models has advanced significantly over recent decades, transitioning from synoptic-scale  $O(10\text{ km})$  grids (Bryan et al. 2003) to convection-permitting  $O(1\text{ km})$  systems (Baldauf et al. 2011; Tang et al. 2013). This progress, driven by computational advancements, enables explicit resolution of critical atmospheric processes such as deep convection and orographic flows, reducing reliance on parameterisations (Lean et al. 2024). Kilometre-scale models are now operational for weather forecasting, regional climate downscaling (Prein et al. 2015; Belušić et al. 2020), and air quality assessments. However, a persistent limitation across NWP frameworks is the unresolved representation of urban morphology-buildings, vegetation, and surface heterogeneities that remain subgrid-scale features, necessitating parameterised approximations.

The urban canopy region is at the lowest level of the atmospheric boundary layer and is characterised by strong heterogeneity and complex interactions between the urban surface and the atmosphere. Traditional urban canopy models (UCMs) aggregate aerodynamic drag, radiative trapping, and energy exchange into bulk parameters such as roughness length, displacement height, and plan area density that connect to the lowest atmospheric level directly (Lipson et al. 2024). Whilst this approach works reasonably well for areas with low-rise buildings, it becomes problematic for areas with tall buildings that extend substantially into the atmospheric boundary layer. Vertically distributed UCMs (Schoetter et al. 2020; Sützl et al. 2021; Lu et al. 2024) improve momentum and scalar flux partitioning by distributing drag, heat over the vertical. However, both approaches assume horizontal homogeneity within grid cells, which breaks down below  $O(1\text{ km})$  resolutions, as urban landscapes exhibit heterogeneity at neighbourhood scales (about 100–500 m) due to variations in building height, street geometry, and land cover (Stewart and Oke 2012). Such fine-scale variability drives localised microclimates, including urban heat islands, airflow channelling, and pollutant dispersion (Masson et al. 2020), which are poorly captured by current parameterisations.

High-resolution simulation solves some of these problems. Recent studies demonstrate that hectometric models (HMs) improve predictions of near-surface winds and turbulence kinetic energy in urban areas by up to 30% compared to coarser models (Lean et al. 2024), highlighting their potential to reduce dependency on empirical parameterisations. Moreover, resolving neighbourhood-scale features partially enters the ‘turbulence grey zone’, where subfilter turbulent fluxes coexist with resolved eddies (Honnert et al. 2020). At these scales, models can better represent terrain-driven phenomena (e.g., cold pools in valleys, slope winds) and urban surface-atmosphere interactions (Brun et al. 2017; Smith et al. 2021).

However, at high horizontal NWP resolutions it will become necessary to relax the UCM assumption of horizontal homogeneity, and to understand how the vertical distribution depends on the horizontal resolution. In order to be able to provide the answer to this question, a multi-resolution framework is required, whereby high-resolution simulations that resolve individual buildings can be coarse-grained to arbitrary NWP resolution, such that it is possible to diagnose a priori what parameterisation is needed to represent the buildings within the NWP model. Coarse-graining is a well-established technique to study, e.g. grey-zone turbulence (Honnert et al. 2020), but there is no established multi-scale framework that can be used for the urban canopy.

Coarse-graining methods apply a spatial averaging, or filtering, operation on a fully resolved field in order to obtain a ‘smoothed’ coarse-resolution field (Wyngaard 2004; Honnert et al. 2011). Spatial averaging has been routinely used for urban canopies, inspired by the formalism presented for vegetation canopies (Raupach and Shaw 1982; Finnigan 2000; Nepf 2012). Within the context of urban canopies, Schmid et al. (2019) provide a spatial-averaging framework based on the volume-averaging framework of Whitaker (1999), which was developed for porous media flows. A unique aspect of NWP models is that near the surface (i.e. in the canopy layer), the horizontal resolution of NWP models is much larger than the vertical resolution, as variations in the vertical tend to be much stronger than in the horizontal. Indeed, typical NWP simulations have a resolution of  $O(1\text{ m})$  in the vertical but  $O(1\text{ km})$  in the horizontal; even in the context of a hectometric model, the cell aspect ratio will be at least 100. This suggests that an appropriate averaging region on the canopy layer can be defined as a thin slab, i.e. reducing a three-dimensional volume averaging to a two-dimensional planar averaging approach, which is the approach used in Schmid et al. (2019) and also for vegetation canopies (Raupach and Shaw 1982). However, the Schmid et al. (2019) framework is not straightforward to apply directly to simulation data, since analyses are typically performed using averages over the entire domain and not for averages of arbitrary length scale  $L$ . In addition, the approach is computationally inefficient and does not explicitly use a planar averaging region.

The aim of this paper is to present a computationally efficient multi-resolution framework based on the Whitaker (1999), Schmid et al. (2019) spatial averaging approach and to apply this framework to results from building-resolving large-eddy simulation. This coarse-graining approach allows us to systematically study heterogeneity in the flow statistics. In the context of NWP models, it facilitates analysis of the required drag and turbulence parameterisation for non-building resolving models at different horizontal resolutions. The paper is structured as follows. Section 2 mainly introduces a multi-resolution spatial averaging framework to grain the coarse field from the original high-resolution one; Section 3 introduces the simulation details of LES; and Section 4 presents and discusses the multi-resolution results on drag force and turbulent momentum stress, respectively. The highlights are remarked in the Section 5.

## 2 Multi-resolution Planar-Averaging Framework

In this section, we derive a multi-resolution framework in terms of convolution filters. The main advantages of using convolution filters are: (1) that it standardises the methodology, bringing it close to the large-eddy simulation formalism (Pope 2000); and (2) that convolutions can be computed efficiently using Fast-Fourier transforms (FFTs). In addition, we formulate the spatial averaging purely in terms of a planar filter. The work in Schmid et al. (2019) is cast in planar form, but the integral representing the effect of the solid boundaries is still presented in the volume-averaging form. Here, we capitalise on recent work by Van Reeuwijk et al. (2021) who derive identities that permit a fully planar formulation. A description of how the volume averaging can be transformed to planar averaging and a derivation of the spatial averaging theorem using convolutions is provided in Appendix A. The main results needed for the convolution formalism are presented here.

## 2.1 Planar Averaging Using Convolution Filters

The domain of interest  $\Omega$  is comprised of a subdomain containing the fluid  $\Omega_f$  and a solid subdomain  $\Omega_b$  such that  $\Omega = \Omega_f \cup \Omega_b$  (Fig. 1). The boundary between the fluid and solid phase is denoted  $\partial\Omega_f$ , which has a (3-D) normal vector  $\mathbf{N}$  pointing into the fluid domain. The horizontal, spanwise and vertical coordinates are denoted  $x, y$  and  $z$ , respectively. Since averaging will take place over the streamwise directions, the coordinate vector  $\mathbf{x}$  is denoted as  $\mathbf{x} = [x_h; z]$ , where  $\mathbf{x}_h = [x, y]^T$ . Denoting the filter associated with the area-averaging  $\mathcal{A}$ , the superficial average (also called comprehensive average, the average normalised by the total area including the solid phase; cf. Xie and Fuka 2018) of an arbitrary scalar  $\varphi$  is given by:

$$\langle \varphi \rangle(\mathbf{x}) \equiv \int_{\Omega_f(\mathbf{x})} \mathcal{A}(\mathbf{x}_h - \mathbf{y}_h) \varphi(\mathbf{y}_h, z) d\mathbf{y}_h. \quad (1)$$

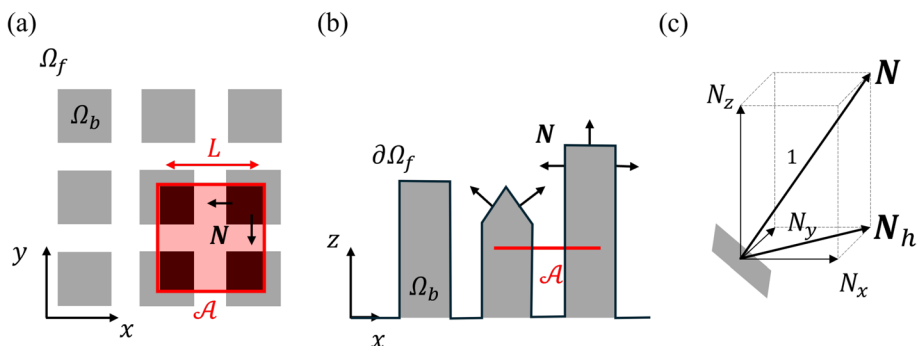
The presence of  $\mathbf{x}$  in the integration domain  $\Omega_f(\mathbf{x})$  means that the integration region comprises all the fluid-occupied areas within the support of the filter. This can be made explicit by defining a mask function  $I_f(\mathbf{x})$  which is 1 when  $\mathbf{x} \in \Omega_f$  and 0 otherwise, which allows the convolution integral above to be expressed as:

$$\langle \varphi \rangle(\mathbf{x}) = \int_{\Omega(z)} \mathcal{A}(\mathbf{x}_h - \mathbf{y}_h) I_f(\mathbf{y}_h, z) \varphi(\mathbf{y}_h, z) d\mathbf{y}_h. \quad (2)$$

This is a standard planar convolution integral over the entire domain  $\Omega$  at given  $z$  which can be evaluated efficiently using the convolution theorem and Fast-Fourier transforms (Bracewell 2000).

Formally, there are few restrictions on the filter, except that it satisfies the normalisation condition  $\int \mathcal{A}(\mathbf{x}_h) d\mathbf{x}_h = 1$ . Here, a symmetric square filter is used, defined as:

$$\mathcal{A}(x, y) \equiv \begin{cases} L^{-2} & \text{if } (-\frac{L}{2} < x < \frac{L}{2}) \cup (-\frac{L}{2} < y < \frac{L}{2}) , \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$



**Fig. 1** Definition sketch of domain and filter  $\mathcal{A}$ . **a** Plan view, **b** elevation view. The domain of  $\Omega$  is comprised of a fluid subdomain  $\Omega_f$  (in white), a solid subdomain  $\Omega_b$  (in grey), and a fluid-solid interface  $\partial\Omega_f$  (solid black lines). The building-surface 3-D normal vectors  $\mathbf{N}$  point into the fluid domain, and the decomposition is shown in (c). The 2-D square filter  $\mathcal{A}$  with averaging length  $L$  is shown in red

where  $L$  is the averaging length scale. Note that  $\langle\langle X \rangle\rangle \neq \langle X \rangle$  in general. However, if the underlying characteristic length scale of the urban surface  $\ell$  is much smaller than the averaging length scale  $L$ , that is  $L \gg \ell$ , we expect  $\langle\langle X \rangle\rangle \approx \langle X \rangle$ . In other words,  $\langle\langle X \rangle\rangle \approx \langle X \rangle$  if there is a separation of scales between the characteristic urban lengthscale  $\ell$  and the averaging lengthscale  $L$  (Schmid et al. 2019).

In the filter formalism, the area (planar) fraction  $\varepsilon = A_f/L^2$ , where  $A_f$  is the fluid surface within the averaging region, is defined as:

$$\varepsilon(\mathbf{x}) \equiv \int_{\Omega_f(\mathbf{x})} \mathcal{A}(\mathbf{x}_h - \mathbf{y}_h) d\mathbf{y}_h. \quad (4)$$

This equation shows the interaction between the filter  $\mathcal{A}$  and the integration over the fluid domain  $\Omega_f$ , which may only cover part of the area over which  $\mathcal{A}$  filters. The porosity  $\varepsilon$  can be used to calculate the intrinsic average  $\langle\varphi\rangle^f$  (the average normalised by only the area of the variable fluid phase), indicative of the average value of  $\varphi$  inside the fluid phase:

$$\langle\varphi\rangle^f = \varepsilon^{-1} \langle\varphi\rangle. \quad (5)$$

## 2.2 Differentiation Rules

Since the solid phase is stationary, the spatial averages commute with time differentiation. The same is not true for spatial differentiation. For the planar filter  $\mathcal{A}$ , the scalar spatial averaging theorem (Whitaker 1999, eqs. 1.2–15) is given by (see Appendix A):

$$\left\langle \frac{\partial \varphi}{\partial x_i} \right\rangle = \frac{\partial \langle \varphi \rangle}{\partial x_i} - \oint_{\partial \Omega_f} \mathcal{A} \varphi \frac{N_i}{|\mathbf{N}_h|} ds. \quad (6)$$

Here, the last term is a Leibniz-like horizontal line integral over the fluid-solid boundary  $\partial \Omega_f$  at given  $z$  within the support of  $\mathcal{A}$ . The negative sign of the line integral is due to the (3D) surface normals  $\mathbf{N}$  pointing into the domain (the direction of the 3D vector  $\mathbf{N}$  is shown by the arrow in Fig. 1c). The quantity  $\mathbf{N}_h = [N_x, N_y]^T$  is the component of  $\mathbf{N}$  in the horizontal  $x - y$  plane. The term  $|\mathbf{N}_h|$  accounts for the local surface orientation that need not be aligned with the horizontal direction. For vertical surfaces,  $|\mathbf{N}_h| = 1$  (e.g., the horizontal vectors on the last building in Fig. 1b), for angled surfaces  $0 < |\mathbf{N}_h| < 1$  (e.g., the vectors on the middle building in Fig. 1b), and for horizontal surfaces, for example, flat roofs,  $|\mathbf{N}_h| = 0$  (the vertical vector on the last building in Fig. 1b). Superficially, it seems the integral is undefined for horizontal surfaces, but it can be shown (Van Reeuwijk et al. 2021) that the integral becomes finite upon integrating over  $z$ . Physically, one can interpret this as a flat surface providing a finite contribution over an infinitesimal distance in  $z$ . Note that the presence of these apparent singularities is an unavoidable consequence of using a planar average rather than a volumetric average, and its evaluation is discussed in Sect. 3.2. To demonstrate the physical meaning of the line integral, it is useful to apply (6) to the mean pressure  $\bar{p}$ . The horizontal pressure gradient then becomes

$$\left\langle \frac{\partial \bar{p}}{\partial x} \right\rangle = \frac{\partial \langle \bar{p} \rangle}{\partial x} - \oint_{\partial \Omega_f} \mathcal{A} \bar{p} \frac{N_x}{|\mathbf{N}_h|} ds. \quad (7)$$

This demonstrates that the filtering operation produces the form drag in the  $x$ -direction within the filter support. Note that under the assumption of flow in the positive  $x$ -direction and cuboids that face the wind head-on, we would have  $N_x/|\mathbf{N}_h| = -1$  on windward vertical surfaces,  $+1$  on leeward vertical surfaces and  $0$  on sidewalls. The top surface of buildings will not contribute to the form drag since  $N_x = 0$ , although care must be taken to show that the apparent singularity caused by  $|\mathbf{N}_h| = 0$  does not cause trouble (see Sect. 3.2).

For the planar filter  $\mathcal{A}$ , the vector form of the spatial averaging theorem (Whitaker 1999, eqs. 1.2–23) is given by (see appendix A):

$$\left\langle \frac{\partial F_i}{\partial x_i} \right\rangle = \frac{\partial \langle F_i \rangle}{\partial x_i} - \oint_{\partial \Omega_f} \mathcal{A} \frac{F_i N_i}{|\mathbf{N}_h|} ds, \quad (8)$$

where repeated indices imply implicit summation. As an example of the application of this theorem, consider filtering the mean continuity equation  $\nabla \cdot \bar{\mathbf{u}} = 0$ . Using (7) with  $F_i = \bar{u}_i$  results in:

$$\left\langle \frac{\partial \bar{u}_i}{\partial x_i} \right\rangle = \frac{\partial \langle \bar{u}_i \rangle}{\partial x_i} - \oint_{\partial \Omega_f} \mathcal{A} \frac{\bar{u}_i N_i}{|\mathbf{N}_h|} ds = \frac{\partial \langle \bar{u}_i \rangle}{\partial x_i}, \quad (9)$$

since  $\bar{u}_i N_i = 0$  on  $\partial \Omega_f$  due to the impermeability of the buildings. Therefore,  $\nabla \cdot \langle \bar{\mathbf{u}} \rangle = 0$ , implying that the filtered velocity field is also divergence-free.

If the averaging length  $L$  is much larger than the characteristic length scale  $\ell$  of the urban surface, the average concentration  $\langle \varphi \rangle$  and horizontal fluxes  $\langle F_x \rangle$  and  $\langle F_y \rangle$  will be independent of  $x, y$ , implying that the spatial averaging theorems reduce to (Schmid et al. 2019):

$$\left\langle \frac{\partial \varphi}{\partial x_i} \right\rangle = \frac{\partial \langle \varphi \rangle}{\partial z} - \oint_{\partial \Omega_f} \mathcal{A} \varphi \frac{N_i}{|\mathbf{N}_h|} ds, \quad (10)$$

$$\left\langle \frac{\partial F_i}{\partial x_i} \right\rangle = \frac{\partial \langle F_z \rangle}{\partial z} - \oint_{\partial \Omega_f} \mathcal{A} \frac{F_i N_i}{|\mathbf{N}_h|} ds. \quad (11)$$

## 2.3 Triple Decomposition

We define the Reynolds average of  $\varphi$  as an ensemble average (Schmid et al. 2019):

$$\bar{\varphi}(\mathbf{x}, t) \equiv \sum_{i=1}^N \varphi^{(i)}(\mathbf{x}, t). \quad (12)$$

Deviations from the average value  $\bar{\varphi}$  are denoted  $\varphi'(\mathbf{x}, t) = \varphi(\mathbf{x}, t) - \bar{\varphi}(\mathbf{x}, t)$ . For statistically steady problems, time-averaging can be used in lieu of an ensemble average. In the current paper, we will assume that Reynolds-averaging has been performed before spatial

averaging, which avoids some of the complexities of simultaneous Reynolds and spatial averaging (Schmid et al. 2019). We thus decompose the average velocity into the local spatial average and its deviation from the average as  $\bar{\varphi} = \langle \bar{\varphi} \rangle^f + \bar{\varphi}''$ . This decomposition is only defined in the fluid domain, and the intrinsic average is used since this represents the average value inside the fluid domain. Substitution into Eq. (10) results in the triple decomposition:

$$\varphi(\mathbf{x}, t) = \underbrace{\langle \bar{\varphi} \rangle^f(\mathbf{x}, t) + \bar{\varphi}''(\mathbf{x}, t)}_{\bar{\varphi}(\mathbf{x}, t)} + \varphi'(\mathbf{x}, t). \quad (13)$$

For products, e.g. of a scalar flux  $u_i \varphi$ , this implies:

$$\begin{aligned} \langle \overline{u_i \varphi} \rangle &= \langle \bar{u}_i \bar{\varphi} \rangle + \langle u'_i \varphi' \rangle \\ &= \langle \langle \bar{u}_i \rangle^f \langle \bar{\varphi} \rangle^f \rangle + \langle \langle \bar{u}_i \rangle^f \bar{\varphi}'' \rangle + \langle \bar{u}_i'' \langle \bar{\varphi} \rangle^f \rangle + \langle \bar{u}_i'' \bar{\varphi}'' \rangle + \langle \overline{u'_i \varphi'} \rangle. \end{aligned} \quad (14)$$

## 2.4 Reynolds- and Area-Averaged Streamwise Momentum Equations

Ignoring buoyancy effects and assuming incompressible airflow, the (unsteady) Reynolds-averaged streamwise horizontal momentum equation is given by:

$$\frac{\partial \bar{u}}{\partial t} + \frac{\partial \bar{u}_j \bar{u}}{\partial x_j} + \frac{\partial \overline{u'_j u'}}{\partial x_j} + \frac{\partial \bar{p}}{\partial x} = f + \nu \frac{\partial^2 \bar{u}}{\partial x_j^2}, \quad (15)$$

where  $\bar{u}$  represents the mean velocity in the streamwise  $x$ -direction,  $\bar{p}$  is the kinematic deviatoric pressure, and  $f$  is the volumetric forcing term, typically due to a pressure gradient. This equation can be written in flux form as:

$$\frac{\partial \bar{u}}{\partial t} + \frac{\partial F_i}{\partial x_i} = f, \quad \text{where} \quad F_i = \bar{u}_i \bar{u} + \overline{u'_i u'} - \nu \frac{\partial \bar{u}}{\partial x_i} + \bar{p} \delta_{i1} \quad (16)$$

where  $\delta_{ij}$  is the Kronecker delta. Superficial averaging and application of the spatial averaging theorem Eq. (7) results in:

$$\frac{\partial \langle \bar{u} \rangle}{\partial t} + \frac{\partial \langle F_i \rangle}{\partial x_i} = \langle f \rangle - \langle f_D \rangle, \quad (17)$$

where  $\langle f \rangle = \varepsilon f$  and the drag force induced by the buildings  $\langle f_D \rangle$  is given by:

$$\langle f_D \rangle = - \oint_{\partial \Omega_f} \mathcal{A} \frac{F_i N_i}{|\mathbf{N}_h|} ds = - \oint_{\partial \Omega_f} \mathcal{A} \left( \bar{p} \frac{N_x}{|\mathbf{N}_h|} - \nu \frac{\partial \bar{u}}{\partial x_j} \frac{N_j}{|\mathbf{N}_h|} \right) ds. \quad (18)$$

This equation explicitly provides the drag force induced by the building surfaces that is comprised of a form drag associated with the pressure and a skin drag associated with the viscosity (Raupach and Shaw 1982; Finnigan 2000; Nepf 2012; Sützl et al. 2020).

Considering the fact that NWP models take into account the net effect of buildings without actually resolving them, the governing equations are posed in terms of *superficial* quantities. Recasting Eq. (15) in the standard form for the streamwise momentum equation, we thus obtain:

$$\frac{\partial \langle \bar{u} \rangle}{\partial t} + \frac{\partial \langle \bar{u}_i \rangle \langle \bar{u} \rangle}{\partial x_i} + \frac{\partial \langle \bar{p} \rangle}{\partial x} = \langle f \rangle - \langle f_D \rangle + \frac{\partial \langle \tau_i \rangle}{\partial x_i}, \quad (19)$$

where:

$$\langle \tau_i \rangle = \langle \bar{u}_i \rangle \langle \bar{u} \rangle - \langle \bar{u}_i \bar{u} \rangle - \langle u'_i u' \rangle, \quad (20)$$

is the unresolved kinematic stress term. Note that we have not decomposed the term  $\langle \bar{u}_i \bar{u} \rangle$  into its components and that the viscous flux has been omitted since it is negligible in magnitude compared to the other fluxes. Equation (17) highlights the two terms that need to be closed in order to represent the effect of buildings in a non-building resolving NWP model: (1) the distributed drag contribution  $f_D$ ; and (2) the unresolved kinematic stress  $\tau_i$ .

## 2.5 Statistically Steady Flow with Periodic Boundary Conditions

The case considered in this paper is a turbulent urban boundary layer flow with periodic boundary conditions and a free-slip top boundary at  $z = h$  in a statistically steady state. In this situation, Eq. (17) becomes:

$$\frac{\partial \langle \bar{u}_i \rangle \langle \bar{u} \rangle}{\partial x_i} + \frac{\partial \langle \bar{p} \rangle}{\partial x} = \langle f \rangle - \langle f_D \rangle + \frac{\partial \langle \tau_i \rangle}{\partial x_i}, \quad (21)$$

which is not much of a simplification. No term except the time derivative can be assumed zero for an arbitrary averaging length  $L$ . However, if the averaging length scale is very large, the problem will become homogeneous in  $x$  and  $y$  and all the horizontal derivatives will vanish. With this in mind, it is useful to rewrite the equation above as:

$$\frac{\partial \langle \bar{u} \rangle \langle \bar{u} \rangle}{\partial x} + \frac{\partial \langle \bar{v} \rangle \langle \bar{u} \rangle}{\partial y} + \frac{\partial \langle \bar{p} \rangle}{\partial x} - \frac{\partial \langle \tau_x \rangle}{\partial x} - \frac{\partial \langle \tau_y \rangle}{\partial y} = \langle f \rangle - \langle f_D \rangle - \frac{\partial \langle \bar{w} \rangle \langle \bar{u} \rangle}{\partial z} + \frac{\partial \langle \tau_z \rangle}{\partial z}, \quad (22)$$

Here, the horizontal terms on the left-hand side will vanish if the problem is homogeneous, leaving the terms on the right-hand side of the equation to balance. The equation above indicates that an imbalance in the terms on the right-hand side implies inhomogeneity in  $x$  and  $y$  via horizontal terms on the left-hand side and vice versa.

Homogeneity is expected when  $L \gg \ell$ , i.e.  $L$  is much larger than the characteristic urban lengthscale  $\ell$ . In this case, the horizontal momentum equation simplifies to:

$$\frac{d \langle \tau_z \rangle_\infty}{dz} = \langle f_D \rangle_\infty - \langle f \rangle_\infty, \quad (23)$$

where:

$$\langle \tau_z \rangle_\infty = -\langle \overline{w} \overline{u} \rangle_\infty - \langle \overline{w'} u' \rangle_\infty = -\langle \overline{w''} \overline{u''} \rangle_\infty - \langle \overline{w' u'} \rangle_\infty, \quad (24)$$

note that  $\langle \overline{w} \rangle_\infty = 0$  due to volume conservation. The subscript on the superficial average indicates the averaging length  $L$ , in this case  $L/\ell \gg 1$  which we denote as  $L = \infty$ . In the limit of  $L \gg \ell$ , the quantity  $\langle \varphi \rangle_\infty$  is identical to the conventional plane-average  $\langle \varphi \rangle_{xy}$ , which is typically used in computational studies (Coceal et al. 2006; Xie et al. 2008; Giometto et al. 2016; Nazarian et al. 2020; Sützl et al. 2020). The superficial plane average of  $\varphi$  is defined as:

$$\langle \varphi \rangle_{xy}(z, t) \equiv \frac{1}{A_T} \int_{\Omega_f(z)} \varphi(\mathbf{x}, t) d\mathbf{x}_h, \quad (25)$$

where  $A_T = L_x L_y$  is the total area of the horizontal plane. Note that the plane average is distributive and that contrary to  $\langle \varphi \rangle$ ,  $\langle \langle \varphi \rangle_{xy} \rangle_{xy} = \langle \varphi \rangle_{xy}$ , i.e., this operator behaves like a regular Reynolds average.

Integrating Eq. (21) from height  $z$  to the domain top at height  $h$  results in the following cumulative stress relation (Sützl et al. 2020):

$$\langle \tau_z \rangle_\infty(z) = \tau_{f;\infty}(z) - \tau_{D;\infty}(z), \quad (26)$$

where

$$\begin{aligned} \tau_{f;L}(\mathbf{x}_h, z) &\equiv \int_z^h \langle f \rangle_L(\mathbf{x}_h, z') dz', \\ \tau_{D;L}(\mathbf{x}_h, z) &\equiv \int_z^h \langle f_D \rangle_L(\mathbf{x}_h, z') dz', \end{aligned} \quad (27)$$

are the cumulative shear stresses due to forcing and drag, respectively. They have been formulated here for an arbitrary length scale  $L$ . For  $L \gg \ell$ , the stresses are a function of  $z$  only. At the ground surface  $z = 0$ , the relation reduces to  $\tau_{f;\infty}(0) = \tau_{D;\infty}(0) \equiv \tau_0$ , where  $\tau_0$  is the kinematic surface shear stress (Pope 2000).

In order to keep consistency in notation with earlier work, we will use that in the absence of a specific filter size in the subscript of a quantity, it is implied that  $L = \infty$ , e.g.  $\tau_f \equiv \tau_{f;\infty}$ ; and we will use the conventional plane average notation  $\langle \varphi \rangle_{xy}$  instead of its equivalent  $\langle \varphi \rangle_\infty$ . With this in mind, the cumulative stress relation Eq. (24) is simply:

$$\langle \tau_z \rangle_{xy} = \tau_f(z) - \tau_D(z). \quad (28)$$

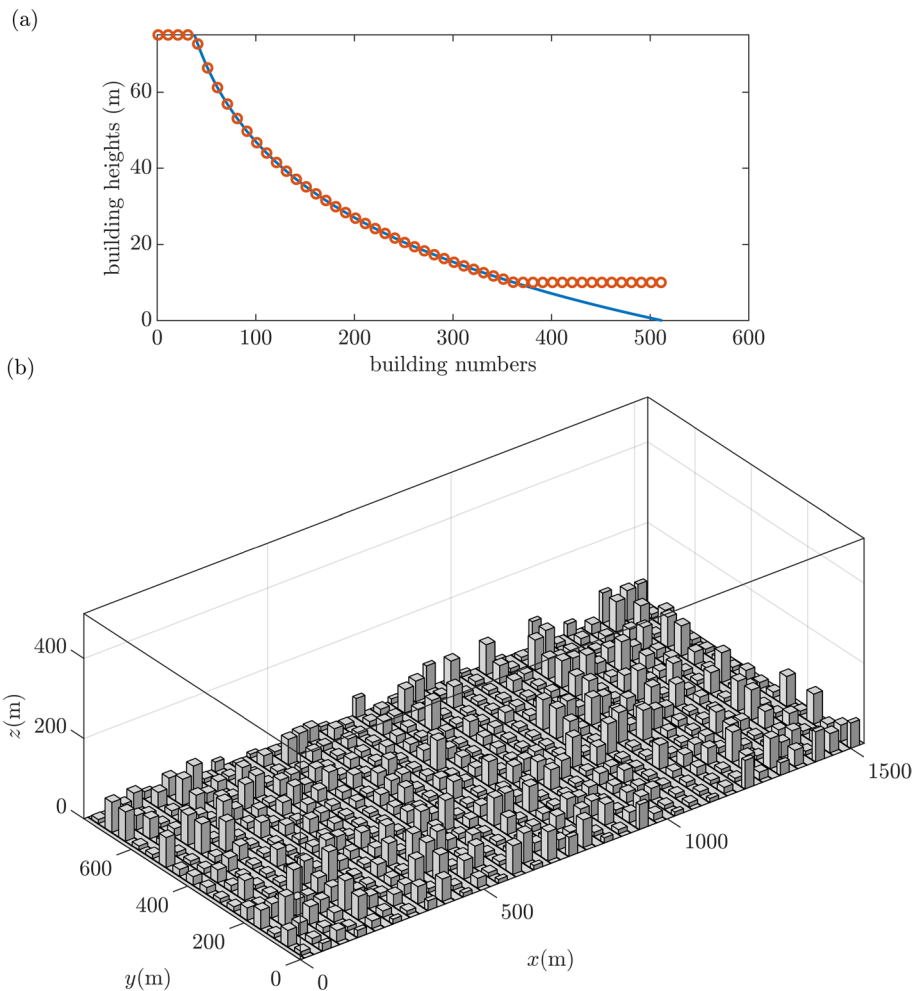
### 3 Simulation Details

#### 3.1 Case Setup

In order to investigate multi-scale aspects of urban flows, we consider an idealised quasi-realistic urban morphology that has a staggered arrangement of cuboidal buildings of breadth and

width  $W = 24$  m, whose height  $H$  is random (Fig. 2). This case is inspired by Xie et al. (2008), but with a much larger simulation domain of  $L_x \times L_y \times L_z = 1536 \text{ m} \times 768 \text{ m} \times 512 \text{ m}$ , and using a height distribution representative for London (Sütl et al. 2021) rather than a Gaussian height distribution.

With a spacing of size  $W$  between the buildings, the domain will contain  $N_b = 32 \times 16 = 512$  buildings. This implies that the plan-area index  $\lambda_p \equiv A_p/A_T = 0.25$  where the plan area and total area are defined as, respectively,  $A_p = N_b W^2$  and  $A_T = L_x L_y$ . For the distribution of building heights  $H$ , we use an empirical relation which was obtained using metre-scale GIS data for the Greater London area (Sütl et al. 2021). It was found that the total building width  $L_b$  (i.e. the sum of the frontal widths of all buildings) at elevation  $z$  can be approximated as:



**Fig. 2** Construction of the urban geometry. **a** Determination of building heights from  $L_b(z)/W$  (solid line) by sampling  $L_b/W$  at integer values (circles; every ten buildings shown); **b** The synthetic urban geometry

$$L_b(z; \alpha) = \frac{A_F}{h_{\max}} \frac{\alpha e^{-\alpha z/h_{\max}}}{1 - e^{-\alpha}}, \quad z \leq h_{\max}, \quad (29)$$

where the coefficient  $\alpha(r)$  was observed to depend on the maximum-to-mean-building height  $r = h_{\max}/h_m$  as:

$$\alpha(r) = 1.355r - 0.7807. \quad (30)$$

Here,  $h_{\max}$  is the maximum building height,  $h_m$  is the mean building height, and  $A_F$  is the total frontal area of the buildings. Note that  $A_F/h_{\max}$  is the total building width  $L_b$  if all buildings would have height  $h_{\max}$ . Note that above the maximum building height  $L_b = 0$ .

The mean building height is set to  $h_m = 30$  m and the ratio is set to  $r = 2.5$ , both of which are close to the average value of London reported by Sützl et al. (2021). Thus the maximum building height is  $h_{\max} = 75$  m and  $\alpha = 2.61$ . At  $z = 0$ , the total building width  $L_b(z; \alpha) = N_b W$ , which implies that the frontal area can be determined from Eq. (26) as:

$$A_F = \frac{1 - e^{-\alpha}}{\alpha} N_b W h_{\max}, \quad (31)$$

resulting in  $A_F = 327\,138 \text{ m}^2$  and thus a frontal area index  $\lambda_f \equiv A_F/A_T = 0.28$ .

In order to assign heights to each cuboidal building, note that Eq. (26) divided by  $W$  can be inverted to show the height  $z$  as a function of the number of buildings  $L_b(z)/W$ , which represents the distribution of heights as a function of the number of buildings (Fig. 2a, solid line). The building heights can be determined by uniformly sampling this function for each  $i = 1, 2, \dots, N_b$ . The first  $X$  buildings, where the function is not defined, are set to  $h_{\max}$ , and the minimum building height is set to 10 m. As a final step, the heights are randomly distributed using a uniform distribution to all 512 buildings as shown in the morphology in Fig. 2b. The total building width  $L_b$  is shown in Fig. 6a against the height, above  $z = 10$  m which is the minimum building height, the building width  $L_b$ , also indicating that the distributive frontal area smoothly decreases to zero at the top of the canopy region. Due to the minimum building height setting, the actual frontal area index is  $\lambda_f = 0.29$ , slightly greater than previously designed.

The urban flows over this morphology are performed by the large-eddy simulation using our open-source code uDALES (Owens et al. 2024), presenting the solid boundary using the immersed boundary method (IBM). The dynamics near the boundary are parametrised by the logarithmic wall functions (Uno et al. 1995; Suter et al. 2022). The turbulent eddy viscosity is calculated following the Vreman (2004) subgrid model. The code employs a second-order central difference scheme on a staggered Arakawa C-grid for spatial discretisation and an explicit third-order Runge–Kutta scheme for time integration.

The simulation is neutral and a constant pressure gradient forcing  $dP/dx = 0.0042 \text{ kg m}^{-2} \text{ s}^{-2}$  is imposed on the streamwise ( $x$ ) direction to drive the wind. The simulation has a grid number of  $N_x \times N_y \times N_z = 1024 \times 512 \times 512$  (i.e., the grid size is  $\Delta x \times \Delta y \times \Delta z = 1.5 \text{ m} \times 1.5 \text{ m} \times 1 \text{ m}$ ), with periodic boundary conditions applied in the lateral sides while the domain top is free-slip. The simulation runs for 10,000 s and the last 8000 s is used to obtain converged time-averaged statistics.

### 3.2 Implementation Details

Efficient computation of the convolution integral Eq. (2) is crucial to carry out a multi-scale analysis. In the spatial domain, the convolution integral needs to be evaluated  $N_x \times N_y \times N_z$  times for a single averaging length  $L$  which is computationally very expensive, particularly for high-resolution simulations: assuming a filter contains  $n^2$  grid cells, the convolution over the entire domain requires the number of operations of  $O(n^2 N_x^2 N_z)$  where we have assumed  $N_x = N_y$ . This is acceptable for small filter scales but becomes computationally inhibiting for large filter sizes. Therefore, we perform the convolution in the Fourier space by making use of the convolution theorem, which states that a convolution becomes a multiplication in Fourier space (Bracewell 2000), thus reducing the computational complexity of a 1-D convolution to  $O(N_x \log N_x)$ . We thus compute the horizontal 2-D FFT of  $I_f \varphi$  and the filter  $\mathcal{A}$ , multiply them, and then transform the product back to physical space using an inverse FFT, obtaining the filtered results for all points in the plane simultaneously. Repeating this for all  $N_z$  planes requires  $O((\log N_x)^2 N_x^2 N_z)$  operations, thus greatly reducing the computational cost, particularly for large simulation domains and large filters. The algorithm can be downloaded from van Reeuwijk and Huang (2025). The filtering is performed for nine different averaging lengths:  $L = 3, 6, 12, 24, 48, 96, 192, 384$ , and  $768$  m.

Calculating the drag force Eq. (16) also requires care as the line integral to evaluate since the building surfaces contain many horizontal surfaces for which  $|\mathbf{N}_h| = 0$ . In order to ensure an appropriate discretisation of this term, note that by definition, for any boundary quantity  $\phi$ ,

$$\int \oint_{\partial\Omega_f(z)} \frac{\phi}{|\mathbf{N}_h|} ds dz = \iint_{\partial\Omega_f} \phi dS, \quad (32)$$

i.e. integrating the line integral over  $z$  results in the total flux exchanged across the solid–fluid surface. In a discrete form, this equation is given by:

$$\sum_k \left( \sum_{m:K_m=k} \frac{\phi_m}{|\mathbf{N}_h|_m} \Delta s_m \right) \Delta z_k = \sum_m \phi_m A_m, \quad (33)$$

where each cell-facet  $m$  has area  $A_m$ , surface normal  $\mathbf{N}_m$ , and its contribution  $\phi_m$  is linked to a grid cell at index  $I_m$ ,  $J_m$  and  $K_m$ . Since both sides of this equation use all cell facets exactly once, it follows that:

$$\frac{\Delta s_m}{|\mathbf{N}_h|_m} = \frac{A_m}{\Delta z_{K_m}}, \quad (34)$$

which clarifies how  $\Delta s_m/|\mathbf{N}_h|_m$  needs to be evaluated; it also shows that this term is always nonzero and finite. In order to be able to straightforwardly incorporate filtering, the surface quantity  $\phi$  is converted to a surface density field  $\rho_\phi$  that contains the boundary term and which is only non-zero in the cells next to the boundary:

$$\rho_{\phi;ijk} = \sum_{m \in M_{ijk}} \frac{\phi_m A_m}{\Delta x \Delta y \Delta z K_m}, \quad (35)$$

where  $M_{ijk} \in \{m : I_m = i, J_m = j, K_m = k\}$  is the set of all cell-facets associated with cell  $i, j, k$ . Using Eq. (32), any surface quantity can be filtered straightforwardly using  $\langle \rho_\phi \rangle(\mathbf{x}) = \int_{\Omega_f} \mathcal{A}(\mathbf{x}_h - \mathbf{y}_h) \rho_\phi(\mathbf{y}_h) d\mathbf{y}_h$ . Here we introduce a volumetric field for two surface quantities:

- the distributed drag term, by setting  $\phi_m = -(\bar{p}_m \mathbf{e}_x - \nu(\nabla \bar{u})_m) \cdot \mathbf{N}_m$ . Denoting the volumetric drag density by  $\rho_D$ , we have that  $\langle f_D \rangle = \langle \rho_D \rangle$ . Here it should be noted that  $\nu(\nabla \bar{u})_m$  follows from the wall shear stress as calculated by the wall functions and not the resolved velocity field (Owens et al. 2024).
- the frontal area term, by setting  $\phi_m = -\min(\mathbf{e}_u \cdot \mathbf{N}_m, 0)$ , where  $\mathbf{e}_u$  is a unit vector for the wind direction (where  $\mathbf{e}_{u;z} = 0$ ). This term can be understood by realising that the projected area of facet area  $m$  is given as  $A_m |\mathbf{e}_u \cdot \mathbf{N}_m|$  and that a contribution is only counted when it involves the windward side, i.e.  $\mathbf{e}_u \cdot \mathbf{N}_m < 0$ . Denoting the resulting frontal-area density  $\rho_L$ , we have that  $\int \rho_L dxdy = L_b$  and  $\int \rho_L dV = A_F$ .

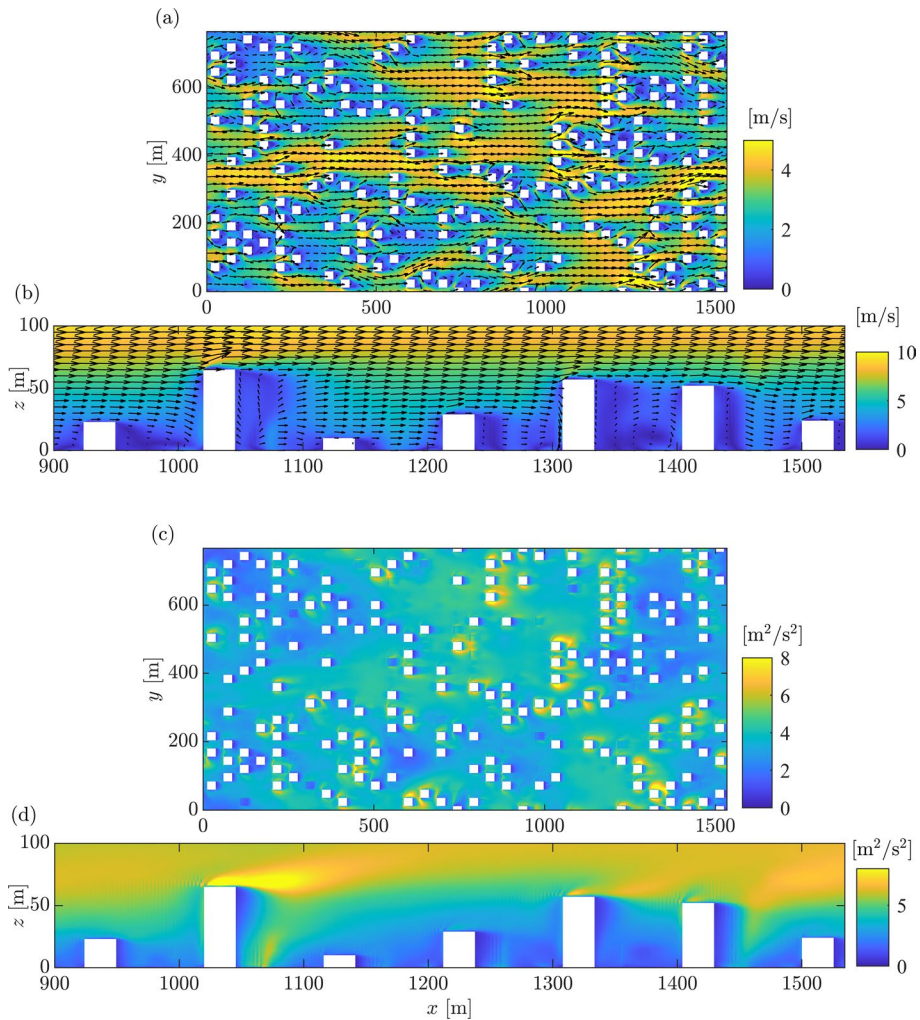
## 4 Results

### 4.1 Classical Plane-Averaged Analysis

The time-averaged wind speed at the mean building height  $z = 30$  m (Fig. 3a) shows large variations in wind speed and channelling in regions where buildings are absent. A strong reduction in wind speed is shown near some of the tall building clusters (Mishra et al. 2023, 2024), with wakes clearly visible downstream of the tall buildings. The elevation view (Fig. 3b) shows the general increase of wind speed with height and also demonstrates the homogenisation of the wind in the  $x$  direction with increasing height, consistent with the extent of the roughness sublayer (the height of this layer is  $h_{\max}$  as will be shown in Fig. 4). The velocity vectors show a clearly visible wake behind the building (in particular behind the first building in Fig. 3b). Accelerations are visible near the upwind top of the building (e.g., for the second cube in Fig. 3b), which is consistent with Coceal et al. (2007). In Fig. 3b, the skimming flow can be observed in places, e.g., near the last cube; and also the flow recirculation, e.g., between the last third and last second cube.

Figure 3c, d show the plane view at  $z = 30$  m and the vertical view at the central plane of time-averaged turbulent kinetic energy (TKE,  $k(\mathbf{x}) \equiv \frac{1}{2} (\overline{u'u'} + \overline{v'v'} + \overline{w'w'})$ ), respectively. Note that the TKE field is correlated with the mean wind speed field: in the building cluster area, e.g. at the top right of Fig. 3c, it can be seen to have low levels of turbulence due to the low velocity in the building wake; while at the upwind top of the building (e.g., the second building in Fig. 3d), large TKE can be observed due to the acceleration and large shear at the top.

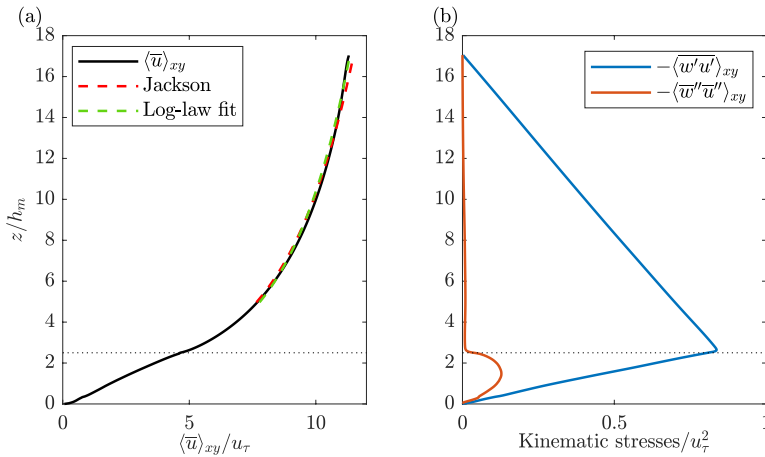
Figure 4 presents the streamwise velocity and turbulent transport, as the pressure gradient forcing is applied solely in the streamwise direction, which governs the dominant flow and transport. Figure 4a shows the profile of plane-averaged superficial streamwise veloc-



**Fig. 3** Time-averaged flow statistics. **a** Wind speed  $\sqrt{\bar{u}^2 + \bar{v}^2}$  at mean building height  $z = 30$  m overlaid with  $(\bar{u}, \bar{v})$  velocity vectors; **b** wind speed  $\sqrt{\bar{u}^2 + \bar{w}^2}$  at the central vertical plane  $y = 384$  m overlaid with  $(\bar{u}, \bar{w})$  velocity vectors; **c** Turbulent kinetic energy  $k$  at mean building height  $z = 30$  m. **d** Turbulent kinetic energy  $k$  at the central vertical plane  $y = 384$  m. The white boxes represent the buildings

ity  $\langle \bar{u} \rangle_{xy}$  and normalised by the friction velocity  $u_\tau = \sqrt{\tau_0}$ . Above the canopy region i.e.,  $z \geq h_{\max}$ , the profile fits the logarithmic law:

$$\langle \bar{u} \rangle_{xy}(z) = \frac{u_\tau}{\kappa} \ln \left( \frac{z - z_d}{z_0} \right), \quad (36)$$



**Fig. 4** Plane-averaged flow statistics. **a** Streamwise velocity  $\langle \bar{u} \rangle_{xy}$  against normalised height  $z/h_m$ , overlaid with logarithmic profiles Eq. (33) estimated using the log-law fitting and the Jackson displacement method; **b** normalised kinematic turbulent shear stress  $-\langle w'u' \rangle_{xy}$  and dispersive stress  $-\langle \bar{w}''\bar{u}'' \rangle_{xy}$ . The dotted line marks the maximum building height  $h_{\max}$

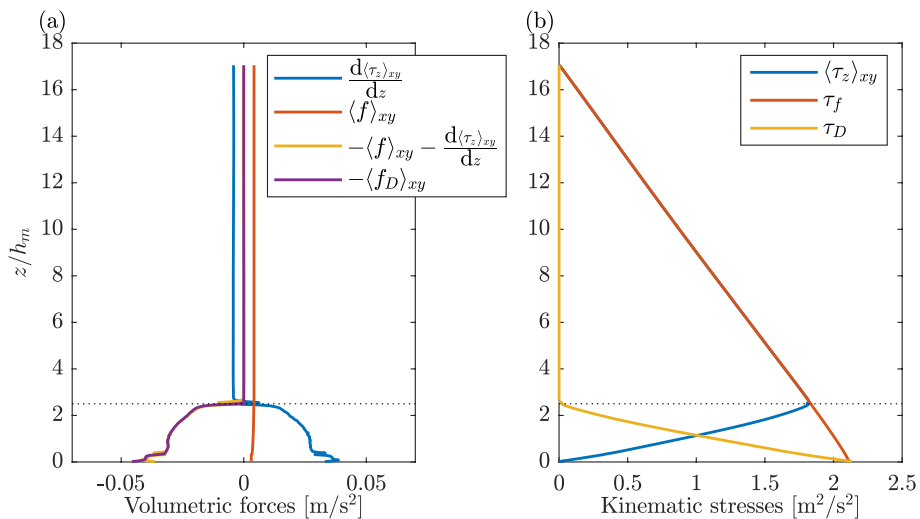
where  $\kappa$  is the Von Karman constant,  $z_d$  is the displacement length and  $z_0$  is the roughness length. Two different methods will be used to estimate the parameters  $z_d$  and  $z_0$ . The first is to find the  $z_d$  and  $z_0$  directly from the profile in the inertial sublayer region (for example, here, we choose the fitted region above  $2h_{\max}$  to the top end) that fits the logarithmic law best. This method takes  $\kappa = 0.41$  to be given and results in  $z_d = 40.0 \text{ m} \approx 1.33h_m$  and  $z_0 = 4.5 \text{ m} \approx 0.15h_m$ . The second method follows a hypothesis by Jackson (1981), that assumes the displacement height  $z_d$  corresponds to the centre of pressure of the forces acting on the buildings, i.e.:

$$z_d = \frac{\int_0^{h_{\max}} z \langle f_D \rangle_{xy} dz}{\int_0^{h_{\max}} \langle f_D \rangle_{xy} dz} = \frac{1}{\tau_0} \int_0^{h_{\max}} \tau_D dz, \quad (37)$$

where  $\langle f_D \rangle_{xy}$  is the distributed drag force and note that  $\tau_D = \tau_{D;\infty}$  (see Sect. 2.5). The displacement length calculated is  $z_d = 35.1 \text{ m} \approx 1.17h_m$ . In this approach,  $\kappa$  is determined together with the roughness length  $z_0$  to best fit the logarithmic law, with results  $\kappa = 0.37$ , and  $z_0 = 6.8 \text{ m} \approx 0.23h_m$ . The velocity profiles fitted using both methods are plotted in Fig. 4a as the dashed lines. A prediction of  $z_d$  and  $z_0$  based on the plane index  $\lambda_p$  (Macdonald et al. 1998; Kanda et al. 2013) results values  $z_d \approx 1.20h_m$ ,  $z_0 \approx 0.15h_m$  for our case. All our results are generally close to this prediction with  $z_0$  from the second method slightly higher. As the displacement height  $z_d$  and roughness length  $z_0$  is typically related to the morphological parameters like plane area index  $\lambda_p$ , frontal area index  $\lambda_f$ , mean building height  $h_m$  and the variance of the building height (Grimmond and Oke 1999) which are all predetermined for this case according to the London morphology profile, so that in general, we do not expect a big variation in  $z_d$  and  $z_0$  among the random building-height distributions, although slight discrepancy would exist for example a distribution with a denser high-rise area may reduce the roughness length due to the sheltering effect (Oke et al. 2017).

Figure 4b shows the vertical evolution of the plane-averaged kinematic turbulent shear stress  $-\langle w'u' \rangle_{xy}$  and dispersive stress  $-\langle \bar{w}''\bar{u}'' \rangle_{xy}$ . It shows that the turbulent shear stress  $-\langle w'u' \rangle_{xy}$  increases with height to a peak near the maximum building height, and then decreases linearly with height as expected from classical boundary layer theory. The dispersive flux  $-\langle \bar{w}''\bar{u}'' \rangle_{xy}$  is significant within the canopy with a peak near the mean building height. Above the canopy, the dispersive stress reduces to a very low value, which indicates that the upper limit of the roughness sublayer, or equivalently, the lower limit of the inertial sublayer, is at about  $z = h_{\max}$ . Note that this is a simple approximation; use of other quantities or definitions for the roughness sublayer, such as that of Placidi and Ganapathisubramani (2015), might work better for varying building heights but are beyond the scope of this study. The fact that the upper limit of the roughness sublayer is identical to the maximum building heights is quite unusual, since it implies that there is only an urban canopy layer but no roughness sublayer. Indeed, for uniform or random-height cube arrays (Coccal et al. 2006; Xie et al. 2008; Sützl et al. 2020; Lu et al. 2024), the roughness sublayer extends above the building heights. This is likely the result of having a very large domain with a specific non-repeating height distribution of buildings. There are 39 buildings out of the total 512 that have height  $h_{\max}$ , which is apparently insufficient to induce significant perturbations to the mean velocity field.

Figure 5a shows the plane-averaged force balance. The drag force  $\langle f_D \rangle_{xy}$  is determined in two ways: from the residual of the other two terms and from its direct definition Eq. (16). The good agreement between both shows that the momentum budget closes and is thus confirmation that the system is in a statistically steady state. Within the canopy region, the magnitude of the drag force decreases with height as expected because less space is occupied by building areas at larger heights. Above the canopy, the drag force vanishes and the pressure gradient matches the turbulent shear stress divergence. Figure 5b illustrates the plane-averaged kinematic stress balance—the integral version of the force balance Eq.



**Fig. 5** Plane-averaged momentum and cumulative stresses as a function of normalised height  $z/h_m$ . **a** Momentum balance Eq. (21). **b** Cumulative kinematic stress balance Eq. (25). Dotted lines indicate the maximum building height

(25). Above the canopy region, the stresses  $\langle \tau_z \rangle_{xy}$  and  $\tau_f$  coincide in the absence of the drag force and both are linear with the height due to the constant gradient shown in Fig. 5a. Within the canopy region, the kinematic shear stress  $\langle \tau_z \rangle_{xy}$  increases roughly linearly with height, and the cumulative drag stress  $\tau_D$  reduces from its maximal value at  $z = 0$  to zero at  $h_{\max}$ .

We consider two methods in which the drag force can be parameterised. The first is to introduce a height-dependent drag coefficient  $C_d(z)$  related to the forcing as (Santiago and Martilli 2010):

$$\langle f_D \rangle_{xy} = \frac{1}{2} \frac{L_b}{A_T} C_d \langle \bar{u} \rangle_{xy}^2, \quad (38)$$

where we recall that  $L_b(z)$  is the frontal width occupied by buildings, such that  $\int_0^{h_{\max}} L_b(z) dz = A_F$ . Note the absence of the fluid density in the expression above since  $\langle f_D \rangle_{xy}$  is a kinematic force. The second method is based on the observation that the cumulative drag  $\tau_D(z)$  can be parameterised as (Sützl et al. 2020):

$$\frac{\tau_D}{\tau_0} = 1.88\zeta^3 - 3.89\zeta^2 + 3.01\zeta, \quad (39)$$

where  $\zeta$  is the cumulative normalised frontal area, defined as:

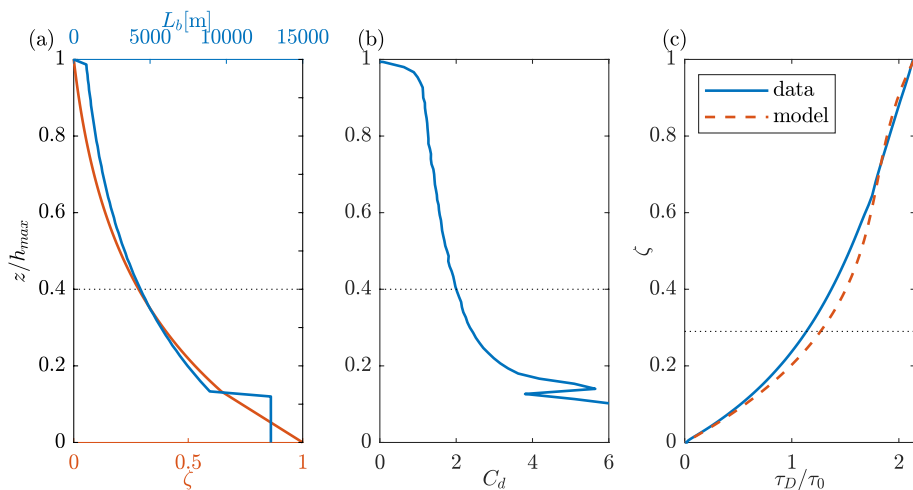
$$\zeta(z) \equiv A_F^{-1} \int_z^{h_{\max}} L_b(z') dz'. \quad (40)$$

Figure 6a shows the vertical evolution of  $L_b$  and  $\zeta$  for the current urban geometry, and shows that  $\zeta$  smoothly decreases from  $\zeta = 1$  at ground level to  $\zeta = 0$  at the maximum building height  $h_{\max}$ .

Figure 6b shows the drag coefficient  $C_d$  calculated from Eq. (35). The coefficient is the largest near the ground due to very low velocities. The coefficient decreases from 4.0 to 0.5 over most of the building heights, i.e.,  $z/h_{\max} \geq 0.2$ . The coefficient profile is generally consistent with the values reported in Coceal and Belcher (2004). The kink in the  $C_d$  near  $0.1 < z/h_{\max} < 0.15$  is due to the rapid decreases in building width  $L_b$ , because due to Eq. (35) and drag and velocity profile, the production of  $C_d$  and  $L_b$  should decrease smoothly with the height. Figure 6c shows the relation between  $\tau_D$  and  $\zeta$  together with Eq. (36). There is reasonably good agreement of the data with the parameterisation, although the current simulation does not show an inflection point near ground level ( $\zeta = 1$ ). This might suggest that Eq. (36), which was obtained from a best fit based on eight independent simulations, may need some further calibration, but this would need many simulations (e.g. using the Lu et al. (2024) dataset) to be done reliably. We will therefore keep using Eq. (36) even though we could construct a better fitting relation.

## 4.2 Multi-resolution Analysis

In this section, we will use the multi-resolution framework developed in Sect. 2. To illustrate the coarse-graining, the mean streamwise velocity  $\bar{u}$  at the mean building height  $z = 30$  m is shown in Fig. 7 at a number of different averaging lengths  $L$ . Figure 7a shows the velocity



**Fig. 6** Distributed drag parameterisations as a function of scaled height  $z/h_{max}$ . **a** Total frontal building width  $L_b$  and scaled total frontal building area  $\zeta$ ; **b** drag coefficient  $C_d$  calculated from Eq. (35); **c** plane-averaged normalised cumulative drag stress  $\tau_D/\tau_0$  against  $\zeta$ , overlaid with parameterisation Eq. (36). The dotted line indicates the location of the mean building height

field at the original resolution. Recall that the resolution is 1.5 m in both directions and that the white boxes represent the buildings where there is no fluid. Figure 7b–e show the superficial velocity  $\langle \bar{u} \rangle_L$  at averaging lengths  $L$  of 6, 24, 96 and 384 m. Figure 7b, c show that the superficial averaging gradually fills in the areas in which no information about the flow is present. Figure 7c shows that the values of  $\langle \bar{u} \rangle_{24}$  are unrealistically low at the original building locations, as can be expected theoretically (Whitaker 1999; Schmid et al. 2019). At  $L = 96$  m however (Fig. 7d), this no longer appears to be a significant effect. At a resolution  $L = 384$  m (Fig. 7e), the velocity  $\langle \bar{u} \rangle_{384}$  is practically homogeneous. Figure 7(f) shows the plane-averaged value  $\langle \bar{u} \rangle_{xy}$  of the plane, which is homogeneous by definition.

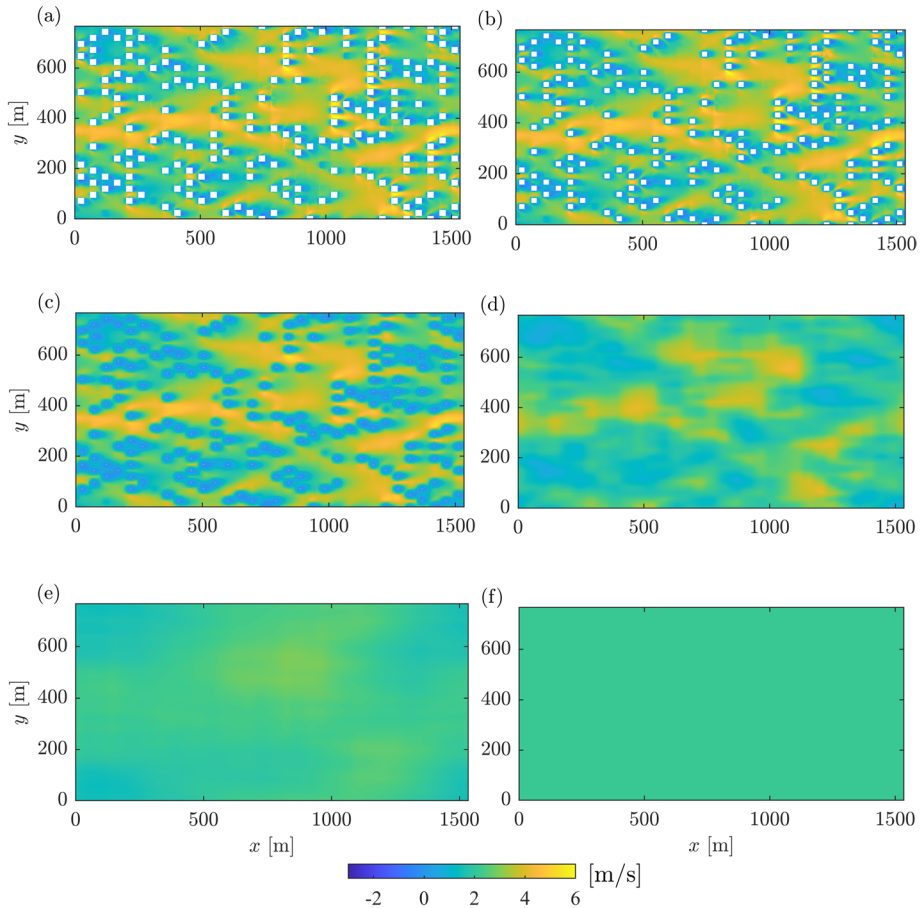
The variance of any quantity  $\bar{\varphi}$  can be used to quantify its spatial heterogeneity (see also Yu et al 2023). The plane-averaged variance of  $\bar{\varphi}$  quantifies the deviation of the original field from its plane average, which is defined as:

$$\sigma^2(z) \equiv \langle (\bar{\varphi}(\mathbf{x}_h, z) - \langle \bar{\varphi} \rangle_{xy})^2 \rangle_{xy}, \quad (41)$$

with the understanding that  $\langle \cdot \rangle_{xy} = \langle \cdot \rangle_\infty$ : this quantity represents the variance relative to the situation when the problem is fully homogenised in the  $x, y$  directions (i.e.  $L \gg \ell$ ). The total variance  $\sigma^2$  can be decomposed as:

$$\sigma^2 = \sigma_R^2 + \sigma_U^2 + \sigma_I^2, \quad (42)$$

where the unresolved variance  $\sigma_U^2$  quantifies the deviation of the original field from the spatial average, the resolved variance  $\sigma_R^2$  quantifies the deviation of the spatial average from the plane average, and (twice) the interaction covariance  $\sigma_I^2$  are defined as, respectively,

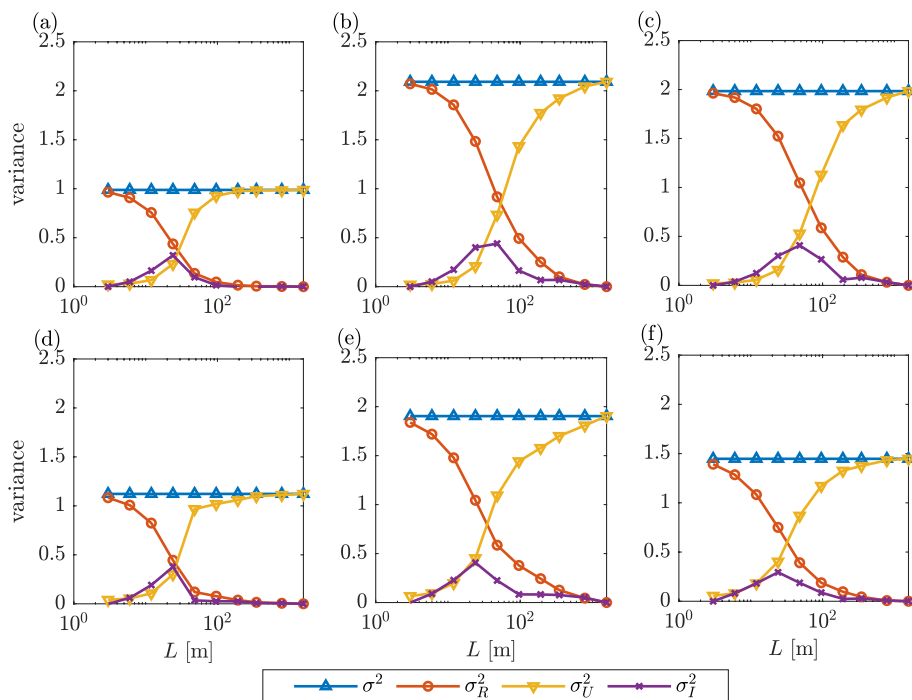


**Fig. 7** The streamwise velocity field  $\bar{u}$  at various averaging lengths at the mean building level  $z = 30$  m. **a** The original field  $\bar{u}$ , **b**  $L = 6$  m, **c**  $L = 24$  m, **d**  $L = 96$  m, **e**  $L = 384$  m, and **f** the plane-averaged field. The white boxes represent the buildings

$$\begin{aligned}
 \sigma_U^2 &\equiv \langle (\bar{\varphi} - \langle \bar{\varphi} \rangle)^2 \rangle_{xy}, \\
 \sigma_R^2 &\equiv \langle (\langle \bar{\varphi} \rangle - \langle \bar{\varphi} \rangle_{xy})^2 \rangle_{xy}, \\
 \sigma_I^2 &\equiv 2 \langle (\bar{\varphi} - \langle \bar{\varphi} \rangle) (\langle \bar{\varphi} \rangle - \langle \bar{\varphi} \rangle_{xy}) \rangle_{xy}.
 \end{aligned} \tag{43}$$

These variances are dependent on the height  $z$  and the averaging length  $L$ .

Figure 8 shows the plane-averaged variances of streamwise velocity  $\bar{u}$  and TKE  $k$  as a function of the averaging length  $L$  at three levels within the canopy region, specifically, at the pedestrian level, at the mean height and at the maximum building height. The velocity and TKE variance data have very similar trends. By definition, the total variance  $\sigma^2$  is independent of the averaging length  $L$ . Generally, for both quantities, the variances are small near the ground, are larger at the mean height and the maximum height levels due to the building height distribution, and should gradually vanish above the canopy region as the flow becomes uniform.



**Fig. 8** The total, resolved and unresolved variances of the streamwise velocity  $\bar{u}$  (top panels) and the TKE  $k$  (bottom panels) as a function of the averaging length  $L$ . The variances are taken at three different heights within the canopy region: **a, d** the pedestrian level  $z/h_m = 0.05$ , **b, e** the mean building height  $z/h_m = 1.0$ , **c, f** the canopy top  $z/h_m = 2.5$

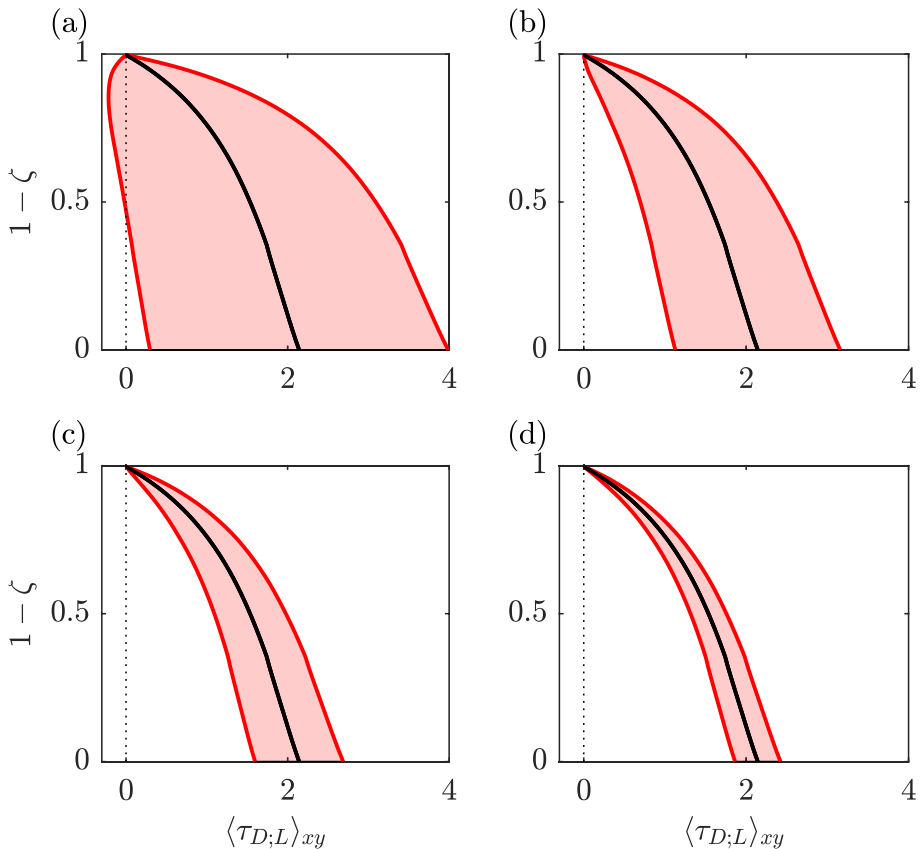
The dependence of the variances on the averaging length  $L$  is similar for all heights and both quantities. The resolved variances  $\sigma_R^2$  decrease while the unresolved variances  $\sigma_U^2$  increase with increasing  $L$ . This is consistent with Fig. 7, which shows inhomogeneity (large resolved variance) at low values of  $L$  and homogeneity (low resolved variance) at high values of  $L$ , explaining the resolved variance.

The interaction term  $\sigma_I^2$  is always negligible at very high or low values of  $L$  as either  $\sigma_R^2$  or  $\sigma_U^2$  would dominate in Eq. (39). Indeed, for very high resolutions, most variance is resolved, i.e., it is not necessary to model the unresolved part; while for very low resolutions, the variance is largely unresolved. The interaction term thus provides information on the scales at which resolved and unresolved scales interact most strongly, and thus provides information on the characteristic urban length scale  $\ell$  introduced in Sect. 2. For all figures,  $\sigma_I^2$  peaks around  $L \approx 24 - 96$  m, and therefore  $\ell \approx 50$  m for this case, which is about twice the building width  $W$ . Note that this value is consistent with the averaging length at which the resolved variance  $\sigma_R^2$  and unresolved variance  $\sigma_U^2$  are equal to each other, which would be an alternative method to infer a characteristic urban lengthscale  $\ell$ . For averaging lengths  $L > 200$  m, most variance is unresolved and homogeneity can be assumed.

### 4.3 Drag Distribution

As discussed in Sect. 2.4, the drag distribution  $\langle f_D \rangle$  or equivalently the cumulative kinematic shear stress  $\langle \tau_D \rangle$  is a key quantity that requires parameterisation for non-building resolving NWP simulations. Figure 9 shows the plane-averaged value of the cumulative drag force  $\langle \tau_{D;L} \rangle_{xy}$  at different averaging lengths  $L$  together with its standard deviation  $\sigma_R$  as a function of the scaled height  $1 - \zeta$ . Here, we have chosen  $1 - \zeta$  over  $\zeta$  so the canopy top is presented at the top of the figures. Note that  $\langle \tau_{D;L} \rangle_{xy}$  is independent of  $L$ , as shown in the black line. The variances are typically small near the canopy top and increase with decreasing height. It is also clearly visible that the spread in  $\tau_{D;L}$  increases as the averaging length  $L$  decreases. This suggests that the plane-averaged drag is not representative of the local value at a specific height.

In the following, we perform an a priori analysis at different averaging lengths  $L$  and test the performance of the Sützl et al. (2020) parameterisation Eq. (36). In order to do this, we need to make the parameterisation local, which can be done as follows:



**Fig. 9** Vertical profile of plane-averaged cumulative drag stress  $\langle \tau_{D;L} \rangle_{xy}$  in the canopy layer against the normalised frontal area  $\zeta$ . **a**  $L = 96$  m. **b**  $L = 192$  m. **c**  $L = 384$  m. **d**  $L = 786$  m. The standard deviation  $\sigma_R$  is marked by the red band

$$\frac{\tau_{D;L}}{\tau_{0;L}} = 1.88\zeta_L^3 - 3.89\zeta_L^2 + 3.01\zeta_L, \quad (44)$$

where  $\tau_{0;L} \equiv \tau_{D;L}(0) = \tau_{f;L}(0)$  is the superficially averaged kinematic surface stress and:

$$\zeta_L(\mathbf{x}_h, z) \equiv \frac{1}{\lambda_{f;L}} \int_z^h \langle \rho_L \rangle_L(\mathbf{x}_h, z') dz', \quad (45)$$

is the scale-dependent cumulative normalised frontal area, where  $\rho_L$  is the frontal area density field described in Sect. 3.2 (note that  $L_{b;L} = L^2 \langle \rho_L \rangle_L$ ). In the above equation, the scale-dependent frontal area index  $\lambda_{f;L}$  is defined as:

$$\lambda_{f;L}(\mathbf{x}_h) \equiv \int_0^h \langle \rho_L \rangle_L(\mathbf{x}_h, z') dz'. \quad (46)$$

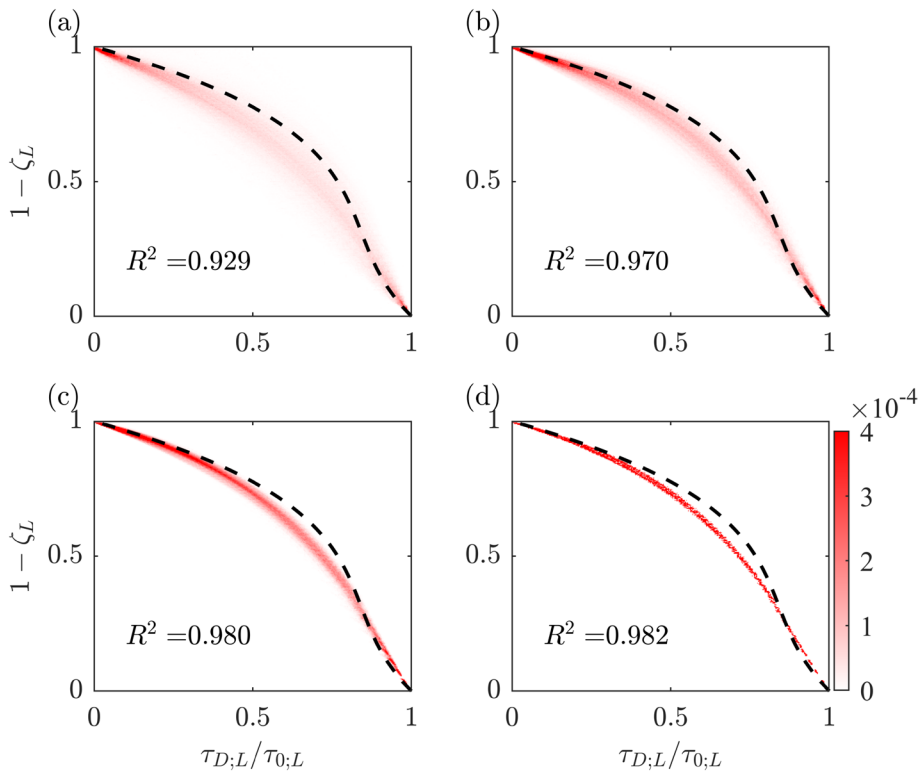
In the limit of  $L \gg \ell$ , we recover the classical quantities  $\tau_0 \equiv \tau_{0;\infty}$  and  $\lambda_f \equiv \lambda_{f;\infty}$ .

Figure 10 shows the probability density function (PDF) value of the distribution of the local cumulative drag stress  $\tau_{D;L}$  against the local scaled frontal area  $\zeta_L$ , at different resolutions and overlaid with the parametrisation Eq. (40). Figure 10c, d show that at averaging lengths  $L = 384$  and  $786$  m, there is very little scatter in the data as the fields are relatively homogeneous and the parametrisation works well with a very high coefficient of determination. This clearly shows that accurate parameterisation requires local quantities rather than plane-averaged quantities (Fig. 9).

As the resolution increases, the scatter increases due to heterogeneity effects. Indeed, Eq. (20) shows that, whenever there are heterogeneity effects, the horizontal terms in the momentum equation become important, which implies that the vertical terms are no longer in balance. The latter is, of course, what Eq. (40) is built on. However, a high coefficient of determination  $R^2$  illustrates that the parameterisation still fits the data reasonably well at  $L = 96$  m (Fig. 10a), indicating that the vertical terms still dominate. However, at averaging lengths below  $96$  m, the agreement between the data and the parameterisation quickly deteriorates, we find that the coefficient of determination dramatically reduces to  $0.23$  at  $L = 48$  m, which means a poor match, indicating that further decreasing  $L$ , the horizontal terms can no longer be ignored. Indeed, below  $L = 96$  m, there is a significant increase in the resolved variance (Fig. 8), and thus the resolved heterogeneity. Note that, the data distribution within Fig. 10 is much narrower than that shown in Fig. 9 which is associated with the same quantity  $\tau_{D;L}$ , the reason is that buildings with different heights would accumulate various of drag stress from their bottom to top, therefore, Fig. 9 shows a wide deviation. However, in Fig. 10, normalising the cumulative stress by its local value greatly reduces the spread of the data.

A potential drawback of Eq. (40) is that, even though it agrees with the data well at averaging lengths larger than  $96$  m, it does not provide the information for the superficially averaged local shear stress  $\tau_{0;L}$ . Here we model  $\tau_{0;L}$  using a single drag coefficient  $\hat{C}_d$  as (Coceal and Belcher 2004; Belcher 2005; Sützl et al. 2021):

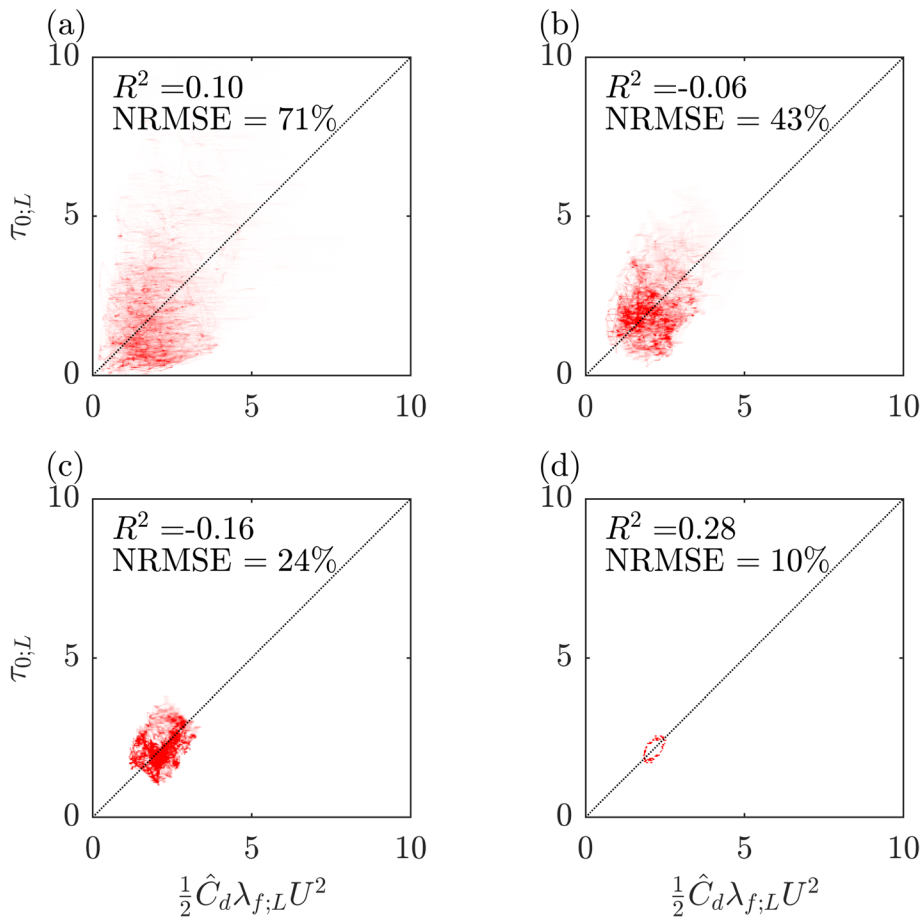
$$\tau_{0;L} = \frac{1}{2} \hat{C}_d \lambda_{f;L} U^2, \quad (47)$$



**Fig. 10** Normalised local cumulative drag stress  $\tau_{D;L}/\tau_{0;L}$  against the local scaled frontal area  $\zeta_L$  at different averaging lengths. The colour represents the statistical probability density function value of the data appearing at each location. **a**  $L = 96$  m, **b**  $L = 192$  m, **c**  $L = 384$  m, **d**  $L = 786$  m. The dashed line represents the parametrisation Eq. (40). The coefficient of determination  $R^2$  is labelled to show the a priori performance of parametrisation

where  $U = \langle \bar{u} \rangle_L(\mathbf{x}_h, z = 30)$  is the reference (superficial) wind velocity at 30 m which coincides with the mean building height. Note that  $\hat{C}_d$  is distinct from  $C_d$  in Fig. 6, which is height-dependent. Using the plane-averaged data, the appropriate value for  $\hat{C}_d$  in the parametrisation above is  $\hat{C}_d = 2.49$ .

In Fig. 11 we show the distribution of  $\tau_{0;L}$  against the parametrisation  $\frac{1}{2}\hat{C}_d\lambda_{f;L}U^2$ . For  $L = 768$  m (Fig. 11d), the data has a very small spread and has the right value, as expected since  $\hat{C}_d$  was inferred from the plane-averaged data. As  $L$  becomes smaller, the spread of the data becomes wider, although still centred around the diagonal. For  $L = 96$  m (Fig. 11a), the parametrisation does not predict the local surface stress well. However, the coefficients of determination  $R^2$  of all four figures are very poor, which is because  $R^2$  becomes less informative when the data show little variation or are very close to their mean value, even for good modellings. With this regard, we also evaluate the normalised root-mean-square error (NRMSE) to identify the a priori performance of the parametrisation—the lower the NRMSE is, the better the model predicts. This error is generally consistent with the data distribution in the figure.



**Fig. 11** Comparison between the cumulative drag  $\tau_{0;L}$  and the shear stress parameterisation Eq. (43) at various averaging lengths. **a**  $L = 96$  m, **b**  $L = 192$  m, **c**  $L = 384$  m, **d**  $L = 786$  m. The dotted line indicates perfect agreement. The coefficient of determination  $R^2$  is labelled to show the performance of the parameterisation

#### 4.4 Canopy Turbulence Closures

As discussed in Sect. 2.4, the other key process that requires parameterisation for non-building resolving NWP simulation is the unresolved stress  $\langle \tau_i \rangle$ . Consistent with (17,18), we consider the vertical flux  $\langle \tau_z \rangle = \langle \bar{w} \rangle \langle \bar{u} \rangle - \langle \bar{w} \bar{u} \rangle - \langle w' u' \rangle$ , which comprises a turbulent momentum flux  $\langle w' u' \rangle$  and a residual term  $\langle \bar{w} \rangle \langle \bar{u} \rangle - \langle \bar{w} \bar{u} \rangle$ . As demonstrated in Eq. (12), this residual term can be further decomposed associated, but for simplicity, we regard the residual as a whole and name it a dispersive-like term.

A common way to represent the unresolved fluxes in the canopy region is based on  $K$  theory, which parameterises the flux by linking the flux to the strain rate tensor  $S_{ij}$  employing an eddy diffusivity  $K_m$  (Pope 2000; Wyngaard 2010). For the vertical turbulence flux of horizontal momentum, this results in:

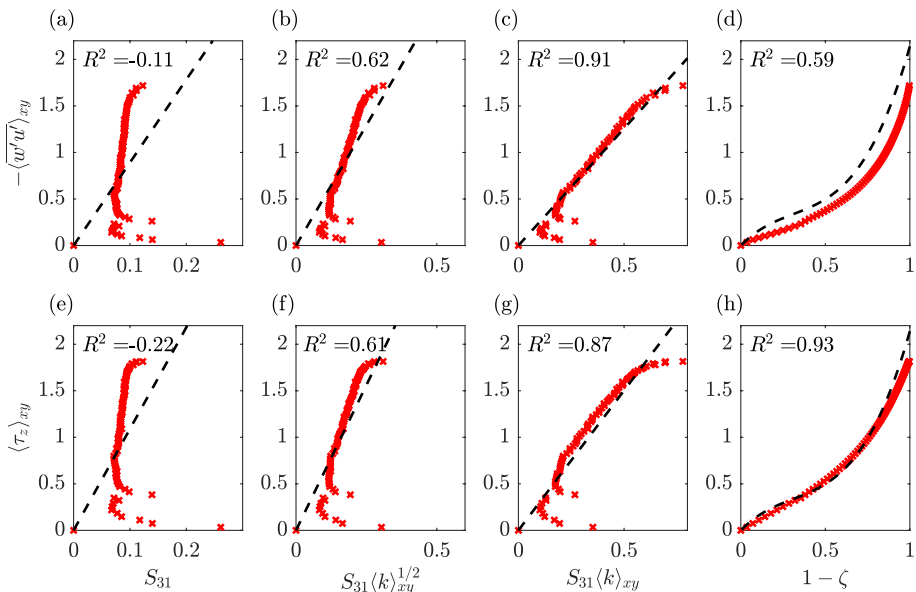
$$\langle \overline{w'u'} \rangle = -2K_m S_{31}, \quad \text{where} \quad S_{ij} \equiv \frac{1}{2} \left( \frac{\partial \langle \overline{u_i} \rangle}{\partial x_j} + \frac{\partial \langle \overline{u_j} \rangle}{\partial x_i} \right). \quad (48)$$

The eddy-diffusivity coefficient  $K_m$  can be modelled in many different ways (Hanjalić and Launder 2022). We investigate two models here, the  $k-l$  model (Deardorff 1972; Bougeault and Lacarrère 1989) and the  $k-\omega$  model (Wilcox 1998):

$$K_m = C_k l_k \langle k \rangle^{1/2} \quad \text{and} \quad K_m = \frac{\langle k \rangle}{\omega}, \quad (49)$$

respectively, where  $C_k$  is a model constant,  $l_k$  is a turbulent mixing length,  $\langle k \rangle$  is the superficially averaged turbulence kinetic energy and  $\omega$  is an inverse turbulence time scale.

First, we will study the closures for plane-averaged statistics. Figure 12a–c show the turbulence fluxes  $\langle \overline{w'u'} \rangle_{xy}$  in the canopy layer with the x-axes chosen such that a straight line starting at the origin indicates agreement with the closure. Figure 12a shows that the turbulence flux in the canopy region is practically independent of  $S_{31}$ , suggesting that the eddy diffusivity  $K_m$  cannot be assumed constant in the canopy layer. The best-fit value (associated with the dashed line) is  $K_m = 8.91 \text{ m}^2/\text{s}$ , which is greater than the value of  $1 \text{ m}^2/\text{s}$  reported by Lu et al. (2024). This is likely due to the difference in geometry. However, the negative coefficient of determination  $R^2$  indicates that the line fitting from the origin shows a poor match. Figure 12b shows that the mixing length model performs better, but although the data is pretty much on a straight line it does not pass through the origin. The best-fit results in  $l_k = 13.00 \text{ m}$ , here we have assumed the model constant  $C_k = 0.4$  (Bougeault



**Fig. 12** Suitability of various turbulence closures to parameterise the turbulence shear stress  $-\langle \overline{w'u'} \rangle_{xy}$  (top row) and the total unresolved stress  $\langle \tau_z \rangle_{xy}$  (bottom row). **a, e** Constant eddy viscosity, **b, f**  $k-l$  model, **c, g**  $k-\omega$  model, and **d, h** parameterisation Eq. (40). Agreement of the data with the dashed lines implies agreement with the model. The data are shown within the canopy region. The coefficient of determination  $R^2$  is labelled to show the a priori performance of parametrisation

and Lacarrère 1989). The ratio  $C_k l_k / h_{\max} = 0.07$  is comparable with (Nazarian et al. 2020) at the same urban density  $\lambda_p$ . Only for the  $k - \omega$  model (Fig. 12c) does the data compare well with a line through the origin (with  $R^2$  very close to 1). The best-fit returns a value  $\omega = 0.40 \text{ s}^{-1}$ .

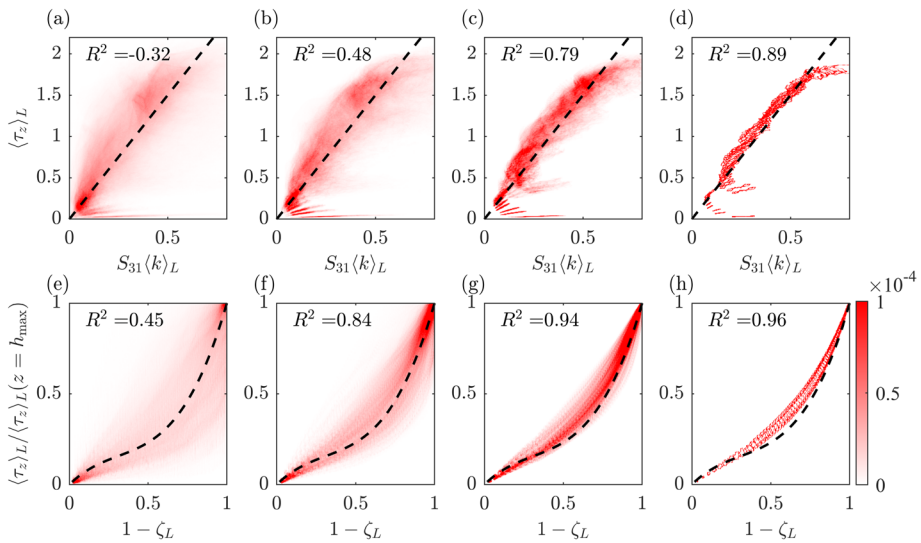
Figure 12e–g present the same figures, but now using the kinematic unresolved shear stress  $\langle \tau_z \rangle_{xy}$  which includes the dispersive-like term. The small difference between Fig. 12a–c and e–g is due to the fact that the turbulence fluxes dominate everywhere in the canopy layer (Fig. 4b). All the figures show that considering the dispersive-like terms slightly reduces the models' performance. However, again, the  $k - \omega$  model can be seen to perform the best with the best-fit values  $\omega = 0.33 \text{ s}^{-1}$ .

Another way to parameterise  $\langle \tau_z \rangle$  without having to rely on auxiliary variables like  $k$  is to use the drag parametrisation Eq. (40). Invoking the constant stress layer assumption often used in boundary layer theory (Schlichting and Gersten 2017) implies that  $\tau_f \approx \tau_0$  in the canopy layer, which after substitution into Eq. (25) and making use of Eq. (40) immediately results in:

$$\frac{\langle \tau_z \rangle_{xy}}{\tau_0} = 1 - \frac{\tau_D}{\tau_0} = 1 - \zeta^3 - 3.89\zeta^2 + 3.01\zeta. \quad (50)$$

The distribution of  $\langle \overline{w'u'} \rangle_{xy}$  and  $\langle \tau_z \rangle_{xy}$  against  $\zeta$  are shown in Fig. 12d, h, respectively. Since the parameterisation is derived for  $\langle \tau_z \rangle_{xy}$ , it always overestimates the  $\langle \overline{w'u'} \rangle_{xy}$  by definition—the total stress is in better agreement with the parameterisation than the turbulence flux only (Fig. 12h). Near the top, for  $1 - \zeta > 0.9$ , the parameterisation slightly overestimates the total stress. Comparing the bottom row, the parameterisations in Fig. 12(g,h) of total unresolved stress have the largest  $R^2$  to fit the data.

Figure 13 shows the two best-performing parameterisations of total unresolved stress at different averaging lengths  $L$ . The top panels show the  $k - \omega$  model predictions for the total unresolved shear stress  $\langle \tau_z \rangle_L$  and the bottom row shows the predictions derived from the drag parameterisation Eq. (46). Note that the stresses in the bottom row are normalised by  $\langle \tau_z \rangle_L$  at maximum building height  $z = h_{\max}$  in order to converge to the unit. However, under the constant stress layer assumption (Schlichting and Gersten 2017),  $\langle \tau_z \rangle_L (z = h_{\max}) \approx \tau_0$  which is consistent with the parameterisation Eq. (46). In order to create a scale-dependent parameterisation for Eq. (46), simply replace  $\langle \tau_z \rangle_{xy}$  with  $\langle \tau_z \rangle_L$ ,  $\tau_0$  with  $\tau_{0;L}$ , and  $\zeta$  with  $\zeta_L$ . As expected, as the averaging length reduces, the data spread around the diagnostic line increases with the reduced coefficient of determination  $R^2$ . However,  $R^2$  is larger in the bottom row than the top row at the same resolution, suggesting that the parameterisation Eq. (46) has higher fidelity than the  $k - \omega$  closure. We also note that, compared with the drag parameterisation (Fig. 10), the coefficients  $R^2$  of stress parameterisation Eq. (46) are lower at the same resolution, and reduce more rapidly with the reducing resolution. This might be because of the assumption of constant stress.



**Fig. 13** Skill of the  $k - \omega$  model (top row) and Eq. (40) (bottom row) to parameterise the total unresolved vertical stress  $\langle \tau_z \rangle_L$  at various averaging lengths. **a, e**  $L = 96$  m, **b, f**  $L = 192$  m, **c, g**  $L = 384$  m, **d, h**  $L = 768$  m. Overlaid with the parameterisation from Fig. 12g, h for the top and bottom row, respectively. The coefficient of determination  $R^2$  is labelled to show the *a priori* performance of parameterisation

## 5 Conclusion

A multi-scale framework was developed for investigating the flows over heterogeneous urban areas. At the core of the framework is a coarse-graining method that employs convolution filters to obtain coarse-grained fields from a building-resolved high-resolution simulation. By employing Fast-Fourier transforms, a computationally efficient procedure was developed to carry out coarse-graining, which is particularly important for the analysis of high-resolution simulations as considered here.

One of the central questions the paper aimed to address was: "What are the requirements for non-building-resolving NWP models as their resolution is increased?" To answer this question, a quasi-realistic case was considered, inspired by the building height distribution of London. The multi-scale framework revealed that the flow has a characteristic length scale  $\ell \approx 50$  m, which is the resolution at which as much variance is resolved as is unresolved. This corresponds to roughly twice the building width. Heterogeneity effects, which challenge the NWP parameterisations for the land surface, become significant at resolutions of about 200 m and below for the case under consideration.

Given that future NWP models will likely run at resolutions at which heterogeneity effects are important, an immediate follow-up question is how to incorporate heterogeneity. This requires a land-surface model that resolves the vertical distribution of the flow and turbulence (Lean et al. 2024) and requires: i) a model for the distribution of the drag over the canopy; and ii) a model for the unresolved turbulence and dispersive stresses in the canopy. Since NWP models do not resolve the urban morphology and it is undesirable to have to modify the governing equations inside the canopy (e.g. the mass conservation equation), superficial (comprehensive) averages are the appropriate formulation to consider, see Eq.

(17). It is therefore highly desirable that the drag distribution and unresolved stress parameterisation are formulated in terms of superficial quantities also.

It was shown that the distributed drag parameterisation of Sützl et al. (2020), which was obtained from plane-averaged analysis of an ensemble of eight different urban landscapes, can be successfully extended to a local formulation, cf Eq. (40). The parameterisation generalises excellently, with an  $R^2$  value of 0.929 at 96 m resolution. The attractive feature of this parameterisation is that it only requires the local distributed frontal area  $\zeta_L$  which can be calculated *a priori* and can be straightforwardly incorporated in NWP models (Sützl et al. 2021). Although this parameterisation is excellent for predicting the local drag distribution, the actual drag requires knowledge of the local surface stress which remains challenging and requires further work; a local version of the Coceal and Belcher (2004); Belcher (2005) parameterisation Eq. (43) has a normalised root-mean-square error of 71% at 100 m resolution.

Several closures were considered for the unresolved and turbulent stresses in the urban canopy layer. The two best-performing models were the  $k - \omega$  model, which requires information about the turbulence kinetic energy  $k$  and the inverse turbulence timescale  $\omega$ , and a new stress parameterisation based on the distributed drag, cf Eq. (46), that does not require further parameters. At a resolution of 96 m, the latter performs better than the former, although both parameterisations have a substantial scatter associated with local heterogeneity effects.

The results reported here indicate that it is possible to incorporate substantial heterogeneity effects in non-building-resolving NWP models with relative ease. However, further work is required. First, it is desirable to ensure the validity of the parameterisations Eq. (36) and Eq. (46) on a much larger flow database, e.g. that of Lu et al. (2024). Furthermore, the atmosphere is non-neutral most of the time, and distributed versions of the sensible and latent heat fluxes, including the interaction with the surface stress, need to be explored.

## Appendix A: Volume-Averaging with Convolution Filters

The Whitaker (1999) volume averaging approach can be cast into a convolution filter approach by introducing a kernel  $\mathcal{G}$  to specify the averaging volume, which is a function that is  $1/V$  within the averaging domain and 0 outside of it, where  $V$  is the total averaging volume. Using this kernel, the superficial volume average of an arbitrary scalar  $\varphi(\mathbf{x})$  is given by:

$$\langle \varphi \rangle(\mathbf{x}) \equiv \int_{\Omega_f(\mathbf{x})} \mathcal{G}(\mathbf{x} - \mathbf{y}) \varphi(\mathbf{y}) d\mathbf{y}, \quad (\text{A1})$$

where  $\mathbf{x}$  is the 3-D location and  $\int \mathcal{G}(\mathbf{x}) d\mathbf{x} = 1$ , i.e. the kernel is normalised. Note that in Eq. (47), integration only takes place over the fluid phase. The fluid volume fraction  $\varepsilon = V_f/V$ , where  $V_f$  is the volume occupied by fluid inside the averaging volume, is given by:

$$\varepsilon(\mathbf{x}) = \int_{\Omega_f(\mathbf{x})} \mathcal{G}(\mathbf{x} - \mathbf{y}) d\mathbf{y}. \quad (\text{A2})$$

## Appendix A.1: Volume to Area Integration

An area-averaging filter can be obtained by defining  $\mathcal{G}$  as:

$$\mathcal{G}(\mathbf{x}) = \mathcal{A}(\mathbf{x}_h)\delta(z), \quad (\text{A3})$$

where  $\delta$  is the Kronecker delta and  $\mathbf{x} = [\mathbf{x}_h, z]^T$ . With this definition of  $\mathcal{G}$ , the volume integral will become an area integral. The planar filter is also normalised:

$$\int \mathcal{G}(\mathbf{x})d\mathbf{x} = \int \int \mathcal{A}(\mathbf{x}_h)\delta(z)d\mathbf{x}_h dz = \int \mathcal{A}(\mathbf{x}_h)d\mathbf{x}_h = 1. \quad (\text{A4})$$

Using Eq. (49), the superficial average Eq. (47) becomes:

$$\langle \varphi \rangle(\mathbf{x}) = \int_{\Omega_f(\mathbf{x})} \mathcal{A}(\mathbf{x}_h - \mathbf{y}_h)\varphi(\mathbf{y}_h, z)d\mathbf{y}_h. \quad (\text{A5})$$

The surface fluid fraction  $\varepsilon = A_f/A$ , where  $A_f$  is the fluid area and  $A$  is the filter area, is given by:

$$\varepsilon(\mathbf{x}) = \int_{\Omega_f(\mathbf{x})} \mathcal{A}(\mathbf{x}_h - \mathbf{y}_h)d\mathbf{y}_h, \quad (\text{A6})$$

## Appendix A.2: Differentiation Rules

We start with deriving the differentiation rule for the divergence. Note that:

$$\langle \nabla \cdot \mathbf{F} \rangle = \int_{\Omega_f} \mathcal{A} \nabla \cdot \mathbf{F} d\mathbf{y}_h = \int_{\Omega_f} \nabla \cdot (\mathcal{A} \mathbf{F}) d\mathbf{y}_h - \int_{\Omega_f} \mathbf{F} \cdot \nabla \mathcal{A} d\mathbf{y}_h \quad (\text{A7})$$

Commuting the integration and differentiation operators in the first term needs to be done with care, since  $\Omega_f$  is dependent on space. It requires the identity (see Van Reeuwijk et al. 2021, for a derivation):

$$\int_{\Omega_f} \nabla \cdot \mathbf{G} d\mathbf{y}_h = \frac{\partial}{\partial z} \int_{\Omega_f} G_z d\mathbf{y}_h - \oint_{\partial\Omega_f} \frac{\mathbf{G} \cdot \mathbf{N}}{|\mathbf{N}_h|} ds. \quad (\text{A8})$$

Using Eqs. (53), (52) becomes:

$$\langle \nabla \cdot \mathbf{F} \rangle = \frac{\partial \langle F_z \rangle}{\partial z} - \int_{\Omega_f} \mathbf{F}_h \cdot \nabla_h \mathcal{A} d\mathbf{y}_h - \oint_{\partial\Omega_f} \mathcal{A} \frac{\mathbf{F} \cdot \mathbf{N}}{|\mathbf{N}_h|} ds, \quad (\text{A9})$$

where use was made of the fact that  $\partial\mathcal{A}/\partial z = 0$ . The first term in this expression is the vertical gradient of the superficial average of  $F_z$ . The second term represents the filtered horizontal gradient (see below). The last term represents the fluxes that are exchanged between

the fluid and the solid domain inside the averaging area. Note that (Leonard 1974; Pope 2000):

$$\begin{aligned}
 \int_{\Omega_f} \mathbf{F}_h(\mathbf{y}_h) \cdot \nabla_h \mathcal{A}(\mathbf{x}_h - \mathbf{y}_h) d\mathbf{y}_h &= \int_{\Omega_f} \mathbf{F}_h(\mathbf{y}_h) \cdot \frac{\partial}{\partial \mathbf{y}_h} \mathcal{A}(\mathbf{x}_h - \mathbf{y}_h) d\mathbf{y}_h \\
 &= - \int_{\Omega_f} \mathbf{F}_h(\mathbf{y}_h) \cdot \frac{\partial}{\partial \mathbf{x}_h} \mathcal{A}(\mathbf{x}_h - \mathbf{y}_h) d\mathbf{y}_h \quad (\text{A10}) \\
 &= - \frac{\partial}{\partial \mathbf{x}_h} \cdot \int_{\Omega_f} \mathbf{F}_h(\mathbf{y}_h) \mathcal{A}(\mathbf{x}_h - \mathbf{y}_h) d\mathbf{y}_h \\
 &= - \nabla_h \cdot \langle \mathbf{F}_h \rangle,
 \end{aligned}$$

which implies that the vector form of the spatial averaging theorem takes the form:

$$\langle \nabla \cdot \mathbf{F} \rangle = \nabla \cdot \langle \mathbf{F} \rangle - \oint_{\partial\Omega_f} \mathcal{A} \frac{\mathbf{F} \cdot \mathbf{N}}{|\mathbf{N}_h|} ds. \quad (\text{A11})$$

Substitution of  $\mathbf{F} = \varphi \mathbf{e}_i$  where  $i \in \{x, y, z\}$  into Eq. (55) results in the the spatial averaging theorem for scalars:

$$\langle \nabla \varphi \rangle = \nabla \langle \varphi \rangle - \oint_{\partial\Omega_f} \mathcal{A} \varphi \frac{\mathbf{N}}{|\mathbf{N}_h|} ds. \quad (\text{A12})$$

**Acknowledgements** The support of the ARCHER2 UK National Supercomputing Service (project ARCHER2-eCSE05-3) and the NERC highlight grant ASSURE: Across-Scale Processes in Urban Environment (NE/W002868/1) is acknowledged.

**Author Contributions** M. provided the paper idea, derived the theory of the framework, reviewed the manuscript and supervised the work. J. performed the simulation, wrote the main manuscript, and prepared figures.

**Data Availability** The data that support the findings of this study, as well as the coarse-graining routines, are openly available in Zenodo at van Reeuwijk and Huang (2025).

## Declarations

**Conflict of interest** The authors report no Conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Baldauf M, Seifert A, Förstner J, Majewski D, Raschendorfer M, Reinhardt T (2011) Operational convective-scale numerical weather prediction with the cosmo model: description and sensitivities. *Mon Weather Rev* 139(12):3887–3905
- Belcher SE (2005) Mixing and transport in urban areas. *Philos Trans A Math Phys Eng Sci* 363:2947–68
- Belušić D, de Vries H, Dobler A, Landgren O, Lind P, Lindstedt D, Pedersen RA, Sánchez-Perrino JC, Toivonen E, van Ulft B, Wang F, Andrae U, Batrak Y, Kjellström E, Lenderink G, Nikulin G, Pietikäinen JP, Rodríguez-Camino E, Samuelsson P, van Meijgaard E, Wu M (2020) Holim38: a flexible regional climate model applicable for different climate zones from coarse to convection-permitting scales. *Geosci Model Dev* 13(3):1311–1333
- Bougeault P, Lacarrère P (1989) Parameterization of orographic induced turbulence in a mesobeta scale model. *Mon Weather Rev* 117:1872–1890
- Bracewell RN (2000) The Fourier transform and its applications. McGraw Hill, New York
- Brun C, Blein S, Chollet JP (2017) Large-eddy simulation of a katabatic jet along a convexly curved slope. Part I: statistical results. *J Atmos Sci* 74(12):4047–4073
- Bryan GH, Wyngaard JC, Fritsch JM (2003) Resolution requirements for the simulation of deep moist convection. *Mon Weather Rev* 131(10):2394–2416
- Coccal O, Belcher SE (2004) A canopy model of mean winds through urban areas. *Q J R Meteorol Soc* 130(599):1349–1372
- Coccal O, Thomas TG, Castro IP, Belcher SE (2006) Mean flow and turbulence statistics over groups of urban-like cubical obstacles. *Bound-Layer Meteorol* 121(3):491–519
- Coccal O, Thomas T, Belcher SE (2007) Spatial variability of flow statistics within regular building arrays. *Bound-Layer Meteorol* 125:537–552
- Deardorff JW (1972) Theoretical expression for the countergradient vertical heat flux. *J Geophys Res* (1896–1977) 77(30):5900–5904
- Finnigan J (2000) Turbulence in plant canopies. *Annu Rev Fluid Mech* 32:519–571
- Giometto M, Christen A, Meneveau C, Fang J, Krafczyk M, Parlange M (2016) Spatial characteristics of roughness sublayer mean flow and turbulence over a realistic urban surface. *Bound-Layer Meteorol* 160(3):425–452
- Grimmond CSB, Oke TR (1999) Aerodynamic properties of urban areas derived from analysis of surface form. *J Appl Meteorol* 38(9):1262–1292
- Hanjalić K, Lauder B (2022) Modelling turbulence in engineering and the environment: rational alternative routes to closure, 2nd edn. Cambridge University Press, Cambridge
- Honnert R, Masson V, Couvreur F (2011) A diagnostic for evaluating the representation of turbulence in atmospheric models at the kilometric scale. *J Atmos Sci* 68(12):3112–3131
- Honnert R, Efstathiou GA, Beare RJ, Ito J, Lock A, Neggers R, Plant RS, Shin HH, Tomassini L, Zhou B (2020) The atmospheric boundary layer and the “gray zone” of turbulence: a critical review. *J Geophys Res Atmos* 125(13):e2019JD030
- Jackson PS (1981) On the displacement height in the logarithmic velocity profile. *J Fluid Mech* 111:15–25
- Kanda M, Inagaki A, Miyamoto T, Gryschka M, Raasch S (2013) A new aerodynamic parametrization for real urban surfaces. *Bound-Layer Meteorol* 148:357–377
- Lean HW, Theeuwes NE, Baldauf M, Barkmeijer J, Bessardon G, Blunn L, Bojarova J, Boutle IA, Clark PA, Demuzere M, Dueben P, Frogner IL, Haan S, Harrison D, Cv H, Honnert R, Lock A, Marsigli C, Masson V, McCabe A, Mv R, Roberts N, Siebesma P, Smolíkova P, Yang X (2024) The hectometric modelling challenge: gaps in the current state of the art and ways forward towards the implementation of 100-m scale weather and climate models. *Q J R Meteorol Soc* 150(765):4671–4708
- Leonard A (1974) Energy cascade in large eddy simulations of turbulent fluid flow. *Adv Geophys* 18A:237–248
- Lipson MJ, Grimmond S, Best M, Abramowitz G, Coutts A, Tapper N, Baik JJ, Beyers M, Blunn L, Boussetta S, Bou-Zeid E, De Kauwe MG, de Munck C, Demuzere M, Fatichi S, Fortuniak K, Han BS, Hendry MA, Kikegawa Y, Kondo H, Lee DI, Lee SH, Lemonsu A, Machado T, Manoli G, Martilli A, Masson V, McNorton J, Meili N, Meyer D, Nice KA, Oleson KW, Park SB, Roth M, Schoetter R, Simón-Moral A, Steeneveld GJ, Sun T, Takane Y, Thatcher M, Tsiringakis A, Varentsov M, Wang C, Wang ZH, Pitman AJ (2024) Evaluation of 30 urban land surface models in the urban-plumber project: phase 1 results. *Q J R Meteorol Soc* 150(758):126–169
- Lu J, Nazarian N, Hart MA, Kravynhoff ES, Martilli A (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(7):2525–2545
- Macdonald R, Griffiths R, Hall D (1998) An improved method for the estimation of surface roughness of obstacle arrays. *Atmos Environ* 32(11):1857–1864

- Masson V, Heldens W, Bocher E, Bonhomme M, Bucher B, Burmeister C, de Munck C, Esch T, Hidalgo J, Kanani-Sühring F, Kwok YT, Lemonsu A, Lévy JP, Maronga B, Pavlik D, Petit G, See L, Schoetter R, Tornay N, Votsis A, Zeidler J (2020) City-descriptive input data for urban climate models: model requirements, data sources and challenges. *Urban Clim* 31(100):536
- Mishra A, Placidi M, Carpentieri M, Robins A (2023) Wake characterization of building clusters immersed in deep boundary layers. *Bound-Layer Meteorol* 189:1–25
- Mishra A, Carpentieri M, Robins A, Placidi M (2024) Experimental study of the turbulent characteristics in the wake of tall building clusters. *Flow* 4:E15
- Nazarian N, Krayenhoff ES, Martilli A (2020) A one-dimensional model of turbulent flow through “urban” canopies (mlucm v2.0): updates based on large-eddy simulation. *Geosci Model Dev* 13(3):937–953
- Nepf HM (2012) Flow and transport in regions with aquatic vegetation. *Annu Rev Fluid Mech* 44:123–142
- Oke TR, Mills G, Christen A, Voogt JA (2017) *Urban climates*. Cambridge University Press, Cambridge
- Owens SO, Majumdar D, Wilson CE, Bartholomew P, Reeuwijk M (2024) A conservative immersed boundary method for the multi-physics urban large-eddy simulation model uDALES v2.0. *EGU sphere* 2024:1–33
- Placidi M, Ganapathisubramani B (2015) Effects of frontal and plan solidities on aerodynamic parameters and the roughness sublayer in turbulent boundary layers. *J Fluid Mech* 782:541–566
- Pope SB (2000) *Turbulent flows*. Cambridge University Press, Cambridge
- Prein AF, Langhans W, Fossler G, Ferrone A, Ban N, Goergen K, Keller M, Tölle M, Gutjahr O, Feser F, Brisson E, Kollet S, Schmidli J, van Lipzig NPM, Leung R (2015) A review on regional convection-permitting climate modeling: demonstrations, prospects, and challenges. *Rev Geophys* 53(2):323–361
- Raupach MR, Shaw RH (1982) Averaging procedures for flow within vegetation canopies. *Bound-Layer Meteorol* 22:79–90
- Santiago JL, Martilli A (2010) A dynamic urban canopy parameterization for mesoscale models based on computational fluid dynamics Reynolds-averaged Navier–Stokes microscale simulations. *Bound-Layer Meteorol* 137(3):417–439
- Schlichting H, Gersten K (2017) *Bound-Layer Theory*. Springer, Berlin
- Schmid M, Lawrence GA, Parlange MB, Giometto MG (2019) Volume averaging for urban canopies. *Bound-Layer Meteorol* 173:349–372
- Schoetter R, Kwok YT, Munck C, Lau KKL, Wong WK, Masson V (2020) Multi-layer coupling between surfex-teb-v9.0 and meso-nh-v5.3 for modelling the urban climate of high-rise cities. *Geosci Model Dev* 13(11):5609–5643
- Smith DK, Renfrew IA, Dorling SR, Price JD, Boutle IA (2021) Sub-km scale numerical weather prediction model simulations of radiation fog. *Q J R Meteorol Soc* 147(735):746–763
- Stewart ID, Oke TR (2012) Local climate zones for urban temperature studies. *Bull Am Meteorol Soc* 93(12):1879–1900
- Suter I, Grylls T, Sützl BS, Owens SO, Wilson CE, Reeuwijk M (2022) uDALES 1.0: a large-eddy simulation model for urban environments. *Geosci Model Dev* 15(13):5309–5335
- Sützl BS, Rooney GG, van Reeuwijk M (2020) Drag distribution in idealized heterogeneous urban environments. *Bound-Layer Meteorol* 178:225–248
- Sützl BS, Rooney GG, Finnenkoetter A, Bohnenstengel SI, Grimmond S, van Reeuwijk M (2021) Distributed urban drag parametrization for sub-kilometre scale numerical weather prediction. *Q J R Meteorol Soc* 147(741):3940–3956
- Tang Y, Lean HW, Bornemann J (2013) The benefits of the met office variable resolution NWP model for forecasting convection. *Meteorol Appl* 20(4):417–426
- Uno I, Cai XM, Steyn DG, Emori S (1995) A simple extension of the louis method for rough surface layer modelling. *Bound-Layer Meteorol* 76:395–409
- Reeuwijk M, Vassilicos JC, Craske J (2021) Unified description of turbulent entrainment. *J Fluid Mech* 908:A12
- van Reeuwijk M, Huang J (2025) Dataset for “multi-scale analysis of flow over heterogeneous urban environments”. <https://doi.org/10.5281/zenodo.17102845>
- Vreman AW (2004) An eddy-viscosity subgrid-scale model for turbulent shear flow: algebraic theory and applications. *Phys Fluids* 16(10):3670–3681
- Whitaker S (1999) *The method of theory and applications of transport in porous media, vol averaging*. Springer, Dordrecht
- Wilcox DC (1998) *Turbulence modeling for CFD*. DCW Industries, La Canada
- Wyngaard JC (2004) Toward numerical modeling in the “terra incognita”. *J Atmos Sci* 61(14):1816–1826
- Wyngaard JC (2010) *Turbulence in the atmosphere*. Cambridge University Press, Cambridge
- Xie ZT, Fuka V (2018) A note on spatial averaging and shear stresses within urban canopies. *Bound-Layer Meteorol* 167:171–179
- Xie ZT, Coceal O, Castro IP (2008) Large-eddy simulation of flows over random urban-like obstacles. *Bound-Layer Meteorol* 129:1–23

Yu T, Sützl BS, Reeuwijk M (2023) Urban neighbourhood classification and multi-scale heterogeneity analysis of greater London. *Environ Plan B-Urban* 50(6):1534–1558

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.